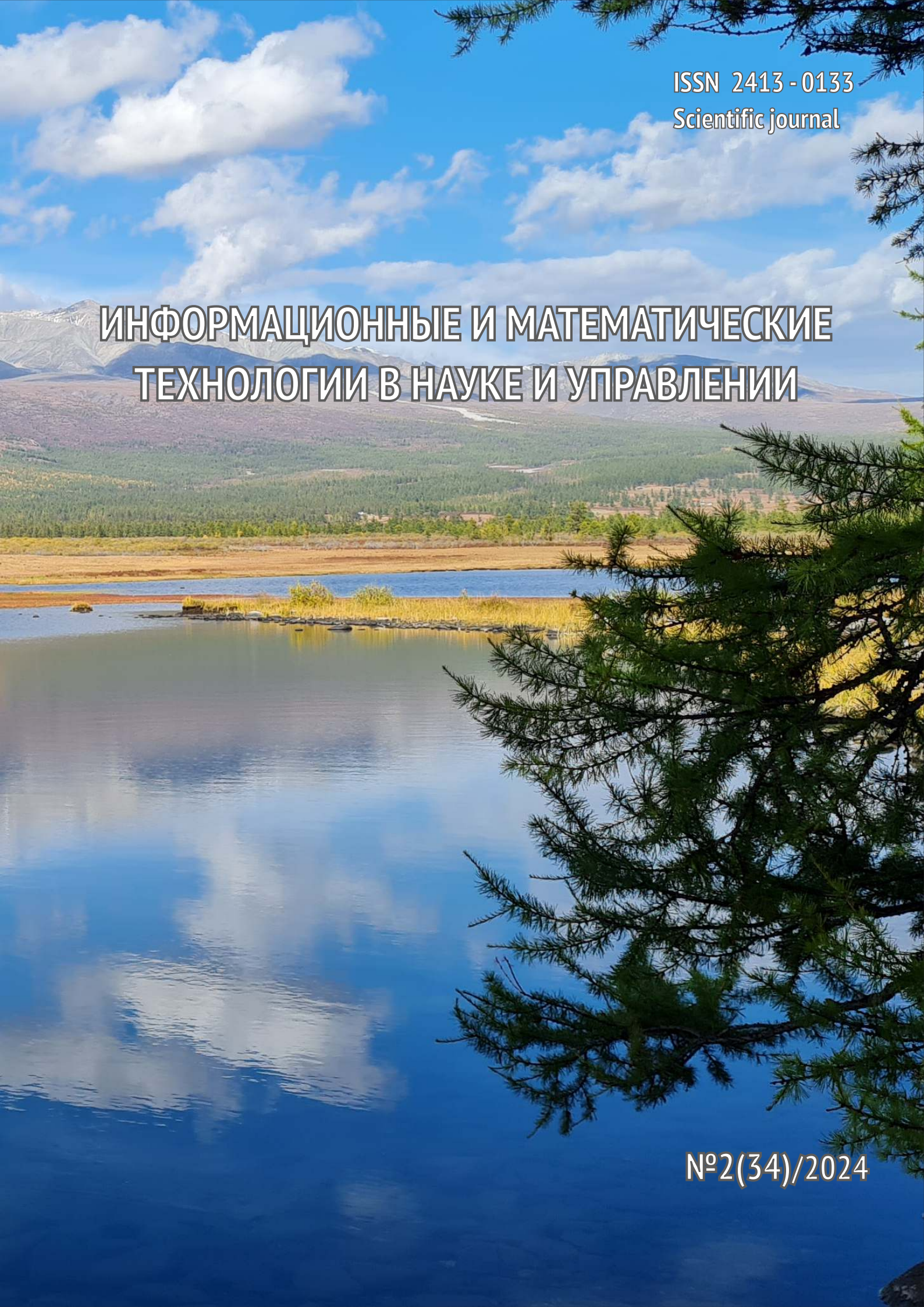
ISSN 2413 - 0133
Scientific journal

ИНФОРМАЦИОННЫЕ И МАТЕМАТИЧЕСКИЕ ТЕХНОЛОГИИ В НАУКЕ И УПРАВЛЕНИИ



№2(34)/2024

Information and mathematical technologies in science and management

ISSN 2413 – 0133
Scientific journal

ИНФОРМАЦИОННЫЕ И МАТЕМАТИЧЕСКИЕ ТЕХНОЛОГИИ В НАУКЕ И УПРАВЛЕНИИ

Научный журнал
№ 2 (34)

© Институт систем энергетики им. Л.А. Мелентьева
Сибирского отделения Российской академии наук

2024

ISSN 2413 – 0133 Индекс Е79114*

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.	Вѐрн Х., Германия, Karlsruhe Institute of Technology (KIT)
Voevodin V.V.	Воеводин В.В., чл.-корр. РАН, Москва, НИВЦ МГУ
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.	Грумпос П., 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, IIASA
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
Executive Editor	Bakhvalova Z.A.	Выпускающий редактор	Бахвалова З.А.	к.т.н., Иркутск, ИСЭМ СО РАН
Editor	Gaskova D.A.	Редактор	Гаськова Д.А.	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»
Адрес типографии: 664046, г. Иркутск, ул. Волжская, 14
*Подписка на журнал доступна со второго полугодия 2021 г.
Дата выхода 19.06.2024 г. Тираж 50 экз.

CC BY
© Издательство ИСЭМ СО РАН

Цена свободная (12+)

Содержание

Методологические аспекты информационных, интеллектуальных и математических технологий

Черкашин А.К. Политическая география и геополитика: герменевтические исследования и математическое моделирование электоральных явлений **5**

Дорогов А.Ю. Быстрые преобразования и самоподобные нейронные сети глубокого обучения. Часть 3. Методы обучения быстрых нейронных сетей **19**

Сенашова М.Ю. Упорядоченность значений GC-состава фрагментов в пространственной структуре геномов органелл **33**

Методы, технологии и приложения искусственного интеллекта

Бойко А.В., Бахшиев А.В., Корсаков А.М. Аппаратная реализация сегментной спайковой модели нейрона CSNM на операционных усилителях **41**

Исаков Т.Т., Бахшиев А.В., Корсаков А.М. Решение задачи классификации с использованием сегментной спайковой модели нейрона на ПЛИС **50**

Доррер М.Г. Оценка качества интерполяции плотности распределения случайной величины при помощи искусственной нейронной сети **59**

Дородных Н.О., Юрин А.Ю. Использование диаграмм переходов состояний для автоматизированного создания баз знаний **69**

Математическое моделирование и его применение в научных исследованиях

Бахвалов С.В., Маланова Т.В., Янхаев Д.В. Транспортная задача с ограничением на время перевозок **82**

Серебровская Е.А., Марчевский И.К., Ерофеева М.А. Алгоритм сглаживания пути на триангулированной поверхности **90**

Математические, информационные и интеллектуальные технологии в энергетике

Погодаева И.Н., Шакиров В.А. Выбор компонентов гибридного энергокомплекса для удаленных районов с использованием методов многокритериальной оценки **99**

Крюков А.В., Черепанов А.В., Нгуен Куок Хиеу. Моделирование процессов плавки гололеда в тяговых сетях постоянного тока **112**

Цифровая экономика и управление

Алдошкина А.М., Пенькова Т.Г., Ноженкова Л.Ф. Поддержка принятия решений в территориальном управлении на основе метода интегрального оценивания качества жизни (на примере Красноярского края) **123**

Сиротинин А.А., Володько О.С. Построение оптимизационной модели беспроводной внутренней сети для использования технологии интернета вещей **135**

Информационные системы и технологии

Малимонов М.И. Система учета заявок для ИТ-подразделений КНЦ СО РАН **144**

Программные системы и комплексы

Опарин Г.А., Богданова В.Г., Пашинин А.А. Инструментальные средства автоматизации разработки и применения пакета прикладных микросервисов **155**

Животова А.А., Бердоносов В.Д. Оптимизационное предредактирование узкоспециальных русскоязычных текстов для их машинного перевода на английский язык **169**

Content

Methodological aspects of information, intellectual and mathematical technologies

Cherkashin A.K. Political geography and geopolitics: hermeneutical research and mathematical modeling of electoral phenomena **5**

Dorogov A.Yu. Fast transformations and self-similar deep learning neural networks. Part 3. Pyramidal neural networks with a deep degree of learning **19**

Senashova M.Yu. Orderliness of GC-content values of fragments in the spatial structure of organelle genomes **33**

Methods, technologies and applications of artificial intelligence

Boyko A.V., Bakhshiev A.V., Korsakov A.M. Operational-amplifier-based hardware implementation of the compartmental spike model of the CSNM neuron **41**

Isakov T.T., Bakhshiev A.V., Korsakov A.M. Solution of the classification task using a compartmental spiking neuron model on an FPGA **50**

Dorrer M.G. Evaluation of the quality of interpolation of the random variable distribution density using an artificial neural network **59**

Dorodnykh N.O., Yurin A.Yu. Using state transition diagrams for automated creation of knowledge bases **69**

Mathematical modeling and its application in scientific research

Bakhvalov S.V., Malanova T.V., Yanhaev D.V. Transport problem with transportation time constraint **82**

Serebrovskaya E.A., Marchevsky I.K., Erofeeva M.A. Path smoothing algorithm for the triangulated surface **90**

Mathematical, information and intellectual technologies in the energy sector

Pogodaeva I.N., Shakirov V.A. Selection of components of a hybrid energy complex for remote areas using multi-criteria assessment methods **99**

Kryukov A.V., Cherepanov A.V., Nguyen Q.H. Modeling modes of ice melting in DC traction networks **112**

Digital economy and management

Aldoshkina A.M., Penkova T.G., Nozhenkova L.F. Decision making support in territory management based on the method of integral estimation of the life quality (in case of Krasnoyarsk region) **123**

Sirotinin A.A., Volodko O.S. Building an optimization model of a wireless internal network for using internet of things technology **135**

Information systems and technologies

Malimonov M.I. Ticketing system for IT-department of KSC SB RAS **144**

Software systems and complexes

Oparin G.A., Bogdanova V.G., Pashinin A.A. Automating tools for the development and application of an applied microservices package **155**

Zhivotova A.A., Berdonosov V.D. Optimizational pre-editing of highly specialized Russian-language texts for its machine translation into English **169**

Методологические аспекты информационных, интеллектуальных и математических технологий

УДК 51-77: 911.3:327:324
DOI:10.25729/ESI.2024.34.2.001

Политическая география и геополитика: герменевтические исследования и математическое моделирование электоральных явлений

Черкашин Александр Константинович

Институт географии им. В.Б. Сочавы СО РАН,
Россия, Иркутск, *akcherk@irnok.net*

Аннотация. На основе критического математического подхода, базирующегося на процедурах расслоения знаний, доказывается различие политической географии как части герменевтической метатеории понимания и геополитики как области интертеории объяснения специфики пространственно распределенной деятельности с учетом или без учета особенностей геоисторической среды. Обсуждается роль и место математики в системе гуманитарных знаний и в решении задач статистического анализа данных и моделирования политических процессов и явлений. Обоснована дополнительность методологических, математических и эмпирических методов исследования. Предложены схемы структуры и организации касательных слоев видов деятельности и уравнения количественного анализа и моделирования политических явлений, применение которых продемонстрировано на примере интерпретации результатов голосования.

Ключевые слова: политические процессы и явления, математический и статистический анализ, организация научных знаний, модели электорального предпочтения

Цитирование: Черкашин А.К. Политическая география и геополитика: герменевтические исследования и математическое моделирование электоральных явлений / А.К. Черкашин // Информационные и математические технологии в науке и управлении, 2024. - № 2. - С. 5-18. – DOI: 10.25729/ESI.2024.34.2.001.

Введение. При реализации технологии математико-статистического анализа данных и знаний приходится учитывать тонкие различия предметных областей исследования отдельных наук, выделять эмпирический, теоретический, метатеоретический и математический уровни представления научной информации. В чем, например, отличие физики, геофизики и физической географии или качественного и количественного подходов к объяснению и пониманию сложных процессов и явлений в природе и обществе?

В частности, в науке существуют две генетически взаимосвязанные дисциплины – геополитика и политическая география, различающиеся объектом, предметом и масштабами исследования [1]. Вместе с тем, имеющиеся отличия недостаточно ясны и обоснованы, в связи с чем базовые положения этих дисциплин подвергаются критике и дополнительному анализу в аспекте существующего политического дискурса [2]. Это способствует критическому обновлению и теоретизации геополитики и политической географии, располагающих собственными системами понятий и методов. Так, в сферу интересов критической геополитики, помимо традиционного внимания к физическим границам, географическому положению, институтам государственной власти и внешним условиям, вошли социальные проблемы, вопросы глобализации, идентичности и суверенитета [3].

Подобный подход отражает общее критическое направление развития науки, нацеленное на формирование инновационного аналитического и рационального мышления, способности оценивать, выбирать и использовать различные источники информации, избегать предвзятости и ошибок интерпретации, обоснованно и непротиворечиво излагать собственное мнение. По этой причине критическое направление подразумевает, в частности, применение математических моделей и методов обработки информации, достоверного обоснования выводов. Такое применение само находится под критикой со стороны гуманитарных исследователей. При-

чинами этого являются незнание математики и нежелание ее осваивать, неумение пользоваться формулами, отсутствие опыта приложения формальных знаний, когда забываются полученные ранее навыки матанализа. Практика научных исследований показывает, что переход на математический уровень доказательного познания позволяет переосмыслить и теоретически обосновать полученные ранее эмпирические знания.

В последние годы геополитика и политическая география, включая электоральную географию, обновляются в части совершенствования аналитического аппарата с применением математических, теоретических, статистических методов, средств пространственно-временного полимасштабного анализа количественных данных наблюдений и опросов. Методы статистической обработки давно применялись в гуманитарных науках [4, 5]. Однако критический взгляд на полученные результаты требует их специальной качественной интерпретации в терминах феноменологии и герменевтики с применением инновационной математики [6] для объяснения выявленных закономерностей.

Основываясь на имеющемся опыте решения подобных задач [7-9], в статье излагается герменевтическая классификация знаний, определяется разграничение уровней знаний геополитики и политической географии, выделяется положение политологии в системе интертеории деятельности, и на примере материалов выборов продемонстрированы модели и методы количественного анализа электоральных предпочтений.

1. Математические основы и их интерпретации. Предлагается критически посмотреть на существующие в общественных науках модели и методы с позиций уравнений дифференциальной геометрии, возникающих при реализации процедур расслоения над многообразием связи явлений и знаний для герменевтического познания целого по частям (слоям).

Расслоением $s=(X,\pi,B)$ в общем случае называется отображение π пространства X на пространство B: $\pi : X \to B$. Пространство X называется пространством (множеством, объектом) расслоения, а B - базой расслоения из набора элементов $b_j \in B$. Обратное отображение $\pi^{-1}=\sigma$: $B \to X$ сечет пространство X , превращая его в расслоенное пространство $TB=\{TB_j\}$ независимых слоев TB_j (возможных миров системного существования). Для любого элемента $b_j \in B$ прообраз $TB_j=\sigma(b_j)$ называется слоем расслоения σ над элементом $b_j \in B$. Простым примером расслоения является сортировка (типизация, районирование) элементов множества, в частности, деление земной поверхности на территории и акватории в признанных границах государств, а также международные воды (открытое море) и земли, имеющие особый международно-правовой статус (Антарктида). Предполагается, что аналогичным образом все научное знание можно разбить на независимые по содержанию теории $T=\{T_j\}$ с базой $B=\{b_j\}$ ее основополагающих законов и аксиом [6].

В дифференциальной геометрии пространство $X=\{X_i\}$ – это многомерное пространство n независимых координат X_i ($i=1,2,3\ldots n$), в котором каждая точка координируется значением $x=\{x_i\}$, $x_i \in X_i$. В зависимости от содержания теории T_j координаты пространства X понимаются по-разному. В этом пространстве априори существует неизвестная функция $F(x)$ связи значений координат $x=\{x_i\} \in X$. Например, политолог Д. Истон [10, 11] предложил системную модель для анализа политических явлений в виде «черного ящика» типа «вход x – выход $F(x)$»[1] со скрытыми от наблюдателей внутренними процессами. На «входе» учитываются политические мотивации людей, определяемые внешними и внутренними по отношению к государству условиями, а на «выходе» – политические действия властей. Теория Д. Истона была модифицирована П. Тейлором и Р. Джонстоном [12] в исследованиях по электоральной географии.

[1] Здесь и далее формулы используются для математической интерпретации мнения других авторов.

Функция $F(x)$ для изучения ее свойств исследуется по частям, представляющим касательные плоскости к поверхности этой функции $F(x)$ в точке $x_0=\{x_{0i}\}\in X$, где $F(x_0)=F_0$ (рис.1). Функция $F(x)=F(x_0)$ задает базу расслоения, а множество касательных слоев $TM(x_0)$ к поверхности этой функции в точках x_0 соответствует расслоенному пространству, которое по частям точно описывает неизвестную функцию $F(x)$ в локальных координатах $y=\{y_i\}\in Y=\{Y_i\}$, $y_i=x_i-x_{0i}$ универсальной функцией $f(y)=F(x)-F(x_0)$ вида [9]:

$$f(y)=a\cdot y=\sum_i a_i y_i=\sum_i \frac{\partial f}{\partial y_i} y_i = Lf(y) \quad , \qquad (1)$$

где $y=\{y_i\}$ – набор (вектор) значений, смещённых $y_i = x_i - x_{0i}$ относительно координат точки касания $x_0=\{x_{0i}\}$ и $F(x)=F(x_0)$; $a=\{a_i\}$ – ковектор переменных коэффициентов чувствительности, определяющих направление, условный вес, цену или ценность относительных показателей $y = x–x_0$ изучаемого явления; $a\cdot y$ - векторное произведение двойственных векторов a и y; L - оператор дифференцирования (действия).

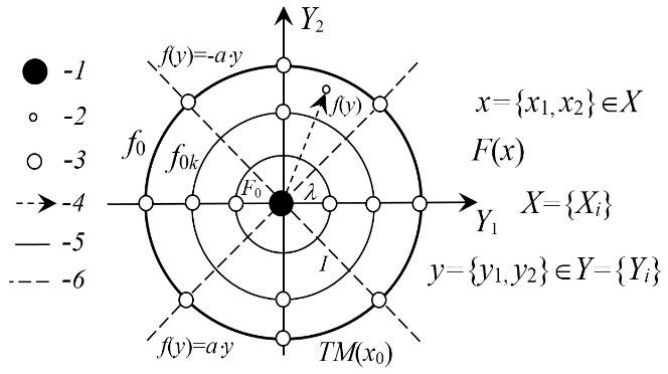


Рис. 1. Качественная концептуальная структура количественных связей в касательном слое. 1 – положение центра; 2 – положение произвольного состояния в плоскости слоя; 3 - выделенные дискретные состояния; 4 – радиус-вектор состояния системы; 5 – граница ядра слоя; 6 – радиусы - линии функциональных связей (пояснения в тексте).

Влияния окружения x на политическую систему $F(x)$ подразделяются на три вида: благоприятствующие $a_i>0$, нейтральные $a_i=0$ и стрессовые $a_i<0$ [11]; последние представляют для системы реальную угрозу, увеличивают риски функционирования.

Предложенные соотношения формируют герменевтическое разложение наблюдаемого явления (феномена)

$$F(x) = f(y) + F(x_0), \qquad (2)$$

на чистое знание $f(y)$ – функцию свободы, формируемую в слое, и функцию $F(x_0)$, определяющую особенность условий местоположения, а именно общезначимый смысл $f(y)$ связи переменных и смысловое значение $F(x_0)$ (основу понимания) феномена или информации о нем. По схожей формуле возникает «мировой порядок, основанный на правилах» в системе межгосударственных и общественных отношений. Он подразумевает индивидуальную трактовку $F(x_0)$ права $f(y)$. На нем базируется современная концепция западных держав, призванная описать желательный для них порядок мироустройства $F(x)$, выражающий интересы Запада. Обоснованное применение на практике правовых законов $f(y)$ заключается в подведении частных случаев жизни $F(x)$ под соответствующие законы $f(y)$, заданные в общем виде, с учетом фактических обстоятельств дела x_0 данного конкретного случая $F(x_0)$. Концептуальный аппарат герменевтики обеспечивает переход от понимания смысла универсальных законов $f(y)$ к правильному варианту его интерпретации $F(x)$ в конкретной средовой ситуации $F(x_0)$.

Критическая геополитика предлагает новый подход, в котором географическое пространство социально сконструировано, отражает идеи и представления различных геополитических акторов [13]. Для этого создается схема базовой структуры слоя, одинаковой для всех научных интерпретаций (рис. 1). Последнее качество обеспечивает аффинную симметрию (подобие) всех слоев, если совмещать их центры и поворачивать элементы структуры относительно центра и друг друга. В частности, эти свойства формально выражают идею равенства суверенных государств, их тождества как противоположностей, идею справедливости. Объяснить – означает отождествить поступающую информацию с универсальной структурой слоя, типовой моделью представления знаний. Понимание же должно быть достигнуто через открытие таких повторяющихся моделей и процессов, испытывающих воздействие географических различий [14].

На рис. 1 структура слоя представлена на плоскости, хотя возможен общий случай многомерной сферической модели представления знаний в прямоугольных или полярных координатах. В структуре выделяется центр $f(y)=f(0)=0$, когда $F(x)=F(x_0)=F_0$, ядро $f(y)\le f_0$ и периферия $f(y)>f_0$, неограниченно распространяющаяся в пространстве слоя. Функция $f(y)$ определяется значениями $y=\{y_i\}$ локальных координат $Y=\{Y_i\}$ и представлена вектором положения в слое или линиями $f(y)=a\cdot y$ (в общем случае гиперплоскостями) сечения слоя через центр. Положения о существовании нормы $F_0=F(x_0)$, норматива отклонения от нормы $f(y)=f_0$ и согласно (1) универсальной связи $f(y)= Lf(y)$ оценки локального состояния системы $f(y)$ с порождающим его действием $Lf(y)$ можно считать аксиомами, задающими структуру любого слоя.

Схематическая модель слоя напоминает азимутальную полярную проекцию Земли, где параллели нормальной сетки выглядят как концентрические окружности f_{0k}, а меридианы – их радиусы $f(y)=a\cdot y$, расходящиеся из общего центра F_0 параллелей под разными углами λ (азимутами): $a=\mathrm{tg}\lambda$. Конец радиус-вектора $f(y)$ в полярной системе координат вырисовывает сложную линию, например, контур антарктического континента. Радиусы $f(y)=a\cdot y$ задают границы секторов, а окружности – границы уровней k отклонения f_{0k} от центра и межуровневые кольца (пояса, оболочки, концентры). Пересечения линий окружностей и радиусов соответствуют выделенным положениям в структуре слоя, а пересечение секторов и колец – ячейкам локального пространства. Получается, что радиусы и окружности - это внутренние расслоения пространства локального слоя на дополняющих базах точек соответственно окружностей и радиусов. Выделенные точки сами становятся элементами базы расслоения внутри слоя, что генерирует иерархическую организацию системы в развитии.

Концентрические (нуклеарные) структуры широко распространены в природе и обществе [15]. Общий план прослеживается в строении Галактик и Солнечной системы, в планетарной модели атома, в эукариотической клетке с генетическим центром (ядром) и клеточной оболочкой, а также в моделях городов с окрестностями, столицы страны с территорией в пределах ее государственных границ. В иерархическом порядке классификации различаются: род и подчиненные виды данного рода, в иерархии расселения – областные города и районные центры. Выделяются пояса (оболочки) и секторы в диаграмме геополитического положения России, на которой по ячейкам распределены ее геополитические интересы и проблемы [1].

В структуре слоя центральное положение $F_0=F(x_0)$ определяет индивидуальные свойства и правила существования всего слоя, подобно тому, как в демократическом обществе народ является материальным источником власти и права, а базовые понятия и аксиомы – источником теоретического знания. Такие представления лежат в основе понимания суверенитета государства, признанного в конкретных границах f_0 и проистекающего из самодостаточности базовой позиции F_0, что включает своеобразие исторического, географического, дипломатического, военного, экономического и культурного суверенитетов [16].

Содержание ядра и периферии всегда рассматривается в направленных отклонениях $f(y)$ от кода центра слоев F_0. В частности, под геополитическими кодами понимаются устойчивые внешнеполитические ориентации государства, в пределах длительного исторического периода свободные от политической конъюнктуры F_0. Периферия страны проявляется в форме трансграничного, экстерриториального влияния по «рыночному» принципу – внешнее предложение (экспансия), положительный спрос (потребление) или активное сопротивление, определяющие характер взаимодействия стран и народов. Под критической геополитикой сначала понимали именно изучение внешней политики в виде анализа политического дискурса [17]. В классической геополитике превалирует государствоцентричный подход, где государство понимается как носитель суверенитета, «вместилища» общества в обозначенных границах [13]. Эквивалентность структурных схем жизненных пространств разных слоев $TM(x_0)$ позволяет поставить в центр F_0 типовой концентрической модели также мысли и деятельность свободных индивидов и независимых сообществ-акторов. Исключение из герменевтического исследования (2) центрального положения F_0 позволяет послойно сопоставлять функции $f(y)$ в чистом виде, т.е. проводить мета-анализ геополитических закономерностей. Это также дает возможность переходить по поверхности многообразия $F(x_0) \subset F(x)$ из слоя в слой $TM(x_0)$ с сохранением инвариантных отношений $f(y)= f(x_0, y)$, что придает политическим процессам циклический характер [18, 19] смены длительно устойчивых состояний (слоев) со сходными структурами и функциями отношений $f(y)$.

Множество центральных позиций $F(x_0)$ лежит на поверхности многообразия функции $F(x)$, касательные слои $TM(x_0)$ к которому в точке x_0 в совокупности задают расслоенное пространство $TM =\{TM(x_0)\}$. В каждом слое $TM(x_0)$ определяется вектор предпочтений $f(x_0, y)$ с началом в точке центра $F(x_0)$ (базового значения) и индивидуальной ориентацией; набор векторов $\boldsymbol{f}=\{f(x_0,y)\}$ разных x_0 организуется в векторное поле типа градиентов (уклонов) рельефной поверхности, указывающей, например, на силовые направления поверхностного стока. Аналогично в конкретных пространственных границах формируются векторы силовых полей, понимаемые как геополитические интересы и соответствующие градиентам изменения $f(x_0, y)$ геополитических потенциалов $F(x_0, x)$ отдельных стран x_0. Геополитическая структура мира отражает соотношение векторов $f(x_0, y)$ силовых полей геопространства $x_0 \in x \in X=\{X_i\}$, что складывается не столько из пространственно-временных параметров, сколько из географических факторов местоположения.

Электоральная география в рамках политической географии изучает территориальные различия в предпочтениях избирателей, географические закономерности результатов выборов и референдумов, воздействия на их итоги различных факторов [12, 20]. Эти итоги имеют большое влияние на государственное, региональное и местное управление и жизнь населения. В циклическом избирательном процессе граждане, как живут – так и голосуют, и как голосуют – так и живут, что наглядно подтверждается новейшей историей на всех территориальных уровнях, где прослеживаются и интерпретируются последствия наличия или отсутствия рационального выбора избирателей. Результаты голосования предоставляют массовый статистический материал для математической обработки.

В политической науке различают два основных подхода к анализу электорального поведения на основе учета: влияния социальных детерминант человека $x=\{x_i\}$ и индивидуального выбора избирателя $f(y)$, расчета в зависимости от своих ожиданий и сравнений предложений $y=\{y_i\}$, которые ему выдвигают кандидаты и партии. А.Зигфрид представил контекстуальную «экологическую» модель анализа голосования, основанную на постулате о том, что окружение $F_0=F(x_0)$, в котором находится избиратель, выступает как интегральный фактор, определяю-

щий его электоральную позицию, предпочтение и поведение. Им выявлена известная устойчивость политической ориентации избирателей в разных районах согласно их самоидентификации [21].

В электоральном отношении в структуре слоя (см. рис.1) центр F_0 соответствует устойчивому голосованию за лидера избирательного процесса, $f(y)$ – ориентация на других кандидатов, отражающая предпочтения избирателей относительно позиции лидера, $f(y)>f_0$ – малозначительное голосование за периферийных кандидатов (электоральный шум). Доля проголосовавших за лидера $F_0=F(x_0)$ изменяется по регионам и в совокупности соответствует функции многообразия связей внешних и внутренних факторов и условий $x_0=\{x_{0i}\}$, воздействующих на избирателей, – электоральному ландшафту. Функция отклонения $f(y)=F(x)-F(x_0)$ определяет степень политических конфликтов в сознании избирателей локальных сообществ, абстрагированную от средового контекста F_0 (модальности деятельности), что наглядно проявляется в протестном голосовании. Смена лидера знаменует революционный процесс, выраженный в изменении положения точки x_0 на многообразии $F(x_0)$ и в соответствующей смене ориентации $f(y)=F(x)–F(x_0)$.

2. Математический анализ и моделирование. Эффективность критического применения математических методов заключается не столько в формальном описании процессов и явлений, сколько в возможности использования всего арсенала математики для исследования полученных формул и уравнений. Так, соотношение (1) задает множество метрик (оценок) расстояния между произвольной точкой x и центральным положением x_0, что широко используется при изучении свойств финслеровых пространств в геометрии. Функции $f(y)$ – это различные количественные оценки, сравнение которых по ситуациям $F(x_0)$ позволяет выбирать наилучший вариант действия.

Соотношение (1) соответствует дифференциальному уравнению Эйлера, что имеет множество решений $f(y)$ – однородных функций первого порядка со свойством $f(ty)=tf(y)$, демонстрирующим зависимость от масштаба t и ориентации $t \to -t$. В связи с этим особой чертой критической геополитики является создание бинарной географии, демонстрирующей образы цивилизации и варварства, традиции и модерна, Запада и Востока, суши и моря, что должно было отразить и закрепить отношения господства и подчинения, роль доминирующих держав как лидеров мирового развития [13]. Вместе с тем, поворот-отрицание одной позиции $f(y)=a\cdot y \to f(-y)=-a\cdot y$ (см. рис.1) другой с разным знаком оценки факторов y не является антагонистическим, исключающим другую противоположность, и в идеале полученное в результате преобразования значение также имеет согласно (1) право на существование, как это происходит при голосовании, выражающем личную свободу волеизъявления или устойчивый нарратив t общественного мнения [9].

Д. Паркер [14] развивал представление о геополитическом процессе не просто как о взаимодействии $f(y)=F(x)–F(x_0)$ множества факторов $y=\{y_i\}$, $y_i=x_i-x_{0i}$, а как об эволюционном процессе, осложненном многофакторностью человеческой истории $F(x_0)$. В данном отношении имеет смысл под функцией $f(y)$ понимать скорость изменения переменных dy/dt, когда из соотношения (1) получается дифференциальное уравнение вида

$$dy/dt=ay(t), \text{ или } d(x-x_0)/dt=a(x-x_0). \qquad (3)$$

В качестве показателя $x(t)$, например, рассматривается переменная величина – размер площади страны $x(t)$ [22], изменяющейся в зависимости от сложившихся геоисторических условий $x_0(t)$. Статистический анализ методом скользящей регрессии временного ряда данных $x(t)$ позволяет оценить изменчивость коэффициента чувствительности a и характеристику среды $x_0(t)$ на момент времени t [22]. Расчеты указывают на явную средовую зависимость от

$x_0(t)$ территориального роста России и экспансивное расширение Британской империи, определенное ее текущим размером $x(t)$.

К этому же типу линейных моделей относится работающая модель Ричардсона [23] механизма регулирования величины уровней вооружений двух противоборствующих стран $x_1(t)$ и $x_2(t)$ в момент времени t, выраженной в относительных единицах $y_i(t)=x_i(t) - x_{0i}(t)$ и $dy_i/dt=d(x_i - x_{0i})/dt = dx_i/dt - dx_{0i}/dt$:

$$dy_1/dt=a_{11}y_1(t)+a_{21}y_2(t),\ dy_2/dt= a_{12}y_1(t)+a_{22}y_2(t),$$
$$a_{11}<0,\ a_{21}>0,\ a_{22}<0,\ a_{12}>0,\ a=\{a_{ik}\},\ i,k=1 \text{ или } 2. \quad (4)$$
$$dx_1/dt = a_{11}x_1(t)+a_{21}x_2(t)+\Phi_1(a),\ \Phi_1(a)= -[a_{11}x_{01}(t)+a_{21}x_{02}(t)-dx_{01}/dt],$$
$$dx_2/dt= a_{12}x_1(t)+a_{22}x_2(t) +\Phi_2(a),\ \Phi_2(a)= -[a_{12}x_{01}(t)+a_{22}x_{02}(t)-dx_{02}/dt],$$

где коэффициент a_{ik} соответствует влиянию величины вооружения $x_i(t)$ одной страны i на изменение вооружения $x_k(t)$ другой k; $\Phi_1(a)$ и $\Phi_2(a)$ – положительные или отрицательные величины в зависимости от того, насколько враждебно $\Phi_i(a)>0$ или дружественно $\Phi_i(a)<0$ настроены государства по отношению друг к другу. Эти величины складываются из коэффициентов взаимного влияния a_{ik} и равновесных величин $dy_i/dt=0$, $y_i(t)=0$, $x_i=x_{0i}$; равенство $x_i=x_{0i}$ характеризует степень напряженности отношения стран. Преимущество варианта (4) этой модели состоит в ее автономности, независимости от средового окружения, в данном случае интегрально учитываемой функцией $\Phi_i(a)$, значение которой необходимо минимизировать, а именно снижать уровень напряженности.

Дальнейшим развитием метода расслоения на многообразиях является переход от линейного расслоения к квадратичному, порождающему зависимости логистического типа:

$$dy/dt=ay(t)+by^2(t),\ dy/dt=ay(t)[1+(b/a)y(t)]=ay(t)[1-y(t)/y_m],\ y_m=-a/b,$$

где параметр a_{ik}характеризует относительную скорость роста переменной $y(t)$; y_m – поддерживающую ёмкость среды, т.е. максимально возможное относительное значение $y=x-x_0$; x_0 - начальное (фоновое) значение переменной x. Такая модель используется при описании роста организмов и численности популяций или развития эпидемии инфекционных заболеваний, а также применяется при описании электоральной активности избирателей по данным почасового хода голосования [24]. В данном случае $x(t)$ – число проголосовавших на момент t, x_0 – число досрочно проголосовавших, y_m – общее число проголосовавших избирателей в день голосования на избирательном участке или округе.

С помощью квадратичных моделей общего типа в химии и биологии учитывается парное взаимодействие и конкурентное отношение частей $x=\{x_i\}$ системы в виде произведения x_ix_k. В последнем случае речь идет об известной модели «хищник x_i – жертва x_k», в которой конкурентные отношения развиваются циклически. Аналогичный подход применяется в ланчестерских моделях боевых действий [25], где $x_i(t)$ определяет боевой потенциал i-ой стороны сражения, который подразделяется $x_i(t)=y_i(t)+x_{0i}(t)$ на активный боевой $y_i(t)$ и обеспечивающий (резерв) $x_{0i}(t)$ потенциалы, изменяющиеся во времени. Уравнения взаимодействия первоначально записываются в относительных единицах $y_i(t)$, величина которых раскрывается через боевые потенциалы $x_i(t)$ и $x_{0i}(t)$, что порождает слагаемое типа $\Phi(a,b)$ в соотношениях (4). Оно зависит от набора коэффициентов $a=\{a_{ki}\}$ небоевых потерь a_{ii} и воздействия a_{ki} противника k на противника i на переднем крае, от взаимодействия сторон конфликта $b=\{b_{ki}\}$ и наличных резервов $x_{0i}(t)$. Системная функция $F(x)$ в данном случае описывает ход сражения в терминах изменения боевого потенциала сторон конфликта, а $F(x_0)$ – скрытый потенциал (резерв), величиной которого в итоге теоретически и эмпирически определяется исход сражения. Таким же образом моделируется конкуренция кандидатов, что ярко проявляется в предвыборных дис-

куссиях с целью убедить избирателей в преимуществах своей программы, выступающей в качестве «резерва» предложений по каждой из поставленных проблем, которые кандидат (в случае избрания) предлагает реализовать на практике. В статьях [26, 27] рассматриваются нелинейные математические модели голосования за разные политические партии.

Еще одной формой обобщения теории расслоения является главное расслоение. Функции $f(y)$ типа (1) обладают групповыми свойствами евклидового линейного пространства в смысле возможности их сложения с получением также линейной функции, что важно, в частности, для свертки информации по иерархии показателей. Надстройка над каждым слоем $f(y)$ – суперпозиция функций $P[f(y)]$, обладающая свойствами элементов теории групп, – порождает главное расслоение, в частности, в случае экспоненциальных функций

$$G[f(y)]=G_0\exp[\alpha f(y)], \quad (5)$$

парное произведение которых также представляют собой экспоненциальную функцию. Примером служит модель Кокса [28] для прогнозирования риска наступления события с учетом влияния независимых переменных (предикторов) $x=\{x_i\}$, где P_0 – базовый риск при $f(y)=0$; α – специфическая константа. Логарифмированием уравнение (5) приводится к линейному виду,

$$\ln[G/G_0]=\alpha f(y), \quad (6)$$

что исследуется методами линейной регрессии.

3. Модель электоральных предпочтений. Для моделирования избирательного процесса и формального представления результатов голосования предлагаются различные уравнения, соответствующие формулам универсальной схемы описания структуры слоев (рис. 1). Здесь необходимо учитывать опыт изучения электоральных предпочтений избирателей, выраженных в итогах выборов с учетом факторов и особенностей регионов [29]. Распространены модели механизмов конкуренции кандидатов и партий и электорального поведения населения [26, 27]. Хорошо согласуется с данными экспоненциальная ординалистская модель вида (5) рангового распределения партий по числу поданных за них голосов избирателей [30], что подчеркивает универсальность так называемой «neutral» модели видовой конкуренции [31]. Обнаружена [30] достоверная логлинейная зависимость типа (6) между долей поданных голосов $G[f(y)]=P(x)$ и рангом x партии при $f(y)=y=x-x_0$, где x – ранг распределения голосов, начиная с $x=x_0$. Линейную форму связи (3) можно также сопоставить с универсальным соотношением (1) при $f(y)=a\cdot y=\ln[P/P_0]=\alpha y=\alpha(x-x_0)$, $\alpha=a$. Значение $P(x_0)=P_0$ соответствует доле проголосовавших за лидирующего кандидата или партию, $\rho(x)=P(x)/P_0$ – индекс предпочтения. Функция $F(x)$ равна $F(x)=\ln P(x)$, а базовая величина $F(x_0)=\ln P_0$ соответствует центральному положению в слое, координаты $(F(x_0), x_0)$ которого меняется по регионам, отражая многообразие результатов победившего кандидата. Это положение становятся ориентиром (началом вектора) $f(y)=\ln[P(x)/P_0]$ результатов голосования за других кандидатов $y=x-x_0$ (с другим рейтингом x) (см. рис. 1). Структура электорального пространства характеризуется направлением α и величиной $\rho(x)$ вектора предпочтения. Граница ядра $f_0=\ln[P_m/P_0]$ определяется долей голосов $P(x_m)=P_m$, поданных за кандидатов-аутсайдеров избирательного процесса, при преодолении которой $f(y)>f_0$ в зоне периферии нарушаются линейные закономерности $f(y)=\ln[P(x)/P_0]=\alpha y=\alpha(x-x_0)$ рангового распределения.

Для количественного анализа использовались итоги выборов Президента Российской Федерации 18 марта 2018 г. по районам Иркутской области, регионам страны и страны в целом[2]. Явка избирателей в Иркутской области составила 55,67%. По результатам голосования в области, как и по всей стране, лидировал В.В.Путин с 73,06% голосов избирателей, что в

[2] ЦИК России. Выборы Президента Российской Федерации. Результаты выборов. - www.irkutsk.vybory.izbirkom.ru

долях составляет $P(x_0)=P_0=0{,}7306$ ($x=x_0=1$ – первое место). Далее следуют результаты поддержки П.Н.Грудинина (15,93%) и В.В.Жириновского (6,44%). На периферии избирательного процесса ($f(y)\geq 0{,}005$) оказались С.Н.Бабурин (0,52%), Б.Ю.Титов (0,51%) и М.А.Сурайкин (0,5%).

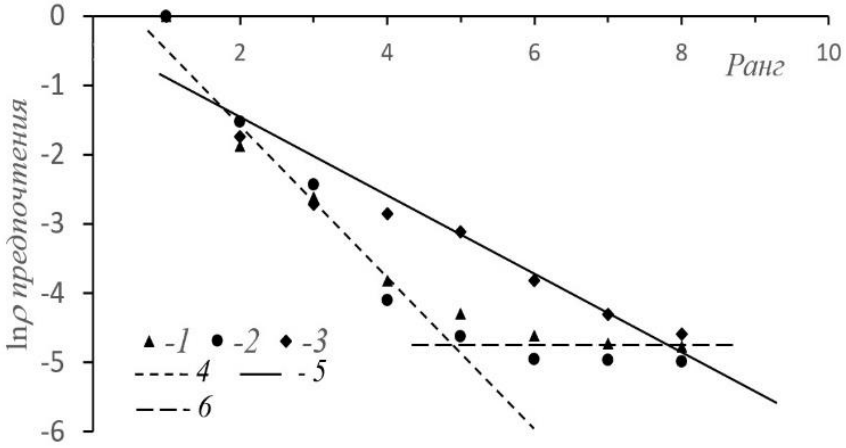


Рис. 2. Линейная зависимость $\ln\rho(x)=\alpha x+\beta$ показателя предпочтения $\rho(x)$ от ранга x кандидата по результатам выборов

1 – данные по Иркутской области; 2 – по Российской Федерации в целом; 3 – по выборам в Москве; 4 – тенденция зависимости $\ln\rho(x)$ для России и Иркутской области; 5 – тенденция зависимости $\ln\rho(x)$ для Москвы; 6 – линия разграничения ядра и периферии результатов голосования.

Данные ранжированы по $x=1, 2, 3\ldots 8$ в порядке уменьшения доли поданных голосов $P(x)$, начиная с P_0. Строится линейная зависимость $\ln\rho(x)=\alpha x+\beta=\alpha(x-x_0)$, $x_0=-\beta/\alpha$ (рис. 2). Для Иркутской области для 5 первых рангов $\ln\rho(x)=-1{,}18x-1{,}01=-1{,}18(x-0{,}86)$ (коэффициент корреляции $R=-0{,}91$), где $\alpha=-1{,}18$, $x_0=0{,}86$. Границы периферийных значений соответствуют $x_m=5$, $P_m\approx 0{,}0051$, $f_0=-5{,}0$. Для России в целом имеются сходные показатели $P_0=0{,}7669$, $R=-0{,}98$, $\alpha=-1{,}05$, $x_0=0{,}64$, $x_m=5$, $P_m\approx 0{,}0091$, $f_0=-4{,}7$. В Москве $P_0=0{,}7087$, $R=-0{,}95$, $\alpha=-0{,}58$, $x_0=-0{,}50$, $x_m=8$, $P_m\approx 0{,}0072$, $f_0=-4{,}6$. В последнем случае все 8 ранговых позиции укладываются в зависимость $\ln\rho(x)=\alpha(x-x_0)$. Расчетная величина x_0 обычно не равна теоретическому значению $x_0=1$, как это получается по результатам голосования в Республике Дагестан, где $P_0=0{,}9076$, $R=-0{,}99$, $\alpha=-2{,}74$, $x_0=1{,}02$, $x_m=3$, $P_m\approx 0{,}002$, $f_0=-6{,}2$.

Сдвиг значения x_0 от единицы в сторону уменьшения соответствует снижению доли проголосовавших за победителя на выборах. По модельной схеме слоя (см. рис.1) сравнение доли голосов P_0, поданных за победившего кандидата, индексирует геоисторические условия существования регионов и страны в целом. Высокие показатели P_0, демонстрирующие политическую лояльность, естественно снижают варьирование результатов голосования за других кандидатов – проявленную степень разнообразия выбора. В России зависимость граничных значений P_m от базовых P_0 заметно варьирует, но обнаруживает главную тенденцию $P_m=-0{,}017P_0+1{,}89=-0{,}017(P_0-1{,}11)$ ($R=-0{,}78$) понижения пороговых значений P_m (нормативов) от нормы голосования P_0 по регионам с тотальным итогом $P_0\approx 1$, когда $P_m\approx 0$. По Иркутской области в сравнении «Бабурин-Путин» эта тенденция описывается похожим равенством $P_m=-0{,}011(P_0-1{,}19)$ ($R=-0{,}80$). На внутрирегиональном уровне зависимость индекса предпочтений $\rho(x)$ от ранга x представлена пучком линий $\ln\rho(x)=\alpha(x-x_0)$ с центом $\rho(x_0)=1$, $x_0=1$ с разными коэффициентами наклона α линий, что подчеркивает универсальность подобной связи для $\rho(x)=\ln[P(x)/P_0]$, исключающей по определению средовую геоисторическую обусловленность

P_0, отраженную в электоральной самоидентификации избирателей на всем многообразии земного пространства политической деятельности.

Заключение и выводы. На примере критически обсуждаемых с математических позиций политических знаний появляется возможность сделать следующие науковедческие обобщения, отражающие классификационные уровни дифференциации знаний с использованием процедур расслоения на многообразиях связи понятий разного рода – качественного и количественного содержания, теоретического (системного) и метатеоретического (герменевтического, метасистемного) обобщения при исследовании явлений по частям (слоям) и в целом через сопоставление слоев.

В системах деятельности политика сосуществует с другими независимыми формами (слоями) культуры: наука, искусство, управление и религия. Герменевтический процесс обеспечивает гармонию (конценсус) частей общественного целого, их тождество как противоположностей через объяснение и понимание в прямом или опосредованном текстами диалогекоммуникации между людьми, сообществами, институтами общества и власти, между суверенными государствами. Политика – вид деятельности, направленной на изменение системы общественных отношений, в том числе отношений к тому или иному политическому явлению, что непосредственно сказывается на результатах народного голосования, когда решается вопрос о власти и обсуждаются другие проблемы общественного развития.

Особо выделяется научная герменевтика, в задачу которой входит специальное исследование действительности с выделением в поступающей информации общезначимого смысла и смыслового значения для ее объяснения и понимания. Политические процессы и явления попадают в сферу интересов науки и в процессе герменевтического анализа преобразуются в научные знания через расслоенное познание целого по частям. Иерархическая классификация научных знаний начинается с научной герменевтики. География является герменевтической наукой и относится к эмпирическому исследованию действительности с учетом состояния окружающей среды с параллельным использованием моделей и методов философской и математической герменевтики. Политическая географическая герменевтика – это метатеория политической географии, что призвана обеспечить толкование явлений по месту и времени в общем контексте состояния мировой политики.

Политология изучает политические явления и процессы в рамках сквозной теории (интертеории) деятельности, исследующей в том числе межличностные, общественные и межгосударственные отношения, универсальные законы морали и нравственности, политической экономии и социального развития в чистом смысловом виде безотносительно к модальности (условиям среды) конкретной деятельности. Геополитика – это пространственная политология, частный раздел интертеории деятельности. Принципиальное отличие теорий от метатеорий заключается в том, что первые основаны на объяснении наблюдаемых событий, вторые – на их понимании посредством учета обстоятельств геоисторической среды, развернутой на многообразии системных связей влияющих факторов.

Политическая система – особый вид социальной системы деятельности, связанной с регулированием общественных отношений, целенаправленным формированием политической деятельности, выражающейся в разработке, принятии и реализации решений. Политические вопросы обсуждаются не только в рамках интертеории деятельности, но и иных теорий – механизмов поведения, динамики, ранжирования (ординации), функционирования или надежности разнородных систем. Каждая из них имеет свое математическое выражение, согласованное с универсальными операциями дифференциальной геометрии – процедурами расслоения на многообразии, что формализуется как функциональная поверхность социально-экономической и природной среды. Каждая автономная политическая система (слой) касается этой среды в определенной точке, идентифицирующей уникальность этой системы.

Схема построения интертеорий соответствует аксиомам структуры слоя (центр, ядро, периферия влияния). Соответствующие законы действуют на всех иерархических системных уровнях, т.е. сквозным образом. В идеале, это одни и те же законы для разных стран и народов в смысле наличия их общих оснований. Одна из основных категорий геополитики – трансграничная экспансия (действие), являющаяся производной от государственных интересов (центра). Унифицированная фундаментальная геополитика предполагает суждения, свободные от обстоятельств проявления общих законов общественных отношений и деятельности. Новая геополитика изучает внутреннее и внешнее пространственное взаимодействие между центрами, ядрами и периферией различных политических образований.

Подобные закономерности наглядно прослеживаются при статистическом анализе результатов выборов в терминах ординалистской интертеории. Центральную позицию занимает победивший кандидат, представляющий интересы большинства избирателей, программа которого, оценка его личности и предшествующей деятельности идентифицирует геоисторическую ситуации в регионе. Ранжированная последовательность изменения функции электорального предпочтения подчиняется экспоненциальной зависимости, нарушение которой указывает на границу ядра функциональной однородности данных голосования, за которой представлены результаты выборов по «периферийным» кандидатам. Итоги голосования в различных регионах за победившего кандидата отражают многообразие средовых условий, которые необходимо принять за основу анализа и прогнозирования ситуации средствами политической географии с использованием моделей и методов геополитики в показателях относительного отклонения параметров состояния системы от состояния среды.

Благодарности. Исследование выполнено за счет средств государственного задания (№ госрегистрации темы: АААА-А21-121012190056-4).

Список источников

1. Колосов В.А. Геополитика и политическая география / В.А. Колосов, Н.С. Мироненко, 2001. – М.: Аспект Пресс. – 479 с.
2. Колосов В.А. Критическая геополитика: основы концепции и опыт ее применения в России / В.А. Колосов // Политическая наука, 2011. – № 4. – С. 31-52.
3. Окунев И.Ю. Критическая геополитика и посткритический сдвиг в исследовательской парадигме геополитики / И.Ю. Окунев // Сравнительная политика, 2014. – № 4 (16-17). – С. 6-14.
4. Taylor A., Tilton L. Humanities data in R: Exploring networks, geospatial data, images, and text. Springer, 2017, 211 p.
5. Humble S. Quantitative analysis of questionnaires: Techniques to explore structures and relationships. Routledge, 2020, 216 p.
6. Черкашин А.К. Инновационная математика: поиск оснований и ограничений моделирования реальности / А.К. Черкашин // Информационные и математические технологии в науке и управлении, 2019. – № 2 (14). – С. 69-87.
7. Черкашин А.К. Политическая география выборов в Иркутской области / А.К. Черкашин // География и природные ресурсы, 1995. – №2. – С. 130-140.
8. Черкашин А.К. Метатеоретические модели политической науки об устойчивом развитии в концепции расслоенных пространств деятельности / А.К. Черкашин // Известия Иркутского государственного университета. Серия Политология. Религиоведение, 2018. – Т. 25. – С. 5-23.
9. Черкашин А.К. Математическое объяснение устойчивого политического развития общества / А.К. Черкашин // Информационные и математические технологии в науке и управлении, 2021. –№ 2 (22). – С.5-20.
10. Easton D.A. Systems analysis of political life. New York, John Wiley & Sons, 1965, 507 p.
11. Истон Д. Категории системного анализа политики / Д. Истон // Антология мировой политической мысли. Т. II. Зарубежная политическая мысль XX в. / Под ред. Г.Ю.Семигина и др. – М.: Мысль, 1997. – С. 630-642.
12. Taylor P.J., Johnston R. Geography of elections. London. Routledge, 2014, 530 p.
13. О'Тоал Дж. Геополитика постмодерна? Геополитические представления модерна и за их пределами / Дж. О'Тоал // Политическая наука, 2009. – № 1. – С.188-223.

14. Parker G. Continuity and change in Western geopolitical thought during the twentieth century. International social science journal, 1991, vol. XLIII, no. 1, pp. 21-33.
15. Ретеюм А.Ю. Земные миры / А.Ю. Ретеюм. – М.: Мысль, 1988. – 266 с.
16. Черных В.В. Государственный суверенитет в истории России: цель и значение / В.В. Черных // Известия Иркутского государственного университета. Серия Политология. Религиоведение, 2023. – Т. 43. – С. 7-18.
17. Воскресенский А.Д. Мировое комплексное регионоведение / А.Д. Воскресенский, Е.В. Колдунова, С.И. Лунев, А.А. Киреева. – М.: Магистр, 2020. – 416 с.
18. Голдстоун Дж.А. Волны революций XXI столетия / Дж.А. Голдстоун, Л.Е. Гринин, А.В. Коротаев // Полис. Политические исследования, 2022. – № 4. – С. 108-119. – DOI:10.17976/jpps/2022.04.09. – EDN:DVNOBB.
19. Худяков А.В. Формирование теоретико-методологических основ циклических концепций в политической сфере: первая половина XX в. / А.В. Худяков // Известия Иркутского государственного университета. Серия Политология. Религиоведение, 2022. – Т. 42. – С. 20-33.
20. Ильин В.А. Электоральная география партийных предпочтений по итогам выборов в Государственную Думу РФ (2007, 2011, 2016 гг.) / В.А. Ильин, В.С. Каминский, Е.Э. Леонидова и др. – Вологда: ФГБУН ВолНЦ РАН, 2018. – 105 с.
21. Желтов В.В. К вопросу об анализе электорального поведения / В.В.Желтов, М.В.Желтов // Территория новых возможностей. Вестник Владивостокского государственного университета экономики и сервиса, 2011. – № 3 (12). – С. 48-63.
22. Черкашин А.К. Геоисторическая точность и исследование русского освоения земель в XVI–XXI вв. / А.К. Черкашин // Известия Иркутского государственного университета. Серия Политология. Религиоведение, 2014. – Т. 9. – С. 175-188.
23. Плотинский Ю.М. Модели социальных процессов / Ю.М. Плотинский. – М.: Логос, 2001. – 296 с.
24. Арбатская М.Н. Организация избирательных процедур как область политико-географического анализа / М.Н. Арбатская // Известия русского географического общества, 2007. – Т. 139. – № 2. – С. 70-77.
25. Lanchester F.W. Aircraft in warfare: the dawn of the fourth arm. London, Lanchester Pr Inc, 1995, 243 p.
26. Bañuelos S., Danet T., Flores C., Ramos A. An epidemiological math model approach to a political system with three parties. CODEE Journal, 2019, vol. 12, pp. 2-13, available at: https://scholarship.claremont.edu/codee/vol12/iss1/8
27. Misra A.K. A simple mathematical model for the spread of two political parties. Nonlinear analysis: modelling and control, 2012, vol.17, no. 3, pp. 343-354.
28. Therneau T.M., Grambsch P.M. Modeling survival data: extending the cox model. New York, Springer Science + Business Media, 2000, 287 p.
29. Евстифеев Р.В. Мелодии электоральных пространств. Политический процесс и электоральные предпочтения избирателей / Р.В. Евстифеев. – Владимир, 2009. – 192 с.
30. Hösel V., Müller J., Tellier A. Universality of neutral models: decision process in politics. Palgrave communications, 2019, no. 2, pp. 1-8.
31. Humble S. Quantitative analysis of questionnaires: techniques to explore structures and relationships. Routledge, 2020, 216 p.

Черкашин Александр Константинович. *Доктор географических наук, профессор, главный научный сотрудник лаборатории теоретической географии Института географии им. В.Б. Сочавы СО РАН, ORCID: 0000-0002-7596-7780, WOS: K-2418-2017, AuthorID: 58425, akcherk@irnok.net, 664033, г. Иркутск, ул. Улан-Баторская 1.*

UDC 51-77: 911.3:327:324
DOI:10.25729/ESI.2024.34.2.001

Political geography and geopolitics: hermeneutical research and mathematical modeling of electoral phenomena

Aleksander K. Cherkashin

V.B. Sochava Institute of Geography SB RAS,
Russia, Irkutsk, *akcherk@irnok.net*

Abstract. The difference between political geography as part of the hermeneutical meta-theory of understanding and geopolitics as an intertheory of explaining the specifics of spatially distributed activity, activities with or without taking into account the feature of the geohistorical environment, is proved on the basis of a critical mathematical approach with the procedures of knowledge stratification (fibration). The role and place of mathematics in the system of humanities and in solving problems of statistical data analysis and modeling of political processes and phenomena are discussed. The complementarity of methodological, mathematical and empirical research methods is substantiated. Schemes of the structure and organization of tangent layers (fibers) of activities and equations of quantitative analysis and modeling of political phenomena are proposed. Their application is demonstrated by the example of the interpretation of the popular election results.

Keywords: political processes and phenomena, mathematical and statistical analysis, organization of scientific knowledge, models of electoral preference

References

1. Kolosov V.A., Mironenko N.S. Geopolitika i politicheskaya geografiya [Geopolitics and political geography]. Moscow, 2001, Aspect Press Publ., 479 p.
2. Kolosov V. A. Kriticheskaya geopolitika: osnovy koncepcii i opyt ee primeneniya v Rossii [Critical Geopolitics: fundamentals of the concept and experience of its application in Russia]. Politicheskaya nauka [Political science], 2011, no. 4, pp. 31-52.
3. Okunev I.Yu. Kriticheskaya geopolitika i postkriticheskij sdvig v issledovatelskoj paradigme geopolitiki [Critical geopolitics and post-critical shift in the research paradigm of geopolitics]. Sravnitelnaya politika [Comparative politics], 2014, no. 4 (16-17), pp. 6-14.
4. Taylor A., Tilton L. Humanities Data in R: exploring networks, geospatial data, images, and text. Springer, 2017, 211 p.
5. Humble S. Quantitative analysis of questionnaires: techniques to explore structures and relationships. Routledge, 2020, 216 p.
6. Cherkashin A.K. Innovacionnaya matematika: poisk osnovanij i ogranichenij modelirovaniya real`nosti [Innovative mathematics: the search for the foundations and limitations of reality modeling]. Informacionny`e i matematicheskie texnologii v nauke i upravlenii [Information and mathematical technologies in science and management], 2019, no. 2 (14), pp. 69-87.
7. Cherkashin A.K. Politicheskaya geografiya vyborov v Irkutskoj oblasti [Political geography of elections in the Irkutsk region]. Geografiya i prirodnye resursy [Geography and natural resources], 1995, no.2, pp. 130-140.
8. Cherkashin A.K. Metateoreticheskie modeli politicheskoj nauki ob ustojchivom razvitii v koncepcii rassloennyx prostranstv deyatelnosti [Metatheoretical models of political science on sustainable development in the concept of stratified spaces of activity]. Izvestiya Irkutskogo gosudarstvennogo universiteta. Seriya Politologiya. Religiovedenie [The Bulletin of Irkutsk State university. Series political science and religion studies], 2018, vol. 25, pp. 5-23.
9. Cherkashin A.K. Matematicheskoe obyasnenie ustojchivogo politicheskogo razvitiya obshhestva [Mathematical explanation of sustainable political development of society]. Informacionnye i matematicheskie texnologii v nauke i upravlenii [Information and mathematical technologies in science and management], 2021, no. 2 (22), pp. 5-20.
10. Easton D.A. Systems analysis of political life, New York, John Wiley & Sons, 1965, 507 p.
11. Easton D.A. Kategorii sistemnogo analiza politiki [Categories of policy system analysis]. Antologiya mirovoj politicheskoj mysli. T. II. Zarubezhnaya politicheskaya mysl XX v. [Anthology of world political thought. Vol. II. Foreign political thought. XX century]. Ed. G.Yu.Semigina et al., Moscow, Mysl publ., 1997, pp. 630-642.
12. Taylor P.J., Johnston R. Geography of elections. London, Routledge, 2014, 530 p.
13. O'Tuathail G. Geopolitika postmoderna? Geopoliticheskie predstavleniya moderna i za ix predelami [Postmodern geopolitics? The modern geopolitical imagination and beyond. Politicheskaya nauka [Political science], 2009, no. 1, pp.188-223.
14. Parker G. Continuity and change in Western geopolitical thought during the twentieth century. International social science journal, 1991, vol. XLIII, no. 1, pp. 21-33.
15. Reteyum A.Yu. Zemnye miry [Terrestrial worlds]. Moscow, Mysl publ., 1988, 266 p.
16. Chernyh V.V. Gosudarstvennyj suverenitet v istorii Rossii: cel i znachenie [State sovereignty in the History of Russia: purpose and meaning]. Izvestiya Irkutskogo gosudarstvennogo universiteta. Seriya Politologiya. Religiovedenie [The Bulletin of Irkutsk State university. Series Political science. Religion studies], 2023, vol. 43, pp. 7-18.
17. Voskresenskij A.D., Koldunova E.V., Lunev S.I., Kireeva A.A. Mirovoe kompleksnoe regionovedenie [World complex Regional Studies]. Moscow, Magistr Publ., 2020, 416 p.

18. Goldstone J.A., Grinin L.E., Korotaev A.V. Volny revolyucij XXI stoletiya [Waves of revolutions in the 21st century]. Polis. Politicheskie issledovaniya [Policy. Political studies], 2022, no. 4, pp. 108-119, DOI:10.17976/jpps/2022.04.09, EDN: DVNOBB.
19. Hudyakov A. V. Formirovanie teoretiko-metodologicheskix osnov ciklicheskix koncepcij v politicheskoj sfere: pervaya polovina XX v. [Formation of theoretical and methodological foundations of cyclic concepts in the political sphere: the first half of the XX century]. Izvestiya Irkutskogo gosudarstvennogo universiteta. Seriya Politologiya. Religiovedenie [The Bulletin of Irkutsk State university. Series political science and religion studies], 2022, vol. 42, pp. 20-33.
20. Il'in V.A., Kaminskii V.S., Leonidova E.E., Morev M.V., Urvanova T.V., Ukhanova Yu.V. Elektoralnaya geografiya partijnyh predpochtenij po itogam vyborov v Gosudarstvennuyu Dumu RF (2007, 2011, 2016 gg.) [Electoral geography of party preferences based on the results of the elections to the State Duma of the Russian Federation (2007, 2011, 2016)]. Vologda, FGBUN VolNCz RAN Publ., 105 p.
21. Zheltov V.V., Zheltov M.V. K voprosu ob analize elektoralnogo povedeniya. Territoriya novyx vozmozhnostej [On the issue of the analysis of electoral behavior]. Vestnik Vladivostokskogo gosudarstvennogo universiteta ekonomiki i servisa [Bulletin of the Vladivostok state university of economics and service], 2011, no. 3 (12), pp. 48-63.
22. Cherkashin A. K. Geoistoricheskaya tochnosti issledovanie russkogo osvoeniya zemelv XVI–XXI vv. [Geohistorical accuracy and research of Russian land development in the XVI–XXI centuries]. Izvestiya Irkutskogo gosudarstvennogo universiteta. Seriya Politologiya. Religiovedenie [The Bulletin of Irkutsk State University. Series Political science and Religion studies], 2014, vol. 9, pp. 175–188.
23. Plotinskij Yu.M. Modeli social'nyh processov [Models of social processes]. Moscow, Logos Publ., 2001, 296 p.
24. Arbatskaia M.N. Organizaciya izbiratelnyx procedur kak oblast politiko-geograficheskogo analiza [Organization of electoral procedures as a field of political and geographical analysis]. Izvestiya russkogo geograficheskogo obshhestva [News of the Russian Geographical society], 2007, vol. 139, no. 2, pp. 70-77.
25. Lanchester F.W. Aircraft in warfare: the dawn of the fourth arm. London, Lanchester Pr Inc, 1995, 243 p.
26. Bañuelos S., Danet T., Flores C., Ramos A. An epidemiological math model approach to a political system with three parties. CODEE Journal, 2019, vol. 12, pp. 2-13, available at: https://scholarship.claremont.edu/codee/vol12/iss1/8.
27. Misra A.K. A simple mathematical model for the spread of two political parties. Nonlinear analysis: modelling and control, 2012, vol.17, no. 3, pp. 343–354.
28. Therneau T.M., Grambsch P. M. Modeling survival data: Extending the Cox model. New York, Springer Science + Business Media, 2000, 287 p.
29. Evstifeev R.V. Melodii elektoralnyh prostranstv. Politicheskij process i elektoralnye predpochteniya izbiratelej [Melodies of electoral spaces. Political process and electoral preferences of voters]. Vladimir, 2009, 192 p.
30. Hösel1 V., Hösel1 V., Müller J., Tellier A. Universality of neutral models: decision process in politics. Palgrave communications, 2019, no. 2, pp. 1-8.
31. Humble S. Quantitative analysis of questionnaires: Techniques to explore structures and relationships, 2020, 216 p.

Aleksander Konstantinovich Cherkashin. *Doctor of geographical sciences, professor, chief researcher of the laboratory of Theoretical geography of V.B. Sochava Institute of Geography SB RAS, ORCID: 0000-0002-7596-7780, WOS: K-2418-2017, AuthorID: 58425, akcherk@irnok.net, 664033, Irkutsk, Ulaanbaatar st. 1.*

Статья поступила в редакцию 08.12.2023; одобрена после рецензирования 17.01.2024; принята к публикации 30.05.2024.

The article was submitted 12/08/2023; approved after reviewing 01/17/2024; accepted for publication 05/30/2024.

УДК 004.032.26
DOI:10.25729/ESI.2024.34.2.002

Быстрые преобразования и самоподобные нейронные сети глубокого обучения.
Часть 3. Пирамидальные нейронные сети с глубокой степенью обучения

Дорогов Александр Юрьевич

ПАО «Информационные телекоммуникационные технологии» («Интелтех»),
Санкт-Петербургский государственный электротехнический университет,
Россия, Санкт-Петербург, *vaksa2006@yandex.ru*

Аннотация. В работе рассматривается класс быстрых нейронных сетей с пирамидальной структурой. Приведены методы топологического построения одномерных и двумерных пирамидальных сетей. Сети рассматриваемого класса представимы линейными операторами, имеют самоподобную структуру и являются частным случаем алгоритма быстрого преобразования Фурье. Предложены топологические модели пирамидальных нейронных сетей прямой и обратной ориентации. В работе показано применение пирамидальных нейронных сетей быстрого обучения для реализации корреляционной цифровой обработки сигналов и изображений, комбинационной логики и элементов памяти. Рассмотрены примеры построения шифратора и дешифратора бинарных кодов. Отмечено, что пирамидальная сеть памяти обеспечивает хранение и точное восстановление образов подобно хранению данных в компьютерной памяти произвольного доступа. Доказано, что быстрая пирамидальная сеть является нейронной сетью глубокого обучения, а самоподобная структура позволяет дообучать сеть к новым данным без необходимости полного переобучения сети. Данная работа является третьей частью обобщающей статьи «Быстрые преобразования и самоподобные нейронные сети глубокого обучения», опубликованной в этом журнале. В первой части рассмотрены стратифицированные модели самоподобных нейронных сетей, во второй части рассмотрены алгоритмы обучения быстрых нейронных сетей и обобщённых спектральных преобразований.

Ключевые слова: быстрое перестраиваемое преобразование; нейронная сеть; пирамидальная структура; глубокое обучение, дешифратор кодов, шифратор кодов, нейросетевая память, пластичность нейронной сети, степени свободы

Цитирование: Дорогов А.Ю. Быстрые преобразования и самоподобные нейронные сети глубокого обучения. Часть 3. Пирамидальные нейронные сети с глубокой степенью обучения / А.Ю. Дорогов // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 19-32 – DOI:10.25729/ESI.2024.34.2.002.

Введение. Как было отмечено в предыдущих частях статьи [1, 2], быстрые нейронные сети (БНС) имеют глубокую связь с алгоритмом быстрого преобразования Фурье и обобщённых быстрых преобразований. Методы обучения БНС, рассмотренные в Части 2 статьи, отличаются высокой скоростью и гарантируемой сходимостью. Тем не менее, было показано, что информационная ёмкость БНС используется не полностью, остаются степени свободы, которые не могут быть непосредственно использованы для обучения. В данной работе будет предложено решение, устраняющее этот недостаток. Ключевая идея заключается в изменении топологии БНС таким образом, чтобы выделить сетевые компоненты, информационная ёмкость которых используется полностью, и отсечь сетевые компоненты, информационная ёмкость которых не используется при обучении. В результате нейронная сеть упрощается, быстродействие её увеличивается, и глубина обучения становится максимально-возможной. Оказалось, что результат может быть достигнут за счёт изменения структурных характеристик сети, при сохранении топологических схем построения БНС. Последовательная реализация данной идеи приводит к построению особого типа нейронных сетей, размерность слоёв в которых изменяется подобно размеру пирамиды при движении от основания к вершине, отсюда появилось название − пирамидальные сети.

Название не новое, интерес исследователей к пирамидальным сетям возник в конце нулевых годов текущего столетия и был мотивирован работами по представлению

изображений в виде иерархических пирамид в задачах декомпозиции, сегментации и сжатия изображений [3]. В работе [4] была предложена архитектура пирамидальной нейронной сети PyraNet, предназначенной для визуального распознавания образов. Пирамидальная сеть PyraNet имеет иерархическую многослойную структуру, обрабатывающую 2D образы. Выходами сети является 1D слой, представляющий категории входных образов. Архитектура пирамиды сети подбирается экспериментально, каких-либо методов целесообразного выбора не предложено. Пирамидальная нейронная сеть данного типа для своего обучения использует тот же арсенал методов обратного распространения ошибок, что и классические многослойные нейронные сети. Общий недостаток этих методов − не гарантируемая сходимость по ошибкам обучения к глобальному минимуму и значительные вычислительные затраты, особенно для глубоких нейронных сетей [5]. Суждение о полезности применения сетей данного типа основано на ряде успешно решенных задач. Количественно информационная емкость пирамидальных нейронных сетей никогда не оценивалась.

В данной работе будут рассмотрены пирамидальные сети с регулярной самоподобной структурой. Как было показано в Части 2, эти сети обладают быстрыми алгоритмами обучения с гарантированной сходимостью к целевой функции. Архитектура пирамиды определяется фиксированным правилом, которое не изменяется и подобрано так, чтобы обеспечить максимальное использование информационной ёмкости сети. Рассмотрены сети с прямой и обратной ориентацией пирамиды. Будет показано, что сети пирамидального типа могут быть использованы для реализации многоканальных корреляторов, сложной комбинационной логики, для хранения и точного восстановления образов. Для самоподобных нейронных сетей существуют количественные оценки информационной емкости сети в зависимости от ее структуры. Будет показано, что предложенный вариант пирамидальных нейронных сетей полностью использует свой потенциал обучения.

1. Построение пирамидальной нейронной сети. В работе [6] (см. также Часть 1) показано, что модель алгоритма быстрого преобразования в топологии «Гуда» определяется кортежами:

$$\begin{aligned} U^m &= \langle v_{m-1} v_{m-2} \ldots v_0 u_{n-1} u_{n-2} \ldots u_m \rangle, \\ V^m &= \langle v_m v_{m-1} \ldots v_0 u_{n-1} u_{n-2} \ldots u_{m+1} \rangle, \\ z^m &= \langle v_{m-1} \ldots v_0 u_{n-1} u_{n-2} \ldots u_{m+1} \rangle. \end{aligned} \quad (1)$$

Здесь m – номер слоя, а локальные переменные u_m, v_m определяют позиционные номера рецепторов и аксонов базовых операций в слое m. Кортежи используются для поразрядного представления чисел в позиционной системе счисления. Например, для системы счисления с основанием 2 имеем:

$$u = \langle u_{n-1} u_{n-2} \ldots u_0 u_1 \rangle = u_{n-1} 2^{n-1} + u_{n-2} 2^{n-2} + \ldots + u_1 2 + u_0 .$$

Из выражения (1) следует, что для нулевого слоя номер базовой операции (далее – нейронного ядра) определяется выражением:

$$z^0 = \langle u_{n-1} u_{n-2} \ldots u_1 \rangle ,$$

а глобальный номер рецептора нулевого нейронного слоя выражением:

$$U^0 = \langle u_{n-1} u_{n-2} \ldots u_0 \rangle . \quad (2)$$

Локальные переменные u_0, v_0 определяют позиционный номер рецептора и позиционный номер аксона в пределах каждого ядра нулевого слоя (нумерация рецепторов и аксонов начинается с нуля). Будем полагать, что значение p_0 определяет число рецепторов ядра нулевого слоя, а значение g_0 − число аксонов. Аналогичные обозначения p_m, g_m, где

$m = 0,1,\ldots,n-1$, будем использовать для обозначения размерностей рецепторных и аксоновых полей остальных слоёв сети.

Выберем значения $g_m = 1$ для $m = 1,2,\ldots,n-1$, а для $m = 0$ сохраним произвольное значение g_0. В этом случае из (1) следует, что аксоны последнего слоя получат глобальные позиционные номера:

$$V^{n-1} = \langle 0_{n-1} 0_{n-2} \ldots 0_1 v_0 \rangle .$$

Таким образом, нейронная сеть будет иметь ровно g_0 выходов, а из (2) следует, что число входов будет равно $N = p_0 \cdot p_1 \cdot \ldots \cdot p_{n-1}$. Построенная сеть является пирамидальной, с размерностью N по входу и g_0 по выходу. На рис. 1 приведён пример трёхслойной пирамидальной сети для структуры $(p_0, p_1, p_2) = (4,2,2)$, $(g_0, g_1, g_2) = (3,1,1)$.

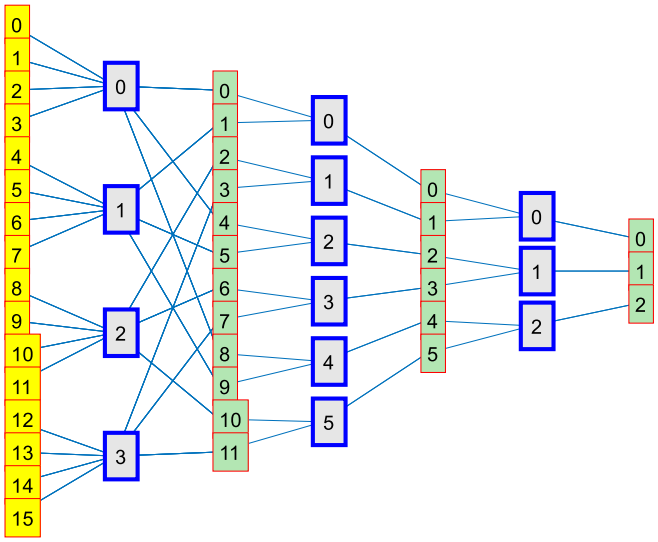


Рис. 1. Трехслойная пирамидальная БНС с регулярной топологией

Детали построения топологического графа БНС изложены в работе [6]. Для БНС доказана обобщённая теорема факторизации (см. Часть 2), согласно которой элементы матрицы преобразования представимы через произведение элементов нейронных ядер:

$$h(U,V) = w^0_{z^0}(u_0, v_0)\, w^1_{z^1}(u_1, v_1) \ldots w^{n-2}_{z^{n-2}}(u_{n-2}, v_{n-2})\, w^{n-1}_{z^{n-1}}(u_{n-1}, v_{n-1}) . \quad (3)$$

Здесь элементы матриц нейронных ядер $W^m_{i^m}$ обозначены через $w^m_{z^m}(u_m, v_m)$, где u_m, v_m – определяют номер строки и номер столбца матрицы, m – номер слоя, и z^m – номер нейронного ядра в пределах слоя m.

В Части 2 приведён алгоритм мультипликативной декомпозиции произвольной функции, заданной на дискретном интервале длиной $N = p_0 p_1 \cdots p_{n-1}$. Мультипликативное представление имеет вид:

$$f(u) = f_0 \langle u_{n-1} u_{n-2} \ldots u_0 \rangle = \phi_{i^0}(u_0)\, \phi_{i^1}(u_1) \ldots \phi_{i^{n-2}}(u_{n-2})\, \phi_{i^{n-1}}(u_{n-1}), \quad (4)$$

где $i^m = \langle u_{n-1} u_{n-2} \ldots u_{m+1} \rangle$. Аргумент функции представляется в виде числа в многоосновной системе счисления:

$$u = \langle u_{n-1} u_{n-2} \ldots u_0 \rangle = u_{n-1} p_{n-2} p_{n-3} \ldots p_0 + \cdots + u_{n-2} p_{n-3} p_{n-4} \ldots p_0 + \ldots + u_1 p_0 + u_0 .$$

Здесь $u_i \in [0,1,\ldots p_i - 1]$ – разрядные переменные. Нетрудно заметить, что выражения (3) и (4) подобны, учитывая, что для построенной пирамидальной сети

$$z^m = \langle 0_{m-1} \ldots 0_1 v_0 u_{n-1} u_{n-2} \ldots u_{m+1} \rangle ,$$

можно записать следующее правило обучения нейронных ядер к эталонным функциям:

$$w^0_{z^0}(u_0, v_0) = \phi^k_{i^0}(u_0), \quad m = 0,$$
$$w^m_{z^m}(u_m, 0) = \phi^k_{i^m}(u_m), \quad m = 1, 2, \ldots, n-1,$$
$$z^m = \langle 0_{m-1} \ldots 0_1 v_0 u_{n-1} u_{n-2} \ldots u_{m+1} \rangle,$$
$$i^m = \langle u_{n-1} u_{n-2} \ldots u_{m+1} \rangle,$$

где k – номер эталонной функции. Взаимно-однозначное соответствие $k \leftrightarrow v_0$ определяет упорядочение эталонных функций по их привязке к нейронам выходного слоя.

2. Многоканальные корреляторы на пирамидальных нейронных сетях. Один из основных способов распознавания сигналов при корреляционных измерениях состоит в сопоставлении сигнала с эталоном. Если сходство между неизвестным сигналом и эталоном велико, то сигнал помечается, как соответствующий эталонному. Простейшей мерой сходства является линейный дискриминант Фишера [7]. В частном случае этот дискриминант представляет собой меру взаимной энергии сигналов и выражается через нормированное скалярное произведение

$$d = (x, e) \Big/ \sqrt{(x, x)(e, e)} > \alpha,$$

где x – неизвестный сигнал, представленный вектором, e – вектор эталонного сигнала, скобки $(,)$ – обозначают скалярное произведение векторов, скаляр α определяет пороговый уровень различения сигналов. В многоканальном дискриминаторе используются несколько эталонов. Будем полагать, что входной и эталонные сигналы нормированы к энергии, так что $(x, x) = 1$ и $(e, e) = 1$. В этом случае дискриминант определяется скалярным произведением

$$d = (x, e).$$

Скалярное произведение можно рассматривать как произведение вектора-строки x на одно-столбцовую матрицу $H = e'$. Для многоканального дискриминатора матрица H будет состоять из g_0 столбцов и вычисление дискриминантов будет определяться выражением:

$$d(k) = \sum_U x(U) h(U, V).$$

Будем полагать, что матрица H является матрицей быстрого пирамидального преобразования с топологией Гуда. Поскольку сеть имеет g_0 выходов, то

$$V^{n-1} = \langle 0_{n-1} 0_{n-2} \ldots 0_1 v_0 \rangle.$$

На каждом выходе будет формироваться значение скалярного произведения, отвечающего эталонному сигналу. Пирамидальная сеть, показанная на рис. 1, может быть обучена к трём сигнальным функциям, заданным на интервале длинной 16.

Замечание. Поскольку для предложенной архитектуры пирамидальной сети выполняются условия: $g_m = 1$ для всех $m = 1, 2, \ldots, n-1$, то из (1) следует

$$U^m = \langle 0_{m-1} 0_{m-2} \ldots 0_1 v_0 u_{n-1} u_{n-2} \ldots u_m \rangle,$$
$$V^m = \langle 0_m 0_{m-1} \ldots 0_1 v_0 u_{n-1} u_{n-2} \ldots u_{m+1} \rangle,$$
$$z^m = \langle 0_{m-1} \ldots 0_1 v_0 u_{n-1} u_{n-2} \ldots u_{m+1} \rangle.$$

При варьировании разрядных переменных u_i при фиксированном значении разряда v_0 значения кортежей выделяют в сети множества рецепторов, аксонов и нейронных ядер, помеченных разрядом v_0. Подчеркнём, что в БНС рецепторные поля нейронных ядер не пересекаются, и каждый нейрон слоя имеет точно один аксон [6], отсюда следует, что

пирамидальная нейронная сеть предложенной архитектуры распадается на независимые группы нейронов, индексированных разрядом v_0. Каждая такая группа нейронов способна обучаться независимо от других групп. Таким образом, пирамидальная нейронная сеть рассматриваемого типа обладает уникальной возможностью последовательного дообучения к эталонам без потери ранее накопленных знаний.

3. Пластичность пирамидальной нейронной сети. В биологии термин пластичность используется, как качественная характеристика способности нейронной сети обучаться при воздействии внешних факторов. Для искусственных нейронных сетей адекватной количественной оценкой может служить число независимых настроек, называемое также числом степеней свободы. Это значение, как правило, меньше полного количества настраиваемых синаптических весов нейронной сети. Для БНС получены точные формулы расчёта числа степеней свободы [6]. Исходными данными для расчёта является структурная модель нейронной сети. Структурная модель – это взвешенный граф, вершинами которого являются нейронные ядра. Вес каждого ядра определяется парой чисел, представляющих размерность по входу и выходу ядра, а вес дуг определяется рангами операторов межъядерной связи. На рис. 2 представлена структурная модель для топологической модели сети, показанной на рис. 1.

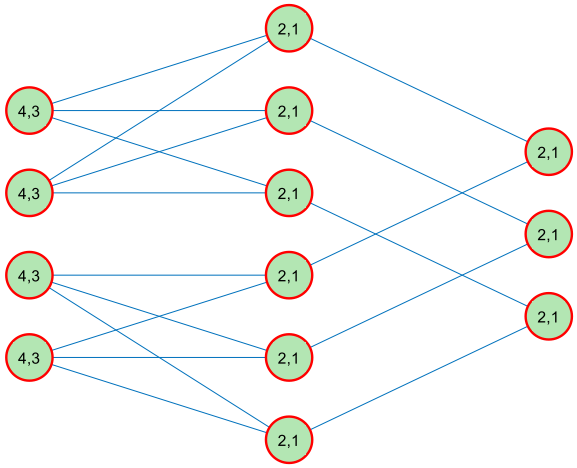


Рис. 2. Структурная модель пирамидальной нейронной сети

Для БНС ранги операторов межъядерных связей равны 1 и на графе структурной модели условно не показаны. Формула расчёта числа степеней свободы для БНС имеет вид:

$$S(H)=\sum_{m=0}^{n-1} p_m g_m q_m - \sum_{m=0}^{n-2} D_m , \qquad (5)$$

где p_m, g_m размерности рецепторного и аксонового поля ядра в слое m, q_m – число ядер в слое, D_m – количество одноранговых связей в межслойном переходе с номером m. Для структурной модели, представленной на рис. 2, непосредственным подсчётом получим $S(H)=4\cdot3\cdot4+2\cdot1\cdot6+2\cdot1\cdot3-12-6=48$. Напомним, что данная сеть может обучиться к трём эталонным функциям, заданным на интервале длиной 16. При произвольном выборе трёх функций необходимо задать 48 значений, что совпадает с числом степеней свободы пирамидальной сети. Таким образом, рассматриваемая сеть является глубокой в том смысле, что её потенциал обучения используется полностью, и покрывает все допустимое многообразие эталонных сигналов.

4. Многомерные пирамидальные сети. Архитектура быстрой пирамидальной нейронной сети легко обобщается на многомерный случай [6]. Выражения (1) для топологии сети в этом случае примут вид:

$$U_*^m = \left\langle 0_{m-1}^* 0_{m-2}^* \ldots 0_1^* v_0^* u_{n-1}^* u_{n-2}^* \ldots u_m^* \right\rangle,$$
$$V_*^m = \left\langle 0_m^* 0_{m-1}^* \ldots 0_1^* v_0^* u_{n-1}^* u_{n-2}^* \ldots u_{m+1}^* \right\rangle,$$
$$i_*^m = \left\langle 0_{m-1}^* 0_{m-2*}^* \ldots 0_1^* v_0^* u_{n-1}^* u_{n-2}^* \ldots u_{m+1}^* \right\rangle,$$

Звездочки «*» в верхнем индексе заменяют обозначения пространственных координат x, y. Из последнего выражения следует правило обучения нейронной сети:

$$w_{i_x^m, i_y^m}^m \left(u_m^y u_m^x; v_m^y v_m^x \right) = f_{i_x^m, i_y^m}^k \left(u_m^y, u_m^x \right), \quad m = 0,$$
$$w_{i_x^m, i_y^m}^m \left(u_m^y u_m^x; 0_m^y 0_m^x \right) = f_{i_x^m, i_y^m}^k \left(u_m^y, u_m^x \right), \quad m > 0,$$
$$i_*^m = \left\langle 0_{m-1}^* 0_{m-2*}^* \ldots 0_1^* v_0^* u_{n-1}^* u_{n-2}^* \ldots u_{m+1}^* \right\rangle.$$

Здесь $w_{i_x^m, i_y^m}^m \left(u_m^y u_m^x; v_m^y v_m^x \right)$ – двумерные нейронные ядра, $f_{i_x^m, i_y^m}^k \left(u_m^y, u_m^x \right)$ – множители мультипликативного разложения эталонного изображения, x, y – пространственные координаты. Взаимно-однозначное соответствие $k \leftrightarrow \left(v_m^y, v_m^x \right)$ определяет упорядочение эталонных функций в выходной плоскости. На рис. 3 представлен топологический граф корреляционной пирамидальной сети для изображения $N \times N = 2^3 \times 2^3$ со структурными параметрами $g_0^x = 4, \quad g_0^y = 2$. Сеть способна выполнять корреляционную обработку для восьми изображений.

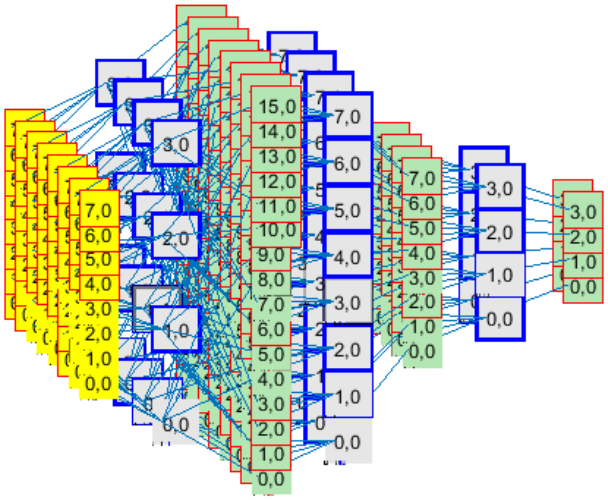


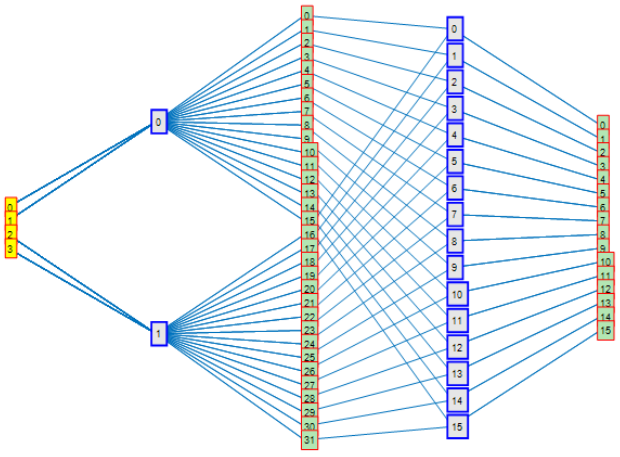
Рис. 3. Топология регулярной пирамидальной 2D сети

5. Реализация комбинационной логики на нейронных сетях. Простые логические элементы комбинационной логики «И», «ИЛИ», «НЕТ» легко реализуются на одиночных нейронах с пороговой активационной функцией. Немного сложней реализуется элемент «Исключающее ИЛИ», который требует несколько нейронов в двухслойной сети. Для реализации сложных логических функций возможен путь, повторяющий схемотехнику логических устройств, но это решение исключает потенциал нейронных сетей, связанный с обучением. Альтернативным вариантом является реализация сложной логики через обучение нейронной сети. Однако в этом направлении возникают проблемы целесообразного выбора структуры нейронной сети так, чтобы она была достаточной и не избыточной для реализации логических функций. Кроме того, классические методы обучения нейронных сетей, построенные на основе алгоритма с обратным распространением ошибки, не гарантируют сходимость, что добавляет неопределённость в выборе сети.

Использование пирамидальных самоподобных сетей, позволяет эффективно решить эти вопросы. Рассмотрим для примера построение логического дешифратора на основе пирамидальной нейронной сети. Как было показано ранее, пирамидальные сети быстрого обучения могут быть точно настроены на произвольные эталонные функции. Количество эталонов определяется выбранной топологией сети. Будем полагать, что настройка сети выполнена к эталонным функциям, нормированным по энергии к единице и представляющим собой логические коды по основанию 2. В рабочем режиме сеть выполняет вычисление скалярных произведений входной вектор-функции с эталонными функциями. Если входная функция нормирована по энергии к единице и совпадает с эталоном, то на соответствующем выходе нейронной сети получим единичное значение, в то время как остальные выходы будут иметь значения меньше единицы. Это обстоятельство можно использовать для построения дешифраторов логических кодов.

Рассмотрим, например, построение дешифратора для четырёхбуквенных кодовых слов. Для кода с основанием 2 дешифратор должен иметь 4 входа и 16 выходов. Нейронная сеть должна иметь такие же размерности по входу и выходу. Выберем одномерную пирамидальную двухслойную нейронную сеть, топология сети показана на рис. 4.

Сеть обучается к кодам, представленным в таблице на этом же рисунке. Коды рассматриваются, как вектор-функции. Обучающие эталонные функции формируются из кодов нормированием по энергии к единице. Сеть имеет структурные характеристики: по рецепторным полям $P=[p_0 p_1]=[2,2]$, по аксоновым полям – $G=[g_0 g_1]=[16,1]$. Результаты классификации входных кодов показаны на рис. 5. Максимальный уровень выхода при классификации ближайшего кода, отличного от собственного, не превышает значения 0.886. Компараторы логического нуля на выходах сети должны быть настроены на значения из интервала [0.886+,1.000-].



Входной код	Номер выхода
0000	0
0001	1
0010	2
0011	3
0100	4
0101	5
0110	6
0111	7
1000	8
1001	9
1010	10
1011	11
1100	12
1101	13
1110	14
1111	15

Рис. 4. Дешифратор логических кодов на пирамидальной нейронной сети

Проблемным является входной код [0000]. Скалярное произведение этого кода с эталоном будет равно нулю, в то время как в классическом дешифраторе оно должно равняться единице.

Решением является видоизменение нулевого кода, например, замена одного из нулей кодового слова [0000] значением (-1). Более радикальным вариантом является замена во всех кодовых словах значения (0) на значение (-1), т.е. использование двух-полярного бинарного кода со значениями (-1,+1). Результаты классификации входных кодов для этого случая показаны на рис. 6. Выходные компараторы логического нуля должны быть настроены на

уровень 0.5+. Подобным образом могут быть построены дешифраторы кодов с основанием больше двух.

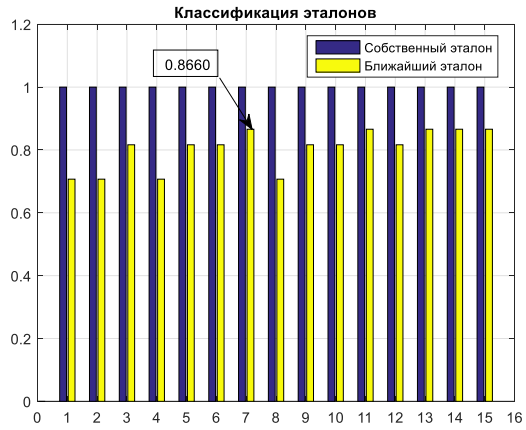


Рис. 5. Классификация логических кодов пирамидальной нейронной сетью

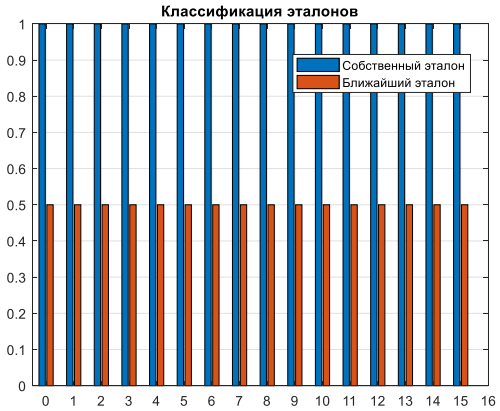


Рис. 6. Классификация двух-полярных бинарных кодов

6. Реализация элементов памяти на нейронных сетях. К нейронным сетям, реализующим память, относится большой класс нейронных сетей с обратными связями, такие, как сети Элмана [8], Хопфилда [9], Хэмминга [10], АРТ [11] и другие. Эти сети решают задачи восстановления искажённых образов, ассоциативной памяти, кратковременной динамической памяти. Нейронные сети данного типа, как правило, используются в контексте распознавания образов, задача точного восстановления образа при этом не ставится.

Рассмотрим построение элементов памяти на обратно-ориентированных пирамидальных нейронных сетях. Возьмём за основу топологическую модель «Кули-Тьюки с прореживанием по времени» (см. Часть 1).

$$U^m = \langle u_{n-1}u_{n-2}\cdots u_{m+1}u_m v_{m-1}v_{m-2}\cdots v_1 v_0 \rangle,$$
$$V^m = \langle u_{n-1}u_{n-2}\cdots u_{m+1}v_m v_{m-1}v_{m-2}\cdots v_1 v_0 \rangle,$$
$$z^m = \langle u_{n-1}u_{n-2}\cdots u_{m+1}v_{m-1}v_{m-2}\cdots v_1 v_0 \rangle.$$

Выберем структурные характеристики следующим образом:

- размерности рецепторных полей сети положим равными $p_0 = p_1 = \cdots = p_{n-2} = 1, \quad p_{n-1} \neq 1$, в этом случае размерность сети по входу будет равна $N = p_0 p_1 \cdots p_{n-1} = p_{n-1}$;
- размерности аксоновых полей сети зададим произвольными натуральными целыми числами $g_0, g_1, \cdots, g_{n-1}$, размерность сети по выходу в этом случае будет равна $M = g_0 g_1 \cdots g_{n-1}$.

Топологическая модель при данных структурных характеристиках будет иметь вид:

$$U^m = \langle u_{n-1}0_{n-2}\cdots 0_{m+1}0_m v_{m-1}v_{m-2}\cdots v_1 v_0 \rangle,$$
$$V^m = \langle u_{n-1}0_{n-2}\cdots 0_{m+1}v_m v_{m-1}v_{m-2}\cdots v_1 v_0 \rangle,$$
$$z^m = \langle u_{n-1}0_{n-2}\cdots 0_{m+1}v_{m-1}v_{m-2}\cdots v_1 v_0 \rangle.$$

Топология сети для структурных параметров $[p_0 p_1 p_2 p_3] = [1112]$, $[g_0 g_1 g_2 g_3] = [2222]$ показана на рис. 7. В соответствии с обобщенной теоремой факторизации (см. Часть 2) элементы матрицы БНС представляются в виде

$$h(U,V) = w_{z^{n-1}}^{n-1}(u_{n-1}, v_{n-1})\, w_{z^{n-2}}^{n-2}(0_{n-2}, v_{n-2})\cdots w_{z^0}^{0}(0_0, v_0). \qquad (6)$$

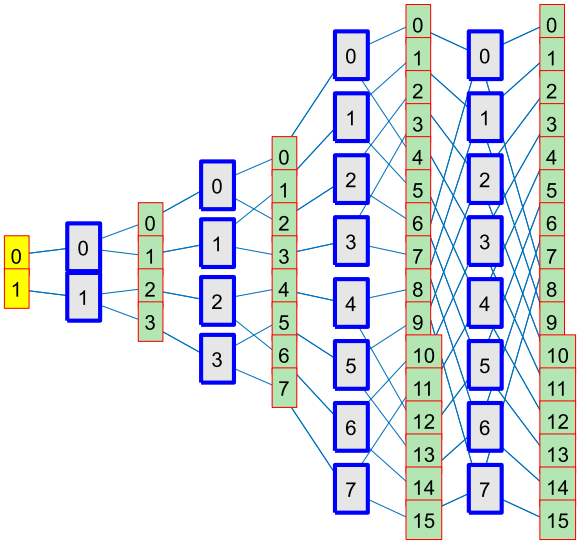


Рис. 7. Обратно-ориентированная пирамидальная сеть для реализации памяти

Пусть запоминаемые образы $y^k(V)$ представляют собой набор, состоящий из p_{n-1} дискретных функций, заданных на интервале длинной M. Выполним мультипликативную декомпозицию каждой функции по переменным v_i, начиная со старшего разряда (см. Часть 2), в результате получим:

$$y^k(V) = \varphi^k_{i^{n-1}}(v_{n-1})\,\varphi^k_{i^{n-2}}(v_{n-2})\cdots\varphi^k_{i^0}(v_0), \tag{7}$$

где $i^m = \langle v_{m-1}v_{m-2}\cdots v_1v_0\rangle$, $k-$ порядковый номер функции. Сравнивая выражения (6) и (7), получим правило обучения нейронной сети:

$$w^{n-1}_{z^{n-1}}(u_{n-1}, v_{n-1}) = \varphi^k_{i^{n-1}}(v_{n-1}) \quad для\ m = n-1,$$
$$w^m_{z^m}(0_m, v_m) = \varphi^k_{i^m}(v_m) \quad для\ m < n-1,$$
$$z^m = \langle u_{n-1}0_{n-2}\cdots 0_{m+1}v_{m-1}v_{m-2}\cdots v_1v_0\rangle,$$
$$i^m = \langle v_{m-1}v_{m-2}\cdots v_1v_0\rangle.$$

Предварительно, для упорядочивания хранимых функций, должно быть установлено взаимно-однозначное соответствие $k \leftrightarrow u_{n-1}$ между порядковым номером функции и разрядной переменной u_{n-1}. Считывание памяти реализуется установкой на входе нейронной сети унитарного кода, в котором только один из разрядов равен 1, а остальные нули. Для сети, показанной на рис. 7, это будут комбинации $[01]$ или $[10]$. При использовании кодов $[11]$ хранимые образы складываются, а при кодах $[1\ -1]$ из первого образа вычитается второй. На рис. 8 показаны результаты считывания образов, представляющих собой дискретные функции синуса, отличающиеся по частоте в два раза.

На рис. 9 показана структурная модель для нейронной сети, представленной на рис.8. Каждая вершина структурной модели является нейронным ядром, пара цифр, записанных в вершине, определяют размерность рецепторного и аксонового поля нейронного ядра, все связи между вершинами имеют ранги, равные единице.

Используя формулу (5) для расчёта числа степеней свободы из структурной модели, непосредственно получим $S(H) = (1\cdot 2)\cdot 14 + (2\cdot 2)\cdot 8 - 28 = 32$. Данная нейронная сеть обеспечивает хранение двух произвольных дискретных функций, заданных на интервале длиной $M = 16$. Для задания двух функций требуется определить 32 значения. Таким образом,

сеть полностью использует свой потенциал для хранения функций и поэтому относится к категории сетей глубокого обучения.

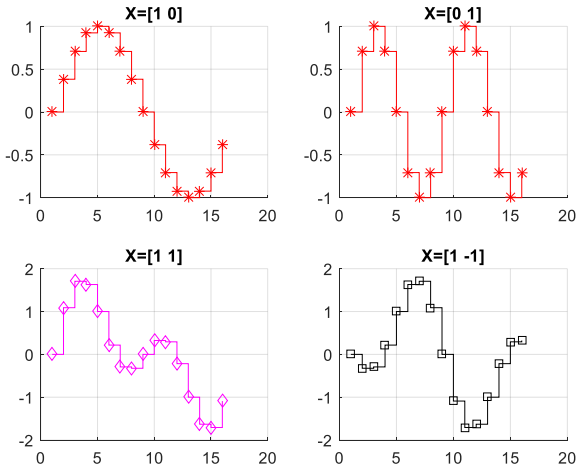


Рис. 8. Результаты считывания памяти, реализованной на пирамидальной БНС

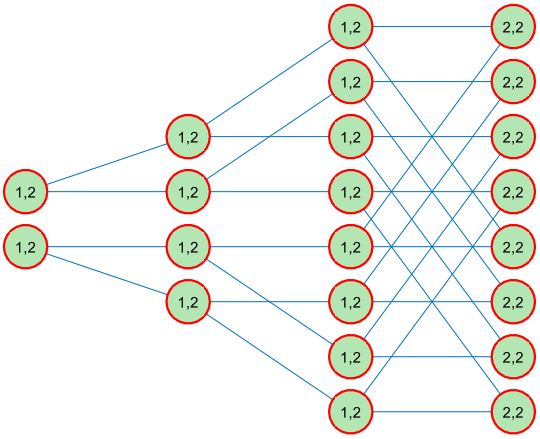


Рис. 9. Структурная модель обратно ориентированной пирамидальной нейронной сети

В памяти на основе БНС могут храниться двумерные и многомерные образы. В отличие от кристаллов цифровой памяти с последовательным хранением данных, где восстановление образа происходит за счёт последовательного опроса ячеек, в памяти на БНС все пиксели образа восстанавливаются одновременно, что потенциально обеспечивает сверхвысокое быстродействие при считывании памяти.

Рассмотрим для примера построение логического шифратора на основе пирамидальной нейронной сети для двоичного четырехразрядного кода. Шифратор выполняет операцию обратную дешифрированию, т.е. преобразует унарный код в кодовые слова. Эту операцию можно выполнить на элементе памяти образов. Разместим в памяти матрицу кодовых слов, тогда логическую функцию шифратора можно реализовать, как умножение унарного вектора-строки длиной 16 на матрицу кодовых слов размерностью 16×4:

$$OUT = [V_3V_2V_1V_0] = [U_0U_1U_2U_3U_4U_5U_6U_7U_8U_9U_{10}U_{11}U_{12}U_{13}U_{14}U_{15}] \begin{bmatrix} 0000 \\ 0001 \\ 0010 \\ 0011 \\ 0100 \\ 0101 \\ 0110 \\ 0111 \\ 1000 \\ 1001 \\ 1010 \\ 1011 \\ 1100 \\ 1101 \\ 1111 \end{bmatrix},$$

где только один элемент вектора-строки равен единице, а остальные равны нулю.

Покажем реализацию памяти образов на основе БНС. Каждую строку матрицы кодовых слов будем рассматривать как функцию $f^k(V) = f^k(v_{n-1}v_{n-2}\cdots v_0)$, где k – номер строки, v_i – разрядные переменные. Выполним мультипликативную декомпозицию (см. Часть 2) каждой функции по переменным v_i, начиная со старшего разряда, в результате получим:

$$f^k(V) = \varphi^k_{i^{n-1}}(v_{n-1})\,\varphi^k_{i^{n-2}}(v_{n-2})\cdots\varphi^k_{i^0}(v_0),$$

где $i^m = \langle v_{m-1}v_{m-2}\cdots v_1v_0\rangle$. Возьмём для примера модель БНС с топологией Кули-Тьюки «с прореживанием по частоте» (см. Часть 1)

$$\begin{aligned} U^m &= \langle v_{m-1}v_{m-2}\ldots v_0u_{n-1}u_{n-2}\ldots u_m\rangle,\\ V^m &= \langle v_mv_{m-1}\ldots v_0u_{n-1}u_{n-2}\ldots u_{m+1}\rangle,\\ z^m &= \langle v_{m-1}\ldots v_0u_{n-1}u_{n-2}\ldots u_{m+1}\rangle. \end{aligned}$$

Выберем структурные характеристики БНС следующим образом:

– размерности рецепторных полей сети положим равными $p_0 = p_1 = \cdots = p_{n-2} = 1, \quad p_{n-1} \neq 1$, в этом случае размерность сети по входу будет равна $N = p_0p_1\cdots p_{n-1} = p_{n-1}$;

– размерности аксоновых полей сети зададим произвольными натуральными целыми числами $g_0, g_1, \cdots, g_{n-1}$, размерность сети по выходу в этом случае будет равна $M = g_0g_1\cdots g_{n-1}$.

Топологическая модель при данных структурных характеристиках будет иметь вид:

$$\begin{aligned} U^m &= \langle v_{m-1}v_{m-2}\ldots v_0u_{n-1}0_{n-2}\ldots 0_m\rangle,\\ V^m &= \langle v_mv_{m-1}\ldots v_0u_{n-1}0_{n-2}\ldots 0_{m+1}\rangle,\\ z^m &= \langle v_{m-1}\ldots v_0u_{n-1}0_{n-2}\ldots 0_{m+1}\rangle. \end{aligned} \tag{8}$$

Сравнивая выражения (6) и (7) с учётом (8), получим правило обучения нейронной сети:

$$\begin{aligned} &w^{n-1}_{z^{n-1}}(u_{n-1}, v_{n-1}) = \varphi^k_{i^{n-1}}(v_{n-1}) \quad для\ m = n-1,\\ &w^m_{z^m}(0_m, v_m) = \varphi^k_{i^m}(v_m) \quad для\ m < n-1,\\ &z^m = \langle v_{m-1}v_{m-2}\cdots v_1v_0u_{n-1}0_{n-2}\cdots 0_{m+1}\rangle,\\ &i^m = \langle v_{m-1}v_{m-2}\cdots v_1v_0\rangle. \end{aligned}$$

Иначе его можно назвать правилом записи образа в нейро-сетевую память. Предварительно, для корректного упорядочивания хранимых функций, должно быть установлено соответствие $k \leftrightarrow u_{n-1}$ между порядковым номером функции и разрядной переменной u_{n-1}. Данная сеть памяти способна хранить p_{n-1} произвольных образов. Считывание памяти реализуется установкой на входе нейронной сети унарного кода, в котором только один из разрядов равен 1, а остальные равны нулю.

Для построения шифратора 16×4 выберем двухслойную нейронную сеть со структурными характеристиками $P = [p_0, p_1] = [1,16]$ и $G = [g_0, g_1] = [2,2]$. Граф топологической модели сети показан на рис. 10.

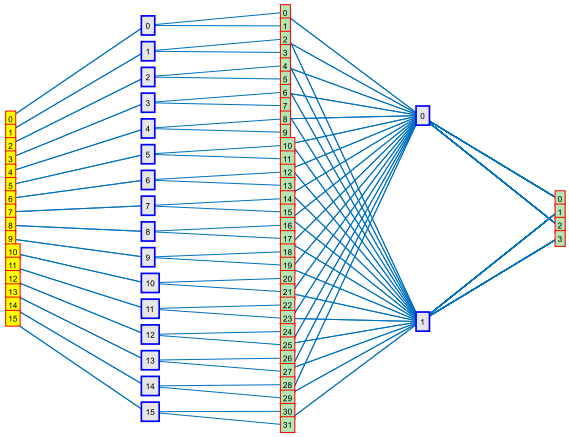


Рис. 10. Нейронная сеть памяти реализующая шифратор 16×4

Двухслойная нейронная сеть по входу имеет размерность 16, а по выходу 4. Ядра нулевого слоя имеют размерности [1, 2], а выходного – [16, 2]. Для сети памяти уровни логических значений могут быть стандартными: 0 и 1.

Заключение. Пирамидальные регулярные нейронные сети можно рассматривать, как частный случай быстрых нейронных сетей, которые, в свою очередь, относятся к широкому классу самоподобных многослойных нейронных сетей. Самодобные сети с регулярной структурой обладают уникальным свойством аналитического представления графа топологической модели, что позволяет разработать быстрые абсолютно сходящиеся алгоритмы обучения с конечным числом вычислительных операций. Кроме того, процесс обучения для каждого эталона может выполняться независимо от других, т.е. пирамидальная сеть данного типа обладает важным качеством накапливать знания, без необходимости переобучения сети при добавлении новых данных.

Пирамидальные сети обладают минимально-достаточной структурой для обучения к заданному многообразию функций. В статье показано, что их информационная ёмкость используется полностью. Сети данного типа удовлетворяют принципу двойственности, они в равной степени могут быть использованы как для реализации многоканальных корреляторов, так и для создания элементов памяти образов. В данной работе этот принцип продемонстрирован на примере построения шифраторов и дешифраторов бинарных кодов. Подобным образом могут быть реализованы другие сложные логические функции, в том числе функции многозначной логики. Последовательное соединение нейросетевого дешифратора и шифратора можно использовать для построения преобразователей кода, например, равномерного двоичного кода в код Грея. Эта же модель может быть легко перенастроена для выполнения обратного преобразования. Более того, регулярный лес пирамидальных сетей позволяет реализовать быстрые нейронные сети глубокого обучения с расширенными возможностями по информационной ёмкости [12].

В памяти, реализованной на основе пирамидальных БНС, могут храниться двумерные и многомерные образы. В отличие от кристаллов цифровой памяти с последовательным хранением данных, где восстановление образа происходит за счёт последовательного опроса ячеек, в памяти на БНС все пиксели образа восстанавливаются одновременно, что потенциально обеспечивает сверхвысокое быстродействие для памяти образов.

Список источников

1. Дорогов А.Ю. Быстрые преобразования и самоподобные нейронные сети глубокого обучения. Часть 1. Стратифицированные модели самоподобных нейронных сетей и быстрых преобразований / А.Ю. Дорогов // Информационные и математические технологии в науке и управлении, 2023. – №4 (32). – С. 5-20.

2. Дорогов А.Ю. Быстрые преобразования и самоподобные нейронные сети глубокого обучения. Часть 2. Методы обучения быстрых нейронных сетей / А.Ю. Дорогов // Информационные и математические технологии в науке и управлении. – 2024. – № 1(33). – С. 5-19. – DOI:10.25729/ESI.2024.33.1.001.
3. Gonzalez R.C., Woods R.E. Digital image processing. Englewood Cliffs, NJ, Prentice-Hall, 2002.
4. Phung S.L., Bouzerdoum A. A pyramidal neural network for visual pattern recognition. IEEE Transactions on Neural Networks, 2007, vol. 18, no. 2, pp. 329-343.
5. Ullahy Ihs., Petrosinoy A. About pyramid structure in convolutional neural networks. arXiv:1608.04064v1 [cs.CV], 2016.
6. Дорогов А.Ю. Теория и проектирование быстрых перестраиваемых преобразований и слабосвязанных нейронных сетей / А.Ю. Дорогов. – СПб.: «Политехника», 2014. – 328с.
7. Фишер Р.А. Статистические методы для исследователей. Москва, Госизадат, 1958. – 267с.
8. Elman J.L. Finding structure in time. Cognitive science, 1990, pp. 179-211.
9. Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities. Proceedings of National academy of sciences, 1982, vol. 79, no. 8, pp. 2554–2558.
10. Lippmann R. An introduction to computing with neural nets. IEEE Assp magazine, 1987.
11. Carpenter G.A., Grossberg S. ART 2: Self-organization of stable category recognition codes for analog input patterns. Applied Optics, 1987, vol. 26, no. 23.
12. Дорогов А.Ю. Быстрые нейронные сети глубокого обучения. / А.Ю. Дорогов // Сборник докладов III Международной научной конференции по проблемам управления в технических системах (CTS'2019). Санкт-Петербург. 30 октября - 1 ноября 2019 г. СПб.: СПбГЭТУ «ЛЭТИ». С. 275-280.

Дорогов Александр Юрьевич. *Доктор технических наук, доцент, профессор кафедры «Автоматики и процессов управления» Санкт-Петербургского государственного электротехнического университета (СПбГЭТУ) «ЛЭТИ», главный научный сотрудник ПАО «Информационные телекоммуникационные технологии». Основные направления исследований автора связаны с интеллектуальным анализом данных, цифровой обработкой сигналов, проектированием быстрых преобразований и нейронных сетей быстрого обучения, разработкой аналитических платформ, моделированием радио-телекоммуникационных систем. AuthorID: 17611, SPIN: 8645-5873, ORCID: 0000-0002-7596-6761, vaksa2006@yandex.ru, г. Санкт-Петербург, ул. Попова, 5.*

UDC 004.032.26
DOI:10.25729/ESI.2024.34.2.002

Fast transformations and self-similar deep learning neural networks. Part 3. Pyramidal neural networks with a deep degree of learning

Alexander Yu. Dorogov

PJSC "Information Telecommunication Technologies" ("Inteltech"),
St. Petersburg State Electrotechnical University,
Russia, St. Petersburg, *vaksa2006@yandex.ru*

Abstract. The paper considers a class of fast neural networks with a pyramidal structure. Methods of topological construction of one-dimensional and two-dimensional pyramidal networks are given. Networks of the class under consideration are representable by linear operators, have a self-similar structure, and are a special case of the fast Fourier transform algorithm. Topological models of pyramidal neural networks of direct and reverse orientation are proposed. The paper shows the use of pyramidal neural networks of fast learning for the implementation of correlative digital signal and image processing, combinational logic and memory elements. Examples of the construction of an encoder and a decoder of binary codes are considered. It is noted that the pyramidal memory network provides storage and accurate recovery of images similar to storing data in random access computer memory. It is proved that a fast pyramid network is a deep learning neural network, and a self-similar structure allows the network to be trained to new data without the need for a complete retraining of the network. This work is the third part of the generalizing article "Fast transformations and self-similar neural networks of deep learning". In the first part, stratified models of self-similar neural networks are considered, in the second part, algorithms for training fast neural networks and generalized spectral transformations are considered.

Keywords: fast tunable transformation; neural network; pyramid structure; deep learning, code decoder, code encoder, neural network memory, neural network plasticity, degrees of freedom

References

1. Dorogov A.Yu. Bystrye preobrazovanija i samopodobnye nejronnye seti glubokogo obuchenija. Chast' 1. Stratificirovannye modeli samopodobnyh nejronnyh setej i bystryh preobrazovanij [Fast transformations and self-similar deep learning neural networks. Part 1. Stratified models of self-similar neural networks and fast transformations]. Informacionnye i matematicheskie tehnologii v nauke i upravlenii [Information and mathematical technologies in science and management], 2023, no. 4 (32), pp. 5-20.
2. Dorogov A.Yu. Bystrye preobrazovanija i samopodobnye nejronnye seti glubokogo obuchenija. Chast' 2. Metody obucheniya bystrykh neyronnykh setey [Fast transformations and self-similar deep learning neural networks. Part 2. Methods of training fast neural networks]. Informacionnye i matematicheskie tehnologii v nauke i upravlenii [Information and mathematical technologies in science and management], 2024, no. 1 (33), pp. 5-19.
3. Gonzalez R.C., Woods R.E. Digital image processing. Englewood Cliffs, NJ, Prentice-Hall, 2002.
4. Phung S.L., Bouzerdoum A. A pyramidal neural network for visual pattern recognition. IEEE Transactions on Neural Networks, 2007, vol. 18, no. 2, pp. 329-343.
5. Ullahy Ihs., Petrosinoy A. About pyramid structure in convolutional neural networks. arXiv:1608.04064v1 [cs.CV], 2016.
6. Dorogov A.Yu. Teorija i proektirovanie bystryh perestraivaemyh preobrazovanij i slabosvjazannyh nejronnyh setej [Theory and design of fast tunable transformations and loosely coupled neural networks]. Spb. Publ "Politehnika", 2014, 328p.
7. Fisher R. A. Statistical methods for research workers. Paperback – January 1, 2017, 267p.
8. Elman J.L. Finding structure in time. Cognitive science, 1990, pp. 179-211.
9. Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities. Proceedings of National academy of sciences, 1982, vol. 79, no. 8, pp. 2554–2558.
10. Lippmann R. An introduction to computing with neural nets. IEEE Assp magazine, 1987.
11. Carpenter G.A., Grossberg S. ART 2: Self-organization of stable category recognition codes for analog input patterns. Applied Optics, 1987, vol. 26, no. 23.
12. Dorogov A.Yu. Bystryye neyronnyye seti glubokogo obucheniya. [Fast Deep Learning Neural Networks]. Sbornik dokladov III Mezhdunarodnoy nauchnoy konferentsii po problemam upravleniya v tekhnicheskikh sistemakh (CTS'2019) [Proceedings of 2019 3rd International Conference on Control in Technical Systems, CTS 2019]. Saint Petersburg, LETI, 2019, pp. 275-280.

Dorogov Alexander Yurievich. *Doctor of technical sciences, associate professor, professor of the department of automation and control Processes of St. Petersburg state electrotechnical university (SPbETU) "LETI", Chief researcher of PJSC "Information telecommunication technologies". The main directions of the author's research are related to data mining, digital signal processing, the design of fast transformations and neural networks of fast learning, the development of analytical platforms, modeling of radio and telecommunications systems. AuthorID: 17611, SPIN: 8645-5873, ORCID: 0000-0002-7596-6761, vaksa2006@yandex.ru, St. Petersburg, st. Popova, 5.*

Статья поступила в редакцию 13.06.2023; одобрена после рецензирования 03.05.2024; принята к публикации 30.05.2024.

The article was submitted 06/13/2023; approved after reviewing 05/03/2024; accepted for publication 05/30/2024.

УДК 57.015 + 573.2
DOI:10.25729/ESI.2024.34.2.003

Упорядоченность значений GC-состава фрагментов в пространственной структуре геномов органелл

Сенашова Мария Юрьевна

Институт вычислительного моделирования СО РАН,
Россия, Красноярск, *msen@icm.krasn.ru*

Аннотация. Рассмотрено пространственное распределение значений GC-состава фрагментов геномов хлоропластов и митохондрий. Под пространственным распределением понимается распределение точек, соответствующих участкам геномов, в пространстве частот триплетов. Обнаружено, что значения GC-состава фрагментов для большинства геномов распределены не хаотически, а упорядоченно. Были обнаружены 2 основных типа распределения: градиентное и центрально-симметричное. У геномов хлоропластов встречается только градиентное распределение. У митохондрий встречаются оба типа распределения. Тип распределения для митохондрий зависит от вида организма. Пространственное распределение GC-состава является устойчивым относительно изменения длины окна считывания.

Ключевые слова: порядок, пространственное распределение, триплеты, частотные словари, главные компоненты

Цитирование: Сенашова М.Ю. Упорядоченность значений GC-состава фрагментов в пространственной структуре геномов органелл / М.Ю. Сенашова // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 33-40. – DOI:10.25729/ESI.2024.34.2.003.

Введение. Изучение особенностей и деталей структуры нуклеотидных последовательностей является важнейшей задачей биологии в настоящее время. Нуклеотидные последовательности являются очень интересным объектом исследования как с точки зрения биологии, так и с точки зрения биоинформатики. Биологи и биоинформатики заинтересованы в обнаружении упорядоченных структур в этих последовательностях. Однако, биология изучает нуклеотидные последовательности с точки зрения их функционирования, выявления тех или иных физических структур, их физико-химического взаимодействия, а биоинформатика рассматривает нуклеотидные последовательности как символьные и анализирует их с использованием математических методов.

Такой параметр, как GC-состав очень часто используется в исследованиях структуры геномов и функций отдельных участков. Этому посвящено много работ, как в случае геномов хлоропластов [1-8], так и геномов митохондрий [9-15]. Под GC-составом понимают долю гуанина и цитозина по отношению к длине рассматриваемой нуклеотидной последовательности. Пара гуанин-цитозин соединена тремя водородными связями, тогда как пара аденин–тимин — двумя. Это обуславливает разные физические свойства у GC-бедных и GC-богатых нуклеотидных последовательностей.

В данной работе нуклеотидные последовательности рассматриваются с точки зрения биоинформатики. Изучается распределение значений GC-состава выделенных фрагментов геномов хлоропластов и митохондрий в пространственной структуре генома, полученной на основе частотных словарей этих фрагментов. Возникает вопрос – имеется ли упорядоченность значений GC-состава в нуклеотидных последовательностях, либо они расположены случайным образом. Естественно было бы ожидать случайное распределение значений GC-состава, но как показывают результаты работы, наблюдается достаточно высокая упорядоченность такого распределения.

1. Материалы и методы. Введем основные понятия. Мы будем рассматривать генетическую последовательность длины L, состоящую из символов алфавита $\aleph=\{A,C,G,T\}$. Для этой последовательности мы будем составлять частотный словарь толщины 3. Частотный

словарь W_3 толщины 3 символьной последовательности – это список всех троек $\omega = \nu_1\nu_2\nu_3$ идущих подряд нуклеотидов с указанием частот этих троек; всего может быть 64 триплета. Мы используем частотный словарь, в котором триплеты подсчитываются таким образом, что они полностью покрывают последовательность и при этом не пересекаются. Частота f_ω – это отношение числа копий n_ω данного слова к общему числу всех триплетов N, где N – сумма всех n_ω:

$$f_\omega = \frac{n_\omega}{N} \tag{1}$$

Всякий частотный словарь W_3 отображает геном в 64-мерное метрическое пространство.

Один из 64 триплетов исключался, поскольку сумма всех частот в словаре равна 1, что порождает линейную связь, которая будет давать ложный сигнал при статистической обработке (корреляционном анализе, определении главных компонент и т.п.).

В нашем случае целесообразнее исключать тот триплет, для которого стандартное отклонение, наблюдаемое по анализируемому набору частотных словарей, является минимальным: такой триплет дает наименьший вклад в различимость объектов (в предельном случае, когда стандартное отклонение равно 0, различий по этому триплету вовсе нет). Таким образом, рассматриваемое нами пространство точек становится 63-мерным.

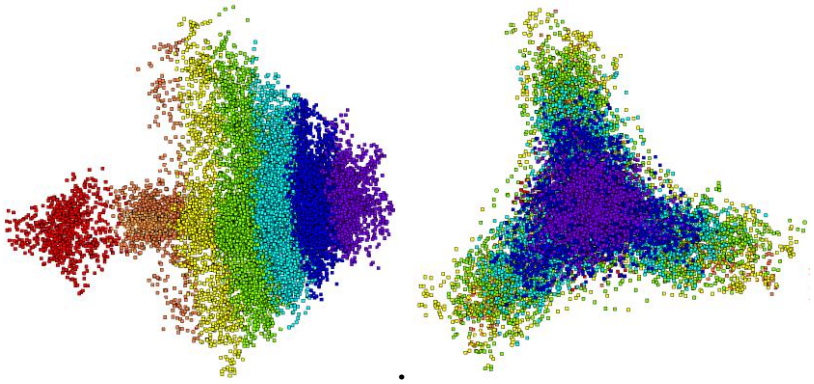
Для выявления структуры в генетической последовательности проводилась предварительная обработка, которая ставила в соответствие данной последовательности множество точек в 63-мерном пространстве триплетов. Делалось это следующим образом: последовательность сканировалась рамкой считывания длины Δ с шагом t. Для каждого положения i рамки определялся участок генетической последовательности, совпадающий с рамкой считывания, для которого вычислялся частотный словарь $W_3^{(i)}$ соответствующий i-ой точке в 63-мерном пространстве. Кроме того, с каждой точкой в 63-мерном пространстве связывались следующие параметры: номер центрального символа рассматриваемого участка и величина GC-состава этого участка.

Данные для исследования брались в базе EMBL-банка. Было отобрано 418 геномов митохондрий растений и животных, в среднем по 25 видов на кладу и 391 геном хлоропластов наземных растений. Для всех генетических последовательностей длина рамки считывания $\Delta = 603$, шаг $t = 11$.

По полученному множеству точек в программе VidaExpert (http://bioinfo-out.curie.fr/projects/vidaexpert/) строился вид данных в пространстве первых трех главных компонент, вычисленных для данного 63-мерного пространства. Рассматривались две проекции на плоскость пространства главных компонент: 1-ой и 2-ой компоненты и 1-ой и 3-ей компоненты или 2-ой и 3-ей компоненты, в зависимости от структуры генома. В обеих проекциях точки окрашивались в соответствии со значениями GC-состава. Окраска выполнялась следующим образом. Строилась гистограмма по величине GC-состава. Все точки в интервале изменения GC-состава делились на семь примерно равных частей пропорционально количеству точек с соответствующими значениями. Точки, распределенные по семи интервалам, раскрашивались в соответствии с цветами радуги. Минимальным значениям GC-состава соответствует фиолетовый цвет, максимальным – красный. Указанные цвета используются для всех рисунков ниже по тексту.

2. Пространственная структура величины *GC*-состава фрагментов геномов хлоропластов. Подавляющее большинство геномов хлоропластов наземных растений имеет

трехлучевую структурут генома [16]. Кроме того, в пространстве первой и второй главных компонент виден отдельный кластер, так называемый «хвост» (рис. 1а слева). На рис 1 а, б показана структура генома хлоропласта с раскрашенным распределением величины *GC* - состава на примере генома *Anthoceros angustus* (идентификатор AB086179 в EMBL). Исключен триплет *gcg*. Было обнаружено, что распределение величины *GC* -состава фрагментов генома по пространственной структуре однотипно. *GC* -состав фрагментов распределен по градиенту вдоль оси симметрии пространственной структуры генома – от меньших значений к большим, причем минимальные значения находятся в вершине трехлучевой структуры, а максимальные – в отдельно расположенном кластере, так называемом «хвосте» (рис. 1, а).



а) б)

Рис. 1. Пространственное расположение *GC* -состава фрагментов генома *Anthoceros angustus* в пространстве 1–2 главных компонент (рис. 1а) и 2–3 главных компонент (рис. 1б)

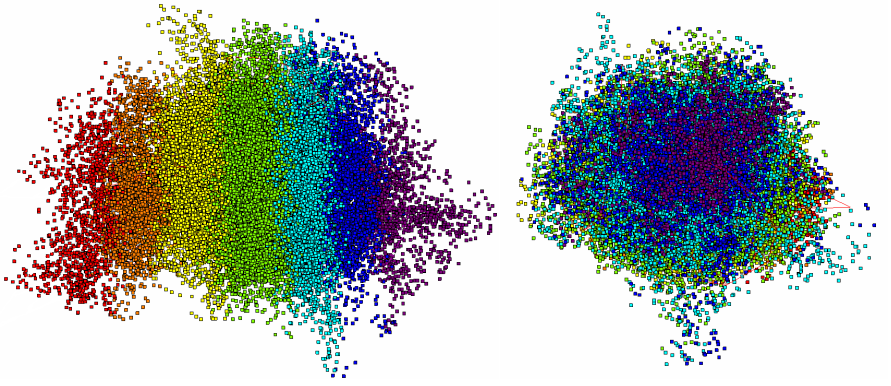
3. Пространственная структура величины *GC* -состава фрагментов геномов митохондрий. Пространственная структура геномов митохондрий имеет преимущественно трехлучевую структуру кроме наземных растений и печеночных мхов, которые имеют структуру близкую к эллипсоиду.

Рассмотрим группы геномов, для которых наблюдается однотипное распределение значений *GC* -состава по пространственной структуре геномов. Рассмотрим сначала геномы митохондрий растений и печеночных мхов, как наиболее отличающихся ото всех остальных геномов митохондрий. Как уже было сказано выше, для наземных растений и печеночных мхов характерна структура геномов митохондрий эллипсоидной формы. Для их геномов характерно ярко выраженное градиентное распределение значений *GC* -состава вдоль оси симметрии структуры. На рис. 2 показано распределение значений *GC* -состава *Zea mays* (идентификатор AB251495 в EMBL, исключен триплет *gcg*). Как видно из рисунка, наблюдается градиентное распределение значений *GC* -состава фрагментов вдоль оси симметрии структуры генома. Аналогично выглядит структура генома и распределение *GC* - состава для печеночных мхов.

Для всех остальных геномов митохондрий характерна трехлучевая структура и далее отдельно это указываться не будет.

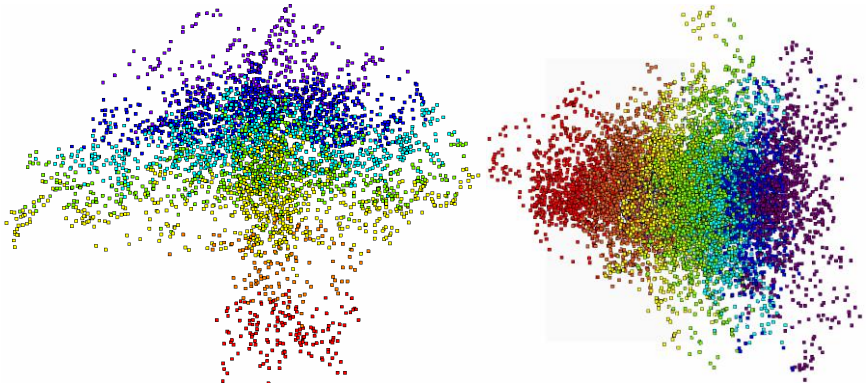
Для одноклеточных водорослей, обычных мхов, мхоподобных лишайников из рода *Cladonia* и высших грибов характерно градиентное распределение значений *GC* -состава фрагментов вдоль оси симметрии структуры генома. На рис. 3 показано распределение значений *GC* -состава фрагментов на примере *Mesostigma viride* (идентификатор AF353999 в EMBL, исключен триплет *ccg*), *Physcomitrella patens* (идентификатор AY506529 в EMBL, исключен триплет *ccg*), *Cladonia petrophila* (идентификатор MG941021 в EMBL, исключен

триплет *ccg*) и *Agaricus bisporus var. bisporus H97* (идентификатор MG941021 в EMBL, исключен триплет *gcg*).

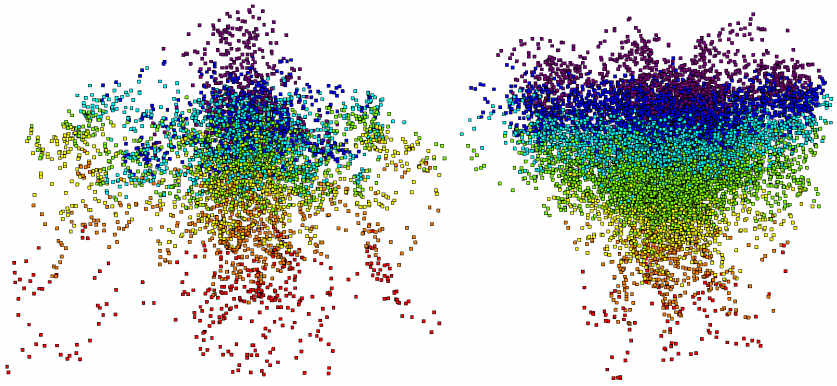


а) б)

Рис. 2. Пространственное расположение *GC* -состава фрагментов генома *Zea mays* в пространстве 1 и 2 главных компонент (рис. 2а) и 2 и 3 главных компонент (рис. 2б)



а) б)

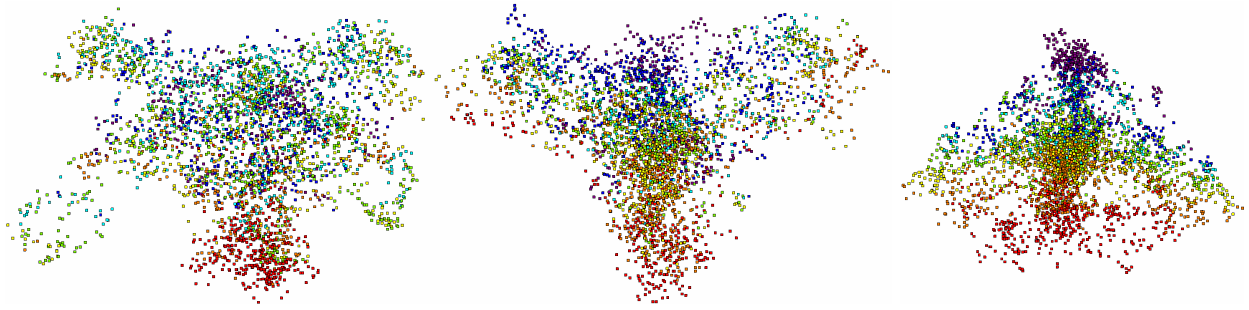


в) г)

Рис. 3. Пространственное расположение *GC* -состава фрагментов генома *Mesostigma viride* в пространстве 1 и 3 главных компонент (рис. 3а), *Physcomitrella patens* в пространстве 2 и 3 главных компонент (рис. 3б), *Cladonia petrophila* в пространстве 1 и 3 главных компонент (рис. 3в) и *Agaricus bisporus var. bisporus H97* в пространстве 1 и 3 главных компонент (рис. 3г)

Многоклеточные водоросли, лишайники и низшие грибы не имеют ярко выраженного градиентного распределения. Тем не менее, можно заметить, что минимальные значения *GC* -состава располагаются большей частью на одном конце оси симметрии структуры, а максимальные на другом. Распределение *GC* -состава фрагментов по структуре геномов показано на рис. 4 на примере *Saccharina japonica* (идентификатор AP011493 в EMBL,

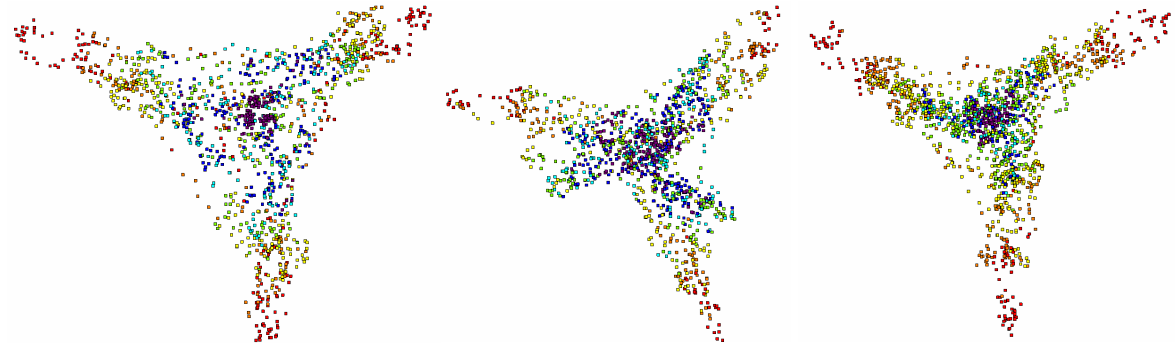
исключен триплет *cgc*), *Hypogymnia vittata* (идентификатор KY362374 в EMBL, исключен триплет *cgc*), *Candida viswanathii* (идентификатор EF536359 в EMBL, исключен триплет *ccg*).



а) б) в)

Рис. 4. Пространственное расположение *GC* -состава фрагментов генома *Saccharina japonica* (рис. 4а), *Hypogymnia vittata* (рис. 4б) и *Candida viswanathii* (рис. 4в) в пространстве 1 и 3 главных компонент

Перейдем к рассмотрению геномов митохондрий животных. Для геномов митохондрий насекомых, паукообразных и ракообразных характерно центральносимметричное распределение значений *GC* -состава фрагментов по пространственной структуре геномов. Минимальные значения *GC* -состава расположены в центре трехлучевой структуры геномов, максимальные по краям. Промежуточные значения не имеют ярко выраженной градации по значениям, но большинство точек предыдущего интервала значений находится ближе к центру, чем большинство точек следующего интервала. На рис. 5 показано центральносимметричное распределение значений *GC* -состава на примере *Homalodisca vitripennis* (идентификатор AY875213 в EMBL, исключен триплет *cgg*), *Liphistius erawan* (идентификатор JQ407803 в EMBL, исключен триплет *cgg*) и *Trigoniophthalmus alternatus* (идентификатор EU016193 в EMBL, исключен триплет *gcg*).



а) б) в)

Рис. 5. Пространственное расположение *GC* -состава фрагментов генома *Homalodisca vitripennis* (рис. 5а), *Liphistius erawan* (рис. 5б) и *Trigoniophthalmus alternatus* (рис. 5в) в пространстве 1 и 2 главных компонент

Следующую группу геномов составляют губки, моллюски, плоские, кольчатые и круглые черви. Для геномов этой группы встречаются распределение значений *GC* -состава в виде неявного градиентного распределения и центральносимметричное. На рис. 6 представлены все варианты на примере *Callyspongia plicifera* (идентификатор EU237477 в EMBL, исключен триплет *cgc*), *Neritina usnea* (идентификатор KU342665 в EMBL, исключен триплет *cgc*) *Urechis caupo* (идентификатор AY619711 в EMBL, исключен триплет *gcg*), *Echinococcus equinus* (идентификатор AF346403 в EMBL, исключен триплет *cgc*).

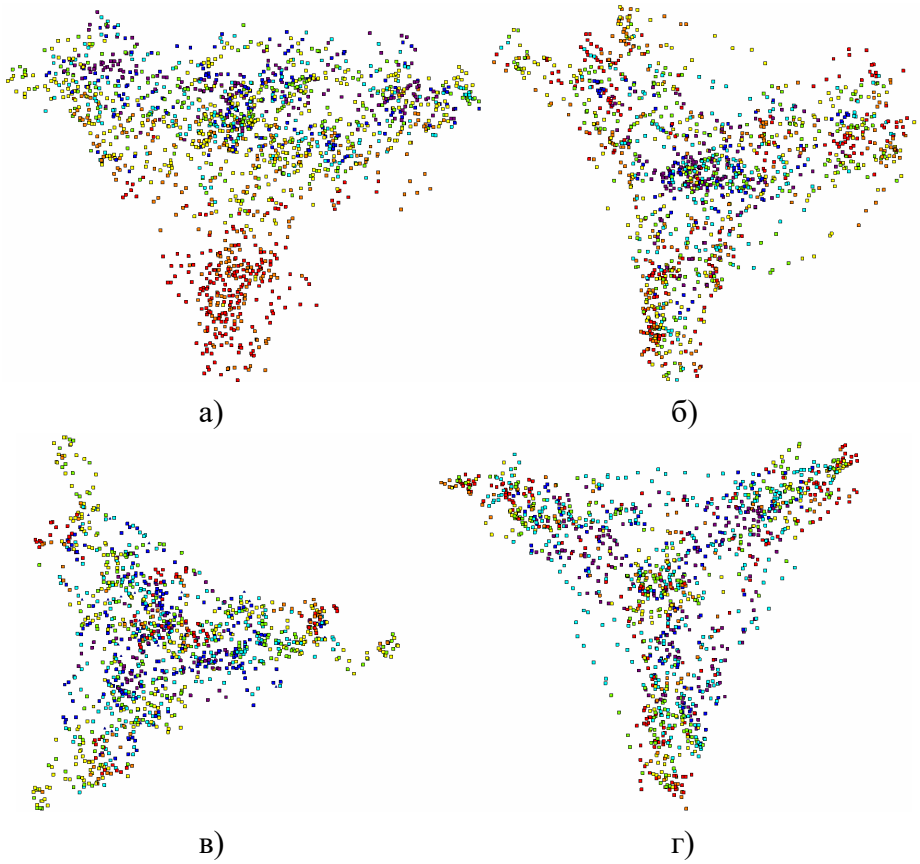


Рис. 6. Пространственное расположение *GC* -состава фрагментов генома *Callyspongia plicifera* (рис. 6а) в пространстве 1 и 3 главных компонент., *Neritina usnea* (рис. 6б), *Urechis caupo* (рис. 6в) и *Echinococcus equinus* (рис. 6г) в пространстве 1 и 2 главных компонент

Для позвоночных животных не удалось выявить какую-либо типичную картину распределения *GC* -состава фрагментов внутри классов и между классами.

Заключение. Результаты, описанные в статье, показывают, что существует упорядоченность в распределении значений *GC* -состава фрагментов различных геномов. Причем эта упорядоченность имеет типичный вид для отдельных групп геномов. Геномы митохондрий обладают большим разнообразием типов распределений *GC* -состава фрагментов. Это обусловлено тем, что митохондрии, рассмотренные в этой работе, принадлежат различным классам организмов с отличающимся строением и функционалом. В то время как геномы хлоропластов достаточно однородны. Обнаруженная упорядоченность – это очень интересный результат, поскольку изначально *GC* -состав описывал физические свойства геномов, но как оказалось, он влияет и на статистические свойства геномов, рассматриваемых как символьные последовательности.

Список источников

1. Shimda H., Sugiuro M. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucleic acids research, 1991, vol. 19, no. 5, pp. 983-995.
2. Young H.A. et al. Chloroplast genome variation in upland and lowland switchgrass. PloS one, 2011, vol. 6, no. 8, pp. e23980.
3. Lockhart P.J., Penny D., Hendy M.D. et al. Controversy on chloroplast origins. FEBS letters, 1992, vol. 301, no. 2, pp. 127-131.
4. Gao L., Yi X., Yang Y.-X. et al. Complete chloroplast genome sequence of a tree fern Alsophila spinulosa: insights into evolutionary changes in fern chloroplast genomes. BMC evolutionary biology, 2009, vol. 9, no. 1, 130 p.
5. Wu Z.Q., Ge S. The phylogeny of the BEP clade in grasses revisited: evidence from the whole-genome sequences of chloroplasts. Molecular phylogenetics and evolution, 2012, vol. 62, no. 1, pp. 573-578.
6. Qian J., Song. J., Gao H. et al. The complete chloroplast genome sequence of the medicinal plant Salvia miltiorrhiza. PloS one, 2013, vol. 8, no. 2, pp. e57607.

7. Zhang T., Fang Y., Wang X. et al. The complete chloroplast and mitochondrial genome sequences of Boea hygrometrica: insights into the evolution of plant organellar genomes. PLoS One, 2012, vol. 7, no. 1, pp. e30531.
8. Yang Y., Zhou T., Duan D. et al. Comparative analysis of the complete chloroplast genomes of five Quercus species. Frontiers in plant science, 2016, vol. 7, 959 p.
9. Behura S.K., Lobo N.F., Haas Br. et al. Complete sequences of mitochondria genomes of Aedes aegypti and Culex quinquefasciatus and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. Insect biochemistry and molecular biology, 2011, vol. 41, no. 10, pp. 770-777.
10. Johnston I.G., Williams B. P. Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. Cell systems, 2016, vol. 2, no. 2, pp. 101-111.
11. Ferla M.P., Thrash J.C., Giovannoni St.J. et al. New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. PLoS One, 2013, vol. 8, no. 12, pp. e83383.
12. Nakamura Y., Sasaki N.V., Kobayashi M. et al. The first symbiont-free genome sequence of marine red alga, Susabi-nori (Pyropia yezoensis). PloS one, 2013, vol. 8, no. 3, pp. e57122.
13. Godel C. et al. The genome of the heartworm, Dirofilaria immitis, reveals drug and vaccine targets. The FASEB Journal, 2012, vol. 26, no. 11, pp. 4650-4661.
14. Imanian B., Pombert J.-F., Dorrell R. et al. Tertiary endosymbiosis in two dinotoms has generated little change in the mitochondrial genomes of their dinoflagellate hosts and diatom endosymbionts. PLoS One, 2012, vol. 7, no. 8, pp. e43763.
15. Wei L. et al. Analysis of codon usage bias of mitochondrial genome in Bombyx mori and its relation to evolution. BMC evolutionary biology, 2014, vol. 14, no. 1, 262 p.
16. Sadovsky M.G., Senashova M.Yu., Malyshev A.V. Amazing symmetrical clustering in chloroplast genomes. BMC Bioinformatics, 2020, 21(Suppl 2):83.

Сенашова Мария Юрьевна. *ИВМ СО РАН, к.ф.-м.н., старший научный сотрудник. Направления исследований: анализ данных, биоинформатика. AuthorID: 133140, SPIN: 1178-6320, ORCID: 0000–0002–1023–7103, msen@icm.krasn.ru, 660036, Красноярск, Академгородок 50, стр.44.*

UDC 57.015 + 573.2
DOI:10.25729/ESI.2024.34.2.003

Orderliness of GC-content values of fragments in the spatial structure of organelle genomes

Maria Yu. Senashova

Institute of computational modelling of the SB RAS,
Russia, Krasnoyarsk, *msen@icm.krasn.ru*

Abstract. The spatial distribution of GC-content values of chloroplast and mitochondrial genome fragments is considered. Spatial distribution refers to the distribution of points corresponding to genome regions in the frequency space of triplets. It was found that the values of the GC-content of fragments for most genomes are distributed not chaotically, but in an orderly manner. Two main types of distribution were found: gradient and centrally symmetric. In chloroplast genomes, only a gradient distribution occurs. In mitochondria genomes, both types of distribution occur. The type of distribution for mitochondria genomes depends on the type of organism. The spatial distribution of the GC-content is stable with respect to changes in the reading window length.

Keywords: order, spatial distribution, triplets, frequency dictionaries, principal components

References

1. Shimda H., Sugiuro M. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucleic acids research, 1991, vol. 19, no. 5, pp. 983-995.
2. Young H.A. et al. Chloroplast genome variation in upland and lowland switchgrass. PloS one, 2011, vol. 6, no. 8, pp. e23980.

3. Lockhart P.J., Penny D., Hendy M.D. et al. Controversy on chloroplast origins. FEBS letters, 1992, vol. 301, no. 2, pp. 127-131.
4. Gao L., Yi X., Yang Y.-X. et al. Complete chloroplast genome sequence of a tree fern Alsophila spinulosa: insights into evolutionary changes in fern chloroplast genomes. BMC evolutionary biology, 2009, vol. 9, no. 1, 130 p.
5. Wu Z.Q., Ge S. The phylogeny of the BEP clade in grasses revisited: evidence from the whole-genome sequences of chloroplasts. Molecular phylogenetics and evolution, 2012, vol. 62, no. 1, pp. 573-578.
6. Qian J., Song. J., Gao H. et al. The complete chloroplast genome sequence of the medicinal plant Salvia miltiorrhiza. PloS one, 2013, vol. 8, no. 2, pp. e57607.
7. Zhang T., Fang Y., Wang X. et al. The complete chloroplast and mitochondrial genome sequences of Boea hygrometrica: insights into the evolution of plant organellar genomes. PLoS One, 2012, vol. 7, no. 1, pp. e30531.
8. Yang Y., Zhou T., Duan D. et al. Comparative analysis of the complete chloroplast genomes of five Quercus species. Frontiers in plant science, 2016, vol. 7, 959 p.
9. Behura S.K., Lobo N.F., Haas Br. et al. Complete sequences of mitochondria genomes of Aedes aegypti and Culex quinquefasciatus and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. Insect biochemistry and molecular biology, 2011, vol. 41, no. 10, pp. 770-777.
10. Johnston I.G., Williams B. P. Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. Cell systems, 2016, vol. 2, no. 2, pp. 101-111.
11. Ferla M.P., Thrash J.C., Giovannoni St.J. et al. New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. PLoS One, 2013, vol. 8, no. 12, pp. e83383.
12. Nakamura Y., Sasaki N.V., Kobayashi M. et al. The first symbiont-free genome sequence of marine red alga, Susabi-nori (Pyropia yezoensis). PloS one, 2013, vol. 8, no. 3, pp. e57122.
13. Godel C. et al. The genome of the heartworm, Dirofilaria immitis, reveals drug and vaccine targets. The FASEB Journal, 2012, vol. 26, no. 11, pp. 4650-4661.
14. Imanian B., Pombert J.-F., Dorrell R. et al. Tertiary endosymbiosis in two dinotoms has generated little change in the mitochondrial genomes of their dinoflagellate hosts and diatom endosymbionts. PLoS One, 2012, vol. 7, no. 8, pp. e43763.
15. Wei L. et al. Analysis of codon usage bias of mitochondrial genome in Bombyx mori and its relation to evolution. BMC evolutionary biology, 2014, vol. 14, no. 1, 262 p.
16. Sadovsky M.G., Senashova M.Yu., Malyshev A.V. Amazing symmetrical clustering in chloroplast genomes. BMC Bioinformatics, 2020, 21(Suppl 2):83.

Senashova Maria Yurievna. *ICM SB RAS, Ph.D., senior researcher. Research direction: data analysis, bioinformatics. AuthorID: 133140, SPIN: 1178-6320, ORCID: 0000–0002–1023–7103, msen@icm.krasn.ru, 660036, Krasnoyarsk, Akademgorodok 50, building 44.*

Статья поступила в редакцию 08.12.2023; одобрена после рецензирования 30.05.2024; принята к публикации 03.06.2024.
The article was submitted 12/08/2023; approved after reviewing 05/30/2024; accepted for publication 06/03/2024.

Методы, технологии и приложения искусственного интеллекта

УДК 004.27+621.382
DOI:10.25729/ESI.2024.34.2.004

Аппаратная реализация сегментной спайковой модели нейрона CSNM на операционных усилителях

Бойко Александр Васильевич[1], Бахшиев Александр Валерьевич[1,2], Корсаков Антон Михайлович[1,3]

[1]Санкт-Петербургский политехнический университет Петра Великого,

[2]ООО Системы Компьютерного зрения,

[3]Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики,

Россия, Санкт-Петербург, *bakhshiev_av@spbstu.ru*

Аннотация. В работе предложен вариант аппаратной реализации сегментной спайковой модели нейрона на операционных усилителях. Актуальность работы обусловлена растущей потребностью как в аппаратных реализациях нейросетевых решений в целом, так и необходимостью развития адаптивных способностей сетей, в первую очередь, к изменяющимся условиям среды. Одним из перспективных направлений представляется реализация спайковых нейронных сетей, в которых основным функциональным элементом является не нейрон, а сегмент мембраны нейрона. Аппаратная реализация таких моделей нейрона на дискретной элементной базе должна позволить облегчить экспериментальные исследования данного направления. В основу предложенного решения положена модель нейрона CSNM. В работе рассмотрены существующие подходы к аппаратной реализации моделей нейронов и выбран подход реализации на операционных усилителях. Разработаны схемы каждого сегмента реализуемой модели нейрона. Проведены тестовые эксперименты и сравнение с математической моделью, результаты которых позволили заявить, что реализация достаточно точно воспроизводит требуемые временные характеристики процессов преобразования сигналов в нейроне. Предложенная реализация позволяет гибко менять структуру дендритного и синаптического аппарата нейрона и удобно интерпретировать сигналы для сопоставления с математической моделью. Недостатком предложенного решения является низкая энергоэффективность, однако для исследовательских целей этот аспект на данном этапе не является критичным.

Ключевые слова: спайковые нейронные сети, глубокое обучение, нейроморфные системы, спайковый нейрон, сегментная модель нейрона, машинное обучение, операционный усилитель, аппаратная реализация

Цитирование: Бойко А.В. Аппаратная реализация сегментной спайковой модели нейрона CSNM на операционных усилителях / А.В. Бойко, А.В. Бахшиев, А.М. Корсаков // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 41-49. – DOI: 10.25729/ESI.2024.34.2.004.

Введение. В современном мире все большее применение находят технологии, основанные на нейронных сетях. Направление спайковых нейронных сетей является важной частью современной нейронауки. Это обусловлено как потребностью в создании энергоэффективных реализацией глубоких нейронных сетей [1], так и поиском новых, более эффективных архитектурных решений. Одним из направлений таких исследований являются попытки реализации вычислений на дендритных деревьях отдельных нейронов, с далеко идущей целью создания архитектур, где функциональным элементом выступает не отдельный нейрон как точечный объект, а сегмент (компартмент) дендритного дерева.

Нейробиологические исследования показали, что дендриты нейронов играют важную роль в интеграции синаптических входов, что может увеличить способность нейронов обрабатывать информацию [2, 3]. Применение модели нейрона с учетом описания структуры дендритного дерева предложено в [4], однако модель не рассматривает сигналы в сети как

спайковые. Модели, учитывающие структуру нейрона, называют сегментными моделями (compartmental). На данный момент можно выделить двухсегментную модель [5], состоящую из модулей сомы и дендритов, и трехсегментную модель, состоящую из модулей сомы, апикальных и базальных дендритов [6]. В работах [7, 8] показывается, что многосегментные модели нейрона, реализованные на нейроморфных чипах, позволяют имитировать динамические процессы нейронных цепей мозга при низких энергозатратах. В работе [9] предлагается алгоритм обучения спайковой нейронной сети, использующей многосегментные нейроны для обработки данных о состоянии системы.

В статье рассматривается аппаратная реализация модели нейрона CSNM, представленная в работах [10, 11, 12]. Модель CSNM позволяет описывать нейрон как древовидную структуру, состоящую из отдельных сегментов, каждый из которых может содержать произвольное число возбуждающих и тормозных синапсов. Особенностью модели является простота реализации на ней нейронов с различной конфигурацией дендритного и синаптического аппаратов с сохранением возможности описания динамических процессов преобразования импульсных потоков в нейроне. Целью работы было реализовать качественно подобное поведение аппаратной реализации на дискретных элементах с возможностью в будущем формировать требуемую структуру мембраны вручную.

В разделе 1 представлен краткий обзор подходов к реализации спайковых моделей нейронов на аналоговой электронике, в разделе 2 описано предлагаемое решение, в разделе 3 демонстрируются результаты симуляции.

1. Обзор подходов к аппаратной реализации моделей нейронов на аналоговой элементной базе. Большинство схем используют транзисторные каскады для реализации уравнений модели [13, 14]. Транзисторы не потребляют много энергии, работают при низких напряжениях, имеют малые габариты и довольно высокий КПД, но при этом чувствительны к температуре и статическому электричеству. Эти недостатки устраняют с помощью обратных связей, в результате чего схемы становятся довольно сложными для восприятия и расчета. Кроме того, схемы на транзисторах чувствительны к параметрам элементов в нем, особенно к качеству транзисторов. Если схема симметрична, то практически обязательным требованием является использование транзисторов из одной партии. При сборке схем на дискретной элементной базе это необходимо учитывать.

В некоторых реализациях кроме транзисторных каскадов используются и операционные усилители (ОУ), чаще – в схемах для обучения нейросетей. ОУ обычно выполняет вспомогательную функцию, например, по объединению нейронов определенным образом или созданию положительной обратной связи. Иногда ОУ используется в качестве компаратора с гистерезисом. В некоторых реализациях ОУ играет главную роль, например, в [15] ОУ используется в качестве интегратора в синапсе. В работе [14] предложен синапс с использованием ОУ для сохранения бистабильности.

Отдельно стоит выделить аппаратную реализацию нейрона Neurogrid [16]. Дифференциальные уравнения модели реализованы с помощью управляемой проводимости, интеграторов на емкости и управляемых источников тока. В основе элементной базы управляемых элементов лежат транзисторы. Главным отличием этой реализации от рассмотренных ранее является ее наибольшая приближенность к биологическому нейрону: наличие "популяции" синапсов, а также отдельно выраженных ионных каналов, дендрита и сомы.

Операционные усилители не пользуются большой популярностью при реализации спайковых нейронных сетей, так как математические модели обычно описывают поведение тока внутри нейрона, а операционные усилители преобразуют именно напряжение.

Исключением является синапс, который, как в модели, так и в реализациях, оперирует входным напряжением, но выходом все равно является синаптический ток.

Схемы на операционных усилителях отличаются постоянством и надежностью характеристик. Параметры такой схемы практически полностью зависят от внешних элементов, а не самого ОУ, поэтому тщательный подбор элементов здесь не нужен. Существуют готовые схемы на ОУ для типовых операций и звеньев теории автоматического управления, которые практически не требуют дополнительной настройки, а также, при каскадном подключении схемы на ОУ, обладают меньшим шумом по сравнению с каскадным подключением транзисторных схем. В случае с сегментной реализацией очень важна повторяемость параметров одинаковых сегментов, что трудно ожидать от транзисторных схем на дискретных элементах.

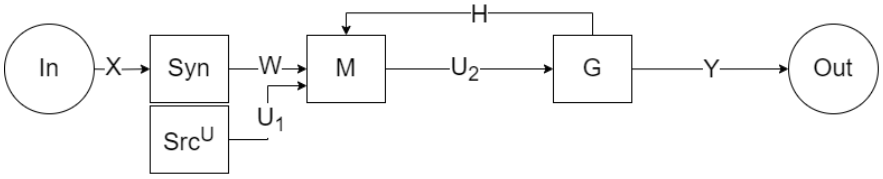
Относительно недавно в реализациях появился новый подход – использование мемристоров. Мемристор – это элемент, который сохраняет определенное значение своего сопротивления в зависимости от проходившего через него тока. В реализациях он обычно используется для хранения веса нейрона, что облегчает обучение нейросети.

В одной из аналоговых реализаций мемристор используется как генератор.

Мемристоры пока являются довольно новыми компонентами, и еще не настолько хорошо изучены и описаны. Они имеют несколько недостатков: от одного цикла переключения к другому они слегка меняют свои характеристики, и, так же, как и транзисторы, обладают вариацией характеристик от одного устройства к другому.

2. Аппаратная реализация модели CSNM на операционных усилителях. Рассматриваемая модель нацелена на реализацию пластичности нейрона через изменение структуры мембраны и числа синапсов, поэтому для сохранения всех функций модели нет необходимости обеспечивать изменение веса синапса. Также, поскольку планировалось использование дискретных элементов, в основу были положены операционные усилители. Вопросы оптимизации энергопотребления и сложности схемотехнических решений в рамках данной работы не рассматривались.

Аппаратная реализация делится на следующие модули аналогично модели CSNM: синапс, сегмент мембраны и генератор импульсов. Далее эти структурные элементы можно комбинировать, формируя нейрон с различной организацией дендритного и синаптического аппарата. Для примера, на рисунке 1 изображена схема простейшего нейрона с одним синапсом и мембраной.



X,Y – входной и выходной сигналы соответственно (прямоугольный импульс)

U1, U2 – потенциал на входе и выходе модели сегмента мембраны

W – вклад синапса в изменение потенциала мембраны

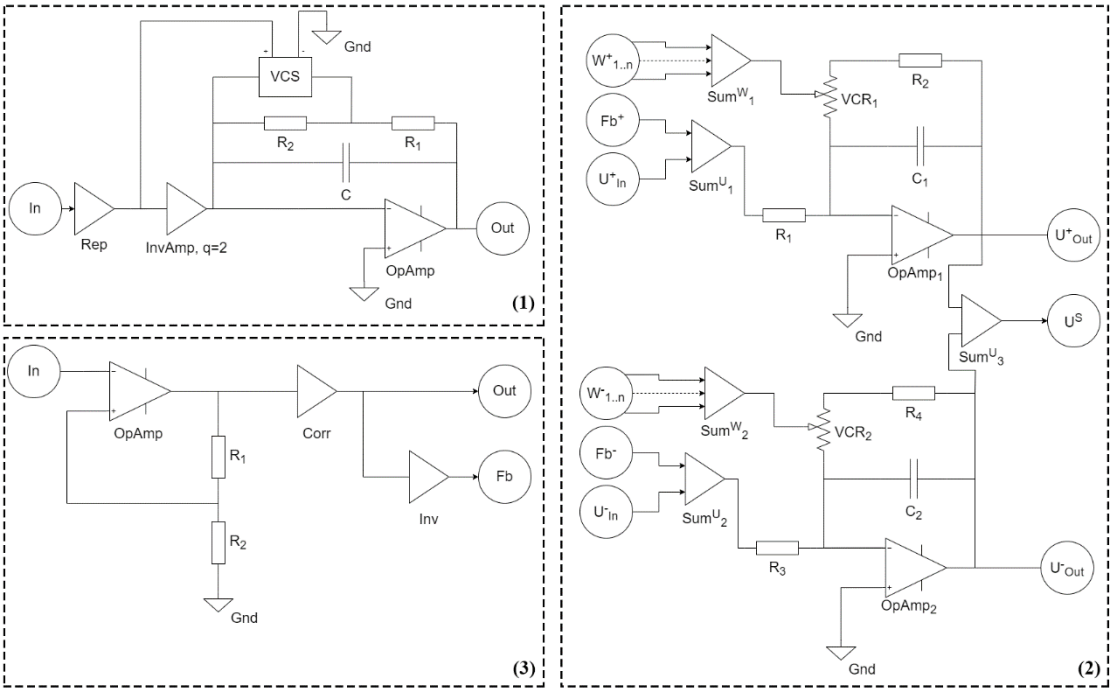
H – влияние генератора на изменение потенциала мембраны (обратная связь)

Рис. 1. Структурная схема аппаратной реализации нейрона

На рисунке 2 приведены схемы функциональных элементов модели. Использование напряжения как носителя сигнала вытекает из факта использования ОУ.

Здесь: In – вход, Out – выход, Rep – повторитель, InvAmp – инвертирующий усилитель, q – коэффициент усиления, VCS – управляемый напряжением ключ, R – сопротивление, C – емкость, OpAmp – операционный усилитель, Gnd – земля, $W_{1..n}$ – сигнал с синапса, Fb –

обратная связь, U_{In} – входное напряжение, Sum^{W} – сумматор сигналов с синапса, Sum^{U}– сумматор напряжений, VCR – управляемое напряжением сопротивление, U_{0ut} – выходное напряжение, U_S – суммарное напряжение, Corr – корректировщик, Inv – инвертор.



(1) – Синапс

(2) – сегмент мембраны

(3) – генератор импульсов

Рис. 2. Схемы функциональных элементов модели

В модели CSNM синапс представлен апериодическим звеном 1-го порядка с логическим управлением постоянной времени. На вход синапса подается положительный прямоугольный импульс длительностью 1 мс и амплитудой 5В. Так как на выходе мы хотим получить положительный отклик синапса, а ОУ инвертирует сигнал, мы предварительно инвертируем импульс с помощью инвертирующего усилителя с коэффициентом усиления 1.

Синапс должен обладать разными постоянными времени во время своего роста (когда импульс активен) и своего спада (когда импульс отсутствует). Постоянная времени апериодического звена на ОУ зависит от сопротивления $R_1\mathrm{R}_1$ и емкости $C\mathrm{C}$ обратной связи ($T = R_1C \quad \mathrm{T} = \mathrm{R}_1\mathrm{C}$), поэтому, чтобы менять постоянную времени, нужно менять сопротивление или емкость. Добавление емкости вызывает «просадки» сигнала при зарядке конденсатора, поэтому было принято решение управлять сопротивлением. Для этого в участок обратной связи был добавлен ключ, управляемый напряжением (VCS), с параллельным ему добавочным сопротивлением $R_2\mathrm{R}_2$. На затвор подается импульс с генератора, таким образом, когда импульс есть – VCS замыкается и шунтирует сопротивление $R_2\mathrm{R}_2$ ($T_{\text{воз}} = R_1C$), когда импульс отсутствует – VCS размыкается и сопротивление $R_2\mathrm{R}_2$ участвует в цепи ($T_{спад} = [R_1 + R_2]C$). Сопротивления были подобраны так, чтобы временные постоянные были приближены к таковым в модели.

Так как напряжение импульса делится на две ветви и уменьшается, коэффициент инвертирующего усилителя был изменен на 2.

Также выход генератора импульсов является выходом операционного усилителя (т.е. низкоомный выход). При подключении следующего нейрона сигнал с него должен идти на

вход другого ОУ в синапсе (т.е. высокоомная нагрузка), и для избежания помех сигнала в синапсе применен повторитель входного сигнала.

Функциональная схема сегмента мембраны представлена на рисунке 2(2). Реализация сегмента состоит из двух частей (ионных механизмов), обеспечивающих условно положительный и отрицательный вклады в потенциал. Разделение на полярность считается условным, потому что играет роль только знак при сложении значений, само напряжение в каждой части положительное. Каждый ионных механизм, по сути, представляет собой апериодическое звено первого порядка с обратной связью. Сигнал синапса участвует в обратной связи и в постоянной времени апериодического звена. Если пересчитать этот элемент с учетом обратной связи, то так же, как для синапса, можно получить интегратор со следующим выражением для выходной характеристики:

$$U_{Out} = \frac{C_m p + g^2 R_m + g}{(g + 1/R_m) p} U_{In},$$

где C_m - емкость мембраны,

R_m - сопротивление мембраны,

g - суммарная текущая проводимость синапсов.

Предложенная реализация не обеспечивает точного соответствия, но решение было подобрано так, чтобы качественно повторять поведение математической модели.

Для реализации изменяемой временной характеристики было использовано сопротивление, управляемое напряжением (управляемое сопротивление). На данный момент в схеме оно представлено виртуальным элементом.

Входными сигналами для сегмента мембраны являются два входа для потенциалов (в случае конечного сегмента дендрита в цепочке – это потенциал с источника напряжения, а в случае последующих – это вклады с предыдущих сегментов), два входа для сигналов с синапсов (тормозящий и возбуждающий) и два входа для обратной связи (ОС) на каждый из потенциалов. Выходными сигналами являются два выхода для потенциала каждого знака и выход для суммы потенциалов. Разделение на полярность считается условным, потому что играет роль только знак при сложении значений, само напряжение в каждой части положительное.

На входе для сигналов с синапса стоят инвертирующие сумматоры, которые позволяют подключать на вход мембраны несколько синапсов с разным весом.

Принцип работы генератора выходных импульсов таков: сигнал обратной связи и выходной импульс начинают генерироваться, когда сигнал на входе выше определенного уровня (P_1), а заканчивает генерироваться, когда ниже другого уровня (P_2, $P_2 < P_1$). Такому поведению полностью соответствует компаратор с гистерезисом на операционном усилителе (ОУ). Функциональная схема генератора импульсов представлена на рисунке 2(3).

Параметры компаратора (значения порогов) задаются через отношения сопротивлений R_1 и R_2, R_1иR_2и максимального и минимального напряжений на ОУ. Компаратор выдает свой отрицательный нижний порог, поэтому он «сдвигается» до нуля суммированием результата компаратора с дополнительным напряжением, противоположным нижнему порогу. Этот элемент схемы представлен как «Corr». Далее сигнал с соответствующим знаком идет на выход и на обратную связь (ОС).

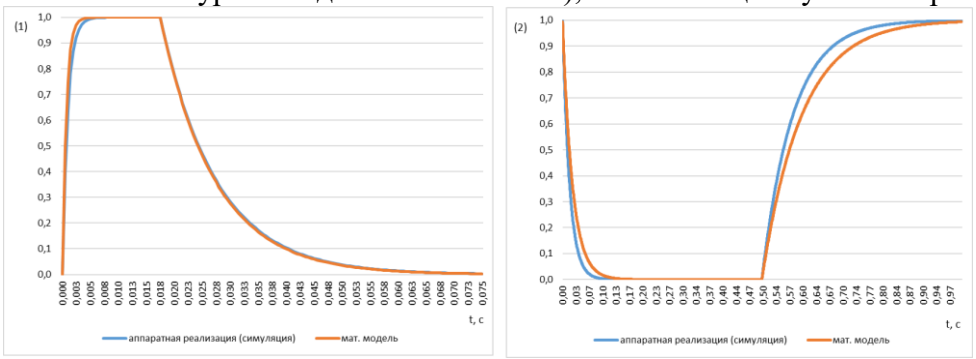
3. Экспериментальное исследование. Для того, чтобы сравнить полученную аппаратную реализацию, был поставлен следующий эксперимент. В качестве входа на синапс и ионный механизм сегмента мембраны подавались одинаковые сигналы: импульс

(прямоугольный сигнал длительностью 1 мс) и ступенька (прямоугольный сигнал длительностью в 1 с). Временем возрастания считалось время, когда сигнал изменяется от 5% до 95% расстояния между экстремумами, время спада, соответственно, наоборот. Результаты экспериментов представлены в таблице 1.

Таблица 1. Время возрастания и спада выходного сигнала реализации синапса и ионного механизма сегмента мембраны, мс

	Синапс						Сегмент мембраны					
	Модель		Аппаратная симуляция		Δ		Модель		Аппаратная симуляция		Δ	
	$T_{воз}$	$T_{спад}$	$T_{воз}$	$T_{спад}$	$T_{воз}$	$T_{спад}$	$T_{воз}$	$T_{спад}$	$T_{воз}$	$T_{спад}$	$T_{воз}$	$T_{спад}$
Импульс	1	29	0,8	29,2	0,2	0,2	102	1,5	97	1,5	7	0
Ступенька	2	28,5	2,9	29,2	0,9	0,7	294	78	228,7	55,3	63,3	22,7

Для наглядности результаты были нормированы и представлены на рисунке 3 (продолжительные плато урезаны для компактности), по оси абсцисс указано время (сек).



(1) – Реакции синапса

(2) – Реакции ионного механизма сегмента мембраны

Рис. 3. Реакции синапса и ионного механизма сегмента мембраны на ступенчатое входное воздействие

Можно видеть, что качественно аппаратная реализация повторяет поведение математической модели. Рассогласование обусловлено спецификой аппаратной реализации, а также некоторыми допущениями в реализации сегмента мембраны. Тем не менее, результаты качественно соответствуют друг другу. Полученная количественная погрешность в длительности переходных процессов может быть компенсирована эмпирической коррекцией параметров модели.

Заключение. Был проведен анализ существующих подходов к аппаратной реализации спайковых нейронов на дискретных аналоговых элементах. Разработаны схемы реализаций функциональных элементов модели CNSM, проведена эмпирическая настройка полученной схемы под типовые параметры математической модели.

Преимуществом предложенного решения является его модульность, то есть возможность собирать произвольные конфигурации структуры мембраны нейрона.

Для оценки качества полученного решения проведены эксперименты со сравнением отклика синапса и сегмента мембраны на типовые входные воздействия. Эксперименты показали, что реализация повторяет временные характеристики, которые наблюдаются в математической модели, что позволит использовать ее для аппаратной реализации предобученных спайковых нейронных сетей.

Из недостатков на данном этапе можно выделить сложность описания схемы, недостаточно эффективное энергопотребление, неопределенность в физической реализации некоторых элементов схемы, необходимость эмпирической подстройки временных параметров схемы.

В качестве перспективных направлений работы можно выделить следующие: необходимость разработки метода сравнения амплитуд в математической модели с амплитудами в её аппаратной реализации, проведение дополнительных экспериментов с отдельными нейронами и цепочками нейронов, оптимизация схемы и сборка прототипов.

Благодарности. Исследование выполнено за счет гранта Российского научного фонда № 23–21-00287, https://rscf.ru/project/23-21-00287.

Список источников

1. Bouvier M., Valentian A., Mesquida T. et al. Spiking neural networks hardware implementations and challenges: A survey. ACM Journal on Emerging Technologies in Computing Systems (JETC), 2019, no. 15(2), pp. 1-35.
2. Smith S. L. et al. Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo. Nature, 2013, vol. 503, no. 7474, pp. 115-120.
3. Gidon A. et al. Dendritic action potentials and computation in human layer 2/3 cortical neurons. Science, 2020, vol. 367, no. 6473, pp. 83-87.
4. Khan G.M., Miller J.F. In search of intelligence: evolving a developmental neuron capable of learning. Connection Science, 2014, vol. 26, no. 4, pp. 297-333.
5. Urbanczik R., Senn W. Learning by the dendritic prediction of somatic spiking. Neuron, 2014, vol. 81, no. 3, pp. 521-528.
6. Sacramento J. et al. Dendritic cortical microcircuits approximate the backpropagation algorithm. Advances in neural information processing systems, 2018, vol. 31.
7. Shrestha A. et al. In-hardware learning of multilayer spiking neural networks on a neuro-morphic processor. 2021 58th ACM/IEEE Design Automation Conference (DAC), IEEE, 2021, pp. 367-372.
8. Kopsick J. D. et al. Robust resting-state dynamics in a large-scale spiking neural network model of area ca3 in the mouse hippocampus. Cognitive computation, 2022, pp. 1-21.
9. Sun Y. et al. Multi-compartment neuron and population encoding improved spiking neural network for deep distributional reinforcement learning. arXiv preprint arXiv:2301.07275, 2023.
10. Бахшиев А.В. Математическое моделирование процессов преобразования импульсных потоков в естественном нейроне / А.В. Бахшиев, С.П. Романов // Нейрокомпьютеры: разработка, применение, 2009. – №3 – С.71-80.
11. Bakhshiev A., Gundelakh F. Mathematical model of the impulses trans-formation processes in natural neurons for biologically inspired control systems development. Supplementary Proceedings of the 4th International conference on analysis of images, social networks and texts (AIST-SUP 2015), Yekaterinburg, Russia, April 9-11, 2015. Published on CEUR-WS, vol. 1452, pp. 1-12, 15.10.2015, available at: http://ceur-ws.org/Vol-1452/
12. Бахшиев А.В. Сегментная спайковая модель нейрона CSNM / А.В. Бахшиев, А.А. Демчева // Известия высших учебных заведений. Прикладная нелинейная динамика, 2022. – Т. 30. – № 3. – С. 299-310. – DOI:10.18500/0869-6632-2022-30-3-299-310.
13. Chicca E., Stefanini F., Bartolozzi C., Indiveri G. Neuromorphic Electronic circuits for building autonomous cognitive systems. Proceedings of the IEEE, 2014, vol. 102, no. 9, pp. 1367-1388.
14. Azghadi M.R., Iannella N., Al-Sarawi S.F., Indiveri G., Abbott D. Spike-based synaptic plasticity in silicon: design, implementation, application, and challenges. Proceedings of the IEEE, 2014, vol. 102, no. 5, pp. 717-737.
15. Millner S., Grubl A., Meier K., Schemmel J., Schwartz M.-O. A VLSI implementation of the adaptive exponential. Advances in neural information processing systems, 2010, vol. 23, pp. 1642-1650.
16. Zohora F.T., Debnath S., Rashid A.H.-u, Memristor-CMOS hybrid implementa-tion of leaky integrate and fire neuron model. International Conference on electrical, computer and communication engineering (ECCE), Cox'sBazar, Bangladesh, 2019.

Бойко Александр Васильевич. *Студент СПбПУ. Направление обучения: мехатроника и робототехника, автономные роботы. ORCID: 0009-0009-7230-3346, boyko02av@mail.ru, 195251, Санкт-Петербург, ул. Политехническая, д. 29.*

Бахшиев Александр Валерьевич. *К.т.н., СПбПУ, доцент. Направления исследований: искусственные нейронные сети, нейроморфные системы, компьютерное зрение. AuthorID: 172135, SPIN: 7055-8676, ORCID: 0000-0002-1284-0088, palexab@gmail.com, 195251, Санкт-Петербург, ул. Политехническая, д. 29.*

Корсаков Антон Михайлович. *Старший научный сотрудник ЦНИИ РТК. Направления исследований: искусственные нейронные сети, нейроморфные системы, компьютерное зрение. AuthorID: 172137, SPIN: 3736-3401, anton_korsakov@mail.ru, 194064, Санкт-Петербург, Тихорецкий пр., д. 21.*

UDC 004.27+621.382
DOI:10.25729/ESI.2024.34.2.004

Operational-amplifier-based hardware implementation of the compartmental spike model of the CSNM neuron

Alexander V. Boyko[1], Alexander V. Bakhshiev[1,2], Anton M. Korsakov[1,3]

[1]Peter the Great St. Petersburg Polytechnic University,

[2]Computer Vision Systems LLC, St. Petersburg,

[3]Russian state scientific center for robotics and technical cybernetics (RTC),

Russia, St. Petersburg, *bakhshiev_av@spbstu.ru*

Abstract. In this paper, a variant of the compartmental spiking neuron model hardware implementation on operational amplifiers is proposed. The relevance of the work is due to the growing need for both hardware implementations of neural network solutions in general and the need to develop adaptive abilities of networks, primarily to changing environmental conditions. One of the promising directions is the implementation of spike neural networks, in which the main functional element is not a neuron, but a compartment of the neuron membrane. The hardware implementation of such neuron models on a discrete element base should make it possible to facilitate experimental research in this area. The proposed solution is based on the CSNM neuron model. The paper considers the existing approaches to the hardware implementation of neuron models and selects the implementation approach on operational amplifiers. Schematics of each compartment of the implemented neuron model have been developed. Test experiments and comparison with a mathematical model were carried out, the results of which allowed us to state that the implementation reproduces the required time characteristics of signal conversion processes in a neuron quite accurately. The proposed implementation makes it possible to flexibly change the structure of the dendritic and synaptic apparatus of a neuron, and it is convenient to interpret signals for comparison with a mathematical model. The disadvantage of the proposed solution is low energy efficiency, however, for research purposes, this aspect is not critical at this stage.

Keywords: spike neural networks, deep learning, neuromorphic systems, spike neuron, compartmental neuron model, machine learning, operational amplifier, hardware implementation

Acknowlegements: The research was carried out at the expense of a grant from the Russian Science Foundation No. 23-21-00287, https://rscf.ru/project/23-21-00287.

References

1. Bouvier M., Valentian A., Mesquida T. et al. Spiking neural networks hardware implementations and challenges: A survey. ACM Journal on Emerging Technologies in Computing Systems (JETC), 2019, no. 15(2), pp. 1-35.
2. Smith S. L. et al. Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo. Nature, 2013, vol. 503, no. 7474, pp. 115-120.
3. Gidon A. et al. Dendritic action potentials and computation in human layer 2/3 cortical neurons. Science, 2020, vol. 367, no. 6473, pp. 83-87.
4. Khan G.M., Miller J.F. In search of intelligence: evolving a developmental neuron capable of learning. Connection Science, 2014, vol. 26, no. 4, pp. 297-333.
5. Urbanczik R., Senn W. Learning by the dendritic prediction of somatic spiking. Neuron, 2014, vol. 81, no. 3, pp. 521-528.
6. Sacramento J. et al. Dendritic cortical microcircuits approximate the backpropagation algorithm. Advances in neural information processing systems, 2018, vol. 31.
7. Shrestha A. et al. In-hardware learning of multilayer spiking neural networks on a neuro-morphic processor. 2021 58th ACM/IEEE Design Automation Conference (DAC), IEEE, 2021, pp. 367-372.
8. Kopsick J. D. et al. Robust resting-state dynamics in a large-scale spiking neural network model of area ca3 in the mouse hippocampus. Cognitive computation, 2022, pp. 1-21.
9. Sun Y. et al. Multi-compartment neuron and population encoding improved spiking neural network for deep distributional reinforcement learning. arXiv preprint arXiv:2301.07275, 2023.
10. Bakhshiev A.V., Romanov S.P. Matematicheskoye modelirovaniye protsessov preobra-zovaniya impul'snykh potokov v yestestvennom neyrone [Mathematical modeling of the pulse flows conversion processes in a natural neuron]. Neyrokomp'yutery. Razrabotka, primeneniye [Neurocomputers: development, application], 2009, vol. 3, pp.71-80.

11. Bakhshiev A., Gundelakh F. Mathematical model of the impulses trans-formation processes in natural neurons for biologically inspired control systems development. Supplementary Proceedings of the 4th International conference on analysis of images, social networks and texts (AIST-SUP 2015), Yekaterinburg, Russia, April 9-11, 2015. Published on CEUR-WS, vol. 1452, pp. 1-12, 15.10.2015, available at: http://ceur-ws.org/Vol-1452/.
12. Bakhshiev A.V., Demcheva A.A. Segmentnaya spaykovaya model' neyrona CSNM [Compartmental spiking neuron model CSNM]. Izvestiya vysshikh uchebnykh zavedeniy. Prikladnaya nelineynaya dinamika [Izvestiya VUZ. Applied nonlinear dynamics], 2022, vol. 30, no. 3, pp. 299-310, DOI:10.18500/0869-6632-2022-30-3-299-310.
13. Chicca E., Stefanini F., Bartolozzi C., Indiveri G. Neuromorphic Electronic circuits for building autonomous cognitive systems. Proceedings of the IEEE, 2014, vol. 102, no. 9, pp. 1367-1388.
14. Azghadi M.R., Iannella N., Al-Sarawi S.F., Indiveri G., Abbott D. Spike-based synaptic plasticity in silicon: design, implementation, application, and challenges. Proceedings of the IEEE, 2014, vol. 102, no. 5, pp. 717-737.
15. Millner S., Grubl A., Meier K., Schemmel J., Schwartz M.-O. A VLSI implementation of the adaptive exponential. Advances in neural information processing systems, 2010, vol. 23, pp. 1642-1650.
16. Zohora F.T., Debnath S., Rashid A.H.-u, Memristor-CMOS hybrid implementa-tion of leaky integrate and fire neuron model. International Conference on electrical, computer and communication engineering (ECCE), Cox'sBazar, Bangladesh, 2019.

Boyko Alexander Vasilyevich. *SPbPU, undergraduate. Study direction: mechatronics and robotics, autonomous robots. ORCID: 0009-0009-7230-3346, boyko02av@mail.ru, 195251, Russia, St.Petersburg, Polytechnicheskaya, 29.*

Bakhshiev Alexander Valeryevich. *Ph.D., SPbPU, docent. Research direction: artificial neural networks, neuromorphic systems, computer vision. AuthorID: 172135, SPIN: 7055-8676, ORCID: 0000-0002-1284-0088, palexab@gmail.com, 195251, Russia, St.Petersburg, Polytechnicheskaya, 29.*

Korsakov Anton Mikhailovich. *RTC, senior researcher. Research direction: artificial neural networks, neuromorphic systems, computer vision. AuthorID: 172137, SPIN: 3736-3401, anton_korsakov@mail.ru, 194064, Russia, St.Petersburg, Tikhoretsky pr., 21.*

Статья поступила в редакцию 08.12.2023; одобрена после рецензирования 30.05.2024; принята к публикации 03.06.2024.
The article was submitted 12/08/2023; approved after reviewing 05/30/2024; accepted for publication 06/03/2024.

УДК 004.27+621.382
DOI:10.25729/ESI.2024.34.2.005

Решение задачи классификации с использованием сегментной спайковой модели нейрона на ПЛИС

Исаков Тим Тимурович[1,3], Бахшиев Александр Валерьевич[1,2], Корсаков Антон Михайлович[3]

[1]Санкт-Петербургский политехнический университет Петра Великого,
[2]ООО Системы Компьютерного зрения, Россия, Санкт-Петербург,
[3]Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики,
Россия, Санкт-Петербург, *t.isakov@rtc.ru*

Аннотация. Спайковые нейронные сети — это класс нейронных сетей, имеющих в своей основе биоподобные модели нейрона. Спайковая природа таких сетей при наличии специализированных вычислителей позволяет достигать показателей энергоэффективности на порядки выше, чем у классических нейронных сетей, что является особо важным для встраивания нейронных сетей в автономные системы. Однако, на данный момент таких вычислителей нет в открытом доступе, поэтому хорошей альтернативой являются ПЛИС. Одним из классов спайковых моделей нейронов являются сегментные модели. Сегментные модели, в отличие от точечных, позволяют учитывать структуру нейрона, что, в свою очередь, позволяет воссоздавать более сложную динамику нейронных структур. Существующие нейроморфные вычислители позволяют реализовать только ограниченный набор спайковых моделей, что также является поводом для использования ПЛИС. На данный момент нет работ по аппаратной реализации сегментных моделей нейрона, поэтому данная работа является актуальной. В ходе работы были выполнены аппроксимация и аппаратная реализация CSNM модели (Compartmental spiking neuron model) на ПЛИС. Для проверки работоспособности полученной реализации был построен классификатор данных IRIS. На основе результатов был сделан вывод, что полученная модель имеет конкурентные показатели по количеству используемых ресурсов ПЛИС, а скорость расчетов при этом на три порядка выше, чем на компьютере. Точность полученной реализации уступает другим работам по причине использования малого количества нейронов и грубой аппроксимации. Дальнейшие исследования методов аппроксимации и алгоритмов инкрементного обучения позволят увеличить точность. Также планируется использовать оперативную память в целях масштабирования моделей и оптимизации вычислений. Еще одним направлением дальнейших работ является реализация обучения на чипе, как в целях ускорения тестирования моделей, так и для исследований обучения с подкреплением.

Ключевые слова: спайковые нейронные сети, классификация, программируемая логика

Цитирование: Исаков Т.Т. Решение задачи классификации с использованием сегментной спайковой модели нейрона на ПЛИС / Т.Т. Исаков, А.В. Бахшиев, А.М. Корсаков // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 50-58. – DOI:10.25729/ESI.2024.34.2.005.

Введение. На сегодняшний день нейронные сети благодаря впечатляющим результатам активно используются в самых разных отраслях жизни. Однако их дальнейшее развитие затруднено из-за различных ограничений. Одним из главных ограничений является низкая энергоэффективность нейронных сетей, что является проблемой при их встраивании в автономные системы. Решением этой и других проблем нейронных сетей занимаются ученые в области спайковых нейронных сетей, в основе которых лежат более биоподобные модели нейронов. Одним из главных отличий таких сетей от традиционных является то, что спайковые нейроны обмениваются не действительными числами, а спайками (единичными импульсами). Принцип работы нейронных сетей, построенных на спайковых моделях нейрона, подразумевает высокую степень асинхронности и параллельности работы сети, что трудновыполнимо на существующих CPU или даже GPU. По этой причине разрабатываются специализированные вычислители для работы со спайковыми нейронными сетями, такие, как TrueNorth [1], SpiNNaker [2], Loihi [3], Tianjic [4]. Однако данные вычислители либо

недоступны, либо способны воспроизводить очень ограниченный набор существующих спайковых моделей нейрона. Отличной альтернативной для исследователей является использование ПЛИС для реализации своих собственных моделей и архитектур. На данный момент существует большое количество реализаций различных моделей нейрона: LIF [5, 6, 7], SRM [8], SRM0 [9], Ижикевич [10], Ходжкин-Хаксли [11]. Существуют также работы по решению посредством ПЛИС задач распознавания образов (pattern recognition) [5, 12], распознавания символов (character recognition) [13], управления (control) [14], а также генерации паттернов (pattern generation) [15, 16].

Модели, учитывающие структуру нейрона, называют сегментными моделями (compartmental). На данный момент можно выделить несколько вариантов сегментных моделей: двухсегментную модель [17], состоящую из модулей сомы и дендритов, трехсегментную модель [18], состоящую из модулей сомы, апикальных и базальных дендритов и CSNM модель [19], состоящую из модулей синапсов, ионного механизма, описывающего участки дендритов, и сомы с низкопороговой зоной. В работе [20] демонстрируется, что многосегментные модели нейрона, реализованные на нейроморфных чипах, позволяют имитировать динамические процессы нейронных цепей мозга при низких энергозатратах. Однако сегментные модели нейрона могут также использоваться и для решения практических задач.

В работе [14] предлагается алгоритм обучения спайковой нейронной сети, использующей многосегментные нейроны для обработки данных о состоянии системы. Данная сеть управляет персонажем в компьютерных играх Atari, и, благодаря использованию сегментных нейронов, имеет значительный прирост игровых очков относительно других нейронных сетей, в том числе спайковых. Авторы делают вывод, что сегментные модели играют большую роль в глубоком распределенном обучении с подкреплением, что является важным основанием для дальнейшего развития сегментных моделей.

В работе [21] представлен алгоритм структурного обучения для решения задачи классификации. Данный метод имеет ряд преимуществ, так как для обучения может быть достаточно одного примера. Целью данной работы является реализация на ПЛИС классификатора для данных IRIS [22] на основе CSNM модели нейрона. CSNM модель была выбрана, так как по сравнению с другими сегментными моделями она позволяет строить сложную структуру дендритного дерева, а дальнейшее развитие алгоритма может позволить одновременно использовать структурное и параметрическое обучение. Полученная реализация позволит выполнять моделирование спайковых нейронных сетей на сегментных моделях нейрона в реальном времени или даже быстрее.

1. Сегментная спайковая модель нейрона. Как уже было сказано выше, сегментные модели учитывают структуру дендритного дерева. Модель CSNM описывается некоторой системой дифференциальных уравнений, подробное описание которой представлено в работе [19]. Для решения данной системы используется метод преобразования Эйлера.

Согласно модели, входные сигналы поступают на синапсы, отражающие влияние синаптического тока на сегмент мембраны нейрона. Поведение синапсов моделируется с помощью полученной после преобразований системы (начальные условия $\rho(0) = 0$):

$$\left.\begin{aligned} \rho_i &= \rho_{i-1} + h\left(\tfrac{x-\rho_{i-1}}{T_s}\right) \\ T_s(t) &= \begin{cases} \tau_s, & x(t) > 0 \\ \tau_d, & x(t) \le 0 \end{cases} \\ i_s &= \rho \tfrac{\varepsilon_s}{R_s}\omega \end{aligned}\right\}, \qquad (1)$$

где ρ – характеризует условную концентрацию медиатора, выделившегося в ответ на импульс,

h – шаг интегрирования,
T_s – постоянная времени выделения или распада медиатора,
R_s – сопротивление синапса,
ε_s – электродвижущая сила синаптической передачи,
ω – вес связи,
i_s – выходной синаптический ток.

Значения синаптического тока подаются на сегменты мембраны. Ионные механизмы, описывающие деполяризацию и гиперполяризацию мембраны, моделируются с помощью системы (начальные условия: $U(0) = 0$):

$$\left.\begin{aligned} U_i &= U_{i-1} + h\left(\frac{U_\Sigma - U_{i-1}(1+g_\Sigma R_m)}{T_I}\right) \\ T_I &= \frac{C_m}{g_\Sigma + R_m^{-1}} \end{aligned}\right\}, \quad (2)$$

где U_i – вклад сегмента в величину внутриклеточного потенциала,
h – шаг интегрирования,
R_m – сопротивление ионного механизма,
C_m – емкость ионного механизма,
g_Σ – суммарная проводимость моделей активных синапсов,
U_Σ – ожидаемый вклад сегмента в величину внутриклеточного потенциала при отсутствии внешнего возбуждения, которая определяется активностью предыдущих сегментов мембраны.

Выходы сегментов мембраны на теле нейрона (сома) суммируются и подаются на генератор выходного сигнала, имеющий некоторый порог срабатывания. Если нейрон срабатывает, на входы сегментов на теле нейрона подается тормозящий сигнал, что позволяет моделировать рефрактерный период.

2. Аппаратная реализация CSNM модели на ПЛИС. Решение дифференциальных уравнений напрямую требует большого количества вычислительных ресурсов. Ресурсы ПЛИС ограничены, поэтому создание нейронных сетей большого объема в таком случае проблематично. Решения дифференциальных уравнений, описывающих спайковые нейронные сети, включают в себя ядра с экспоненциальной временной зависимостью. Существуют различные варианты аппроксимации таких экспоненциальных функций, направленные на определенную точность и вычислительные затраты. В данной работе используется аппроксимация похожая на методы base-2 и CORDIC, так как операции умножения и деления заменяются на сдвиги и операции сложения. Важным отличием является то, что для полученной модели не нужно использовать память, что позволяет сохранить ресурсы ПЛИС. Также вычисления с плавающей запятой приведены к целочисленным вычислениям, что также экономит ресурсы ПЛИС. Минусом такого подхода является подбор диапазонов значений рассматриваемых величин и констант. Также из-за использования сдвигов теряется гибкость настройки модели, однако на данном этапе это не является проблемой.

Логика ПЛИС позволяет детально аппаратно реализовать сегментную модель нейрона. На рисунке 1 представлена RTL-схема сегментной модели нейрона. Для работы модели используются 3 входа: **clk**, регулирующий синхронную работу элементов; **reset**, обнуляющий значения блоков при запуске; **input spike**, отвечающий за входной сигнал нейрона. Входной сигнал передается в блок **Synapse**, реализующий динамику, описанную в системе уравнений (1). Значения, генерируемые синапсами, передаются в возбуждающий канал **Exc_Chennel** (динамика описана в системе уравнений (2)). В данной реализации на блок **Inh_Chennel**, моделирующий тормозный ионный механизм, сигнал с синапса не подается. Выходы возбуждающего и тормозного канала передаются в блок **Membrane**, внутри которого эти

значения суммируются. Как только выход блока мембраны **input potential** становится больше порога в блоке **LTZone**, описывающего низкопороговую зону, нейрон выдает выходной спайк **Output spike**. Вместе с этим запускается обратная связь, которая записывает значение обратной связи в регистры **Feedback**. Обратная связь уменьшает проходимость сигнала от синапса через ионные механизмы, из-за чего нейрон некоторое время перестает выдавать выходной спайк. Полученная схема полностью описывает цикл работы сегментной спайковой модели нейрона.

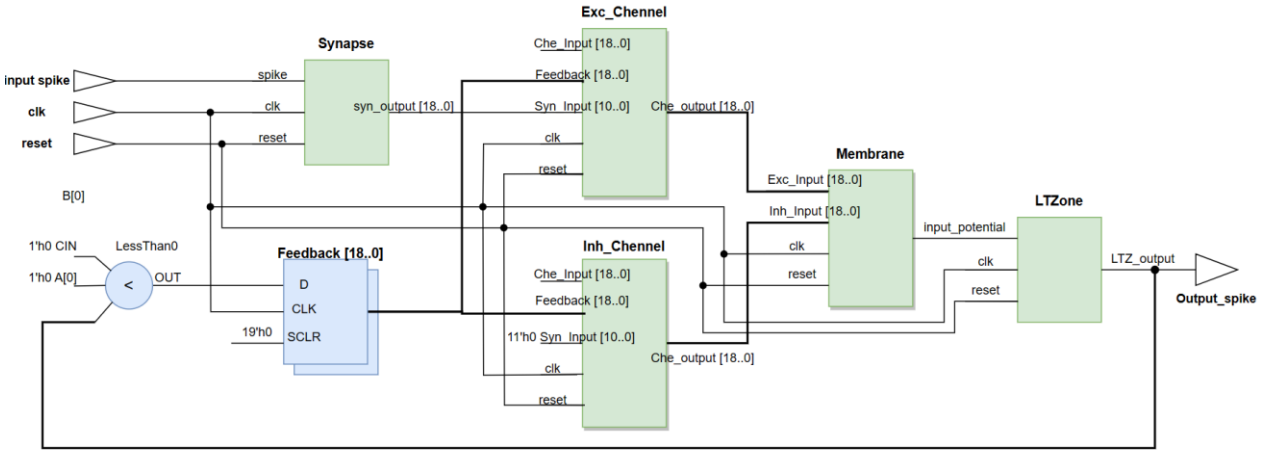


Рис. 1. RTL-схема сегментной модели нейрона

Таким образом, используя описанные модули и добавляя блоки ионных механизмов, можно построить сегментную модель нейрона любой сложности, ограничиваясь только ресурсами ПЛИС. В таблице 1 представлено сравнение полученной реализации CSNM модели с другими реализациями спайковых моделей нейронов в ПЛИС. Реализации сравниваются по количеству используемых ресурсов ПЛИС, количеству реализованных нейронов и среднеквадратичному отклонению реализации.

Таблица 1. Сравнение полученной реализации CSNM модели со связанными работами

Модель	FF (регистры) + LUT (логические элементы)	DSP	Количество нейронов	Память	Среднеквадратичное отклонение
LIF model (2011) [5]	14098 + 22815	-	648	126Kb (BRAM)	-
LIF model (2020) [6]	135 + 218	-	1	15Kb (-)	0,0015
SRM model (2012) [7]	202 + 378	-	1	-	-
LIF model (2006) [8]	4581	-	1,024	320Kb	LIF model (2006) [8]
SRM0 model (2020) [9]	296 + 314	4	512	1,003x64b (DRAM)	4×10^{-12}-1×10^{-10}
CSNM (Данная работа)	67 + 245 (16607)	-	1(3)	-	0,004 – 0,01

По значениям, представленным в таблице, можно сделать вывод, что модель имеет конкурентные показатели используемых ресурсов. Стоит отметить, что количество ресурсов, занимаемое нейроном, сильно зависит от структуры самого нейрона. Для сравнения, в таблице указаны ресурсы, используемые для одного базового необученного нейрона, а в скобках для 3

нейронов, обученных на данных IRIS [22]. В дальнейшем планируется использование памяти, что позволит сэкономить ресурсы ПЛИС при масштабировании моделей.

3. Классификатор данных IRIS. С помощью среды NeuroModeler [23] на трех примерах (по одному для каждого класса) из данных IRIS [22] была обучена модель классификатора. Полученная схема (рисунок 2) из трех нейронов была реализована в ПЛИС на чипе Cyclone IV EP4CE22F17C6N с 22.320 логическими элементами (рисунок 2). В данных IRIS рассматриваются 4 признака, которые преобразуются во входные паттерны нейронов. Также дополнительно используется калибровочный сигнал, чтобы нейроны оценивали абсолютные значения задержек, а не относительные. Обратная связь при срабатывании нейронов распространяется на все нейроны, чтобы избежать нескольких срабатываний. Таким образом «победителем» будет являться нейрон, испустивший спайк раньше.

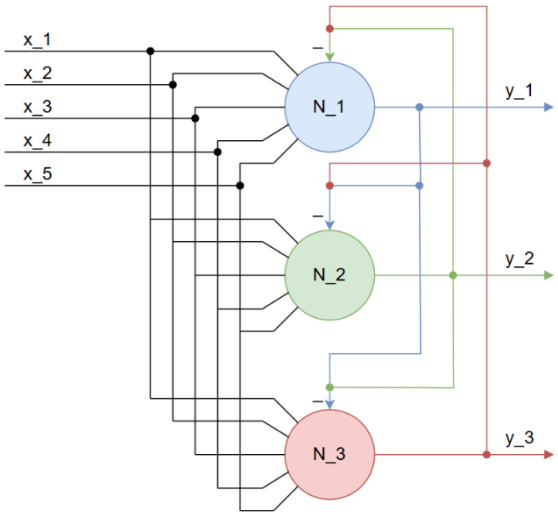


Рис. 2. Схема построенного классификатора

Сравнение представленной реализации с другими решениями задачи классификации данных IRIS с использованием импульсных нейронов представлено в таблице 2. Для описания архитектур используется формат Ni – Nh – No, указывающий количество нейронов на входном, скрытом и выходном слоях модели. Модели SRESN и DoB имеют по 2 слоя, однако количество нейронов выходного слоя может варьироваться (значения в скобках). Реализация CSNM модели также двухслойная, однако в реализации, представленной в данной работе, используется только один слой из трех нейронов.

Таблица 2. Сравнение существующих спайковых классификаторов данных IRIS

Модель	Архитектура	Кол-во настраиваемых параметров	Точность (%)
SpikeProp (Bohte et al. 2002) [24]	25-10-3	4480	96,1
SWAT (Wade et al. 2010) [25]	24-312-3	936	95,3
SRESN (2016) [26]	24-(5-11)	120-264	97,0
DoB-SNN (2021) [27]	5-(5-8)	25-40	97,7
CSNM (2021) [21]	12-3	120	83,3
CSNM (Данная работа)	3	30	76,6

Из таблицы видно, что полученная модель имеет сравнительно низкие показатели точности. Снижение точности относительно оригинального эксперимента [24] вызвано упрощением модели с 15 до 3 нейронов для экономии ресурсов ПЛИС, а также аппроксимацией вычислений. Положительной стороной модели является относительно малое количество настраиваемых параметров. Другим важным достижением является скорость работы алгоритма. Реакция одного нейрона на компьютере посредством CPU измеряется в

секундах или долях секунд (в зависимости от мощности CPU), в то время как реакция такого же нейрона на ПЛИС измеряется в долях миллисекунд. Таким образом, в данной реализации мы получаем прирост в тысячу раз, что особо ощутимо при оценке относительно больших моделей на больших данных, что на CPU может занимать несколько суток. Стоит отметить, что на данном этапе обучение нейрона выполняется с помощью одного примера, что также отражается в низкой точности. Инкрементное обучение, позволяющее дообучать нейроны на нескольких примерах, является одним из направлений будущих исследований.

Заключение. В ходе работы была выполнена аппроксимация и аппаратная реализация CSNM модели нейрона. На основе полученной модели для оценки работы реализации был построен классификатор, ресурсные характеристики которого не уступают схожим работам, а вычислительные характеристики превосходят таковые на CPU. Точность полученной модели ниже, чем в аналогичных работах, однако дальнейшие исследования в области инкрементного обучения и методов аппроксимации позволят увеличить точность. Также планируется использовать оперативную память в целях масштабирования моделей и оптимизации вычислений. Еще одним направлением дальнейших работ является реализация обучения на чипе как в целях ускорения тестирования моделей, так и для исследований обучения с подкреплением.

Благодарности. Результаты получены в рамках выполнения государственного задания Минобрнауки России №075-00697-24-00 от 27.12.2023 «Исследование и разработка биоподобной системы управления поведением мобильных роботов на базе энергоэффективных программно-аппаратных нейроморфных средств» (FNRG-2022-0016 1021060307690-3-1.2.1;2.2.2).

Список источников

1. Merolla P.A. et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. Science, 2014, vol.345, pp. 668-673.
2. Van Albada S.J. et al. Performance comparison of the digital neuromorphic hardware SpiNNaker and the neural network simulation software NEST for a full-scale cortical microcircuit model. Frontiers in neuroscience, 2018, vol.12, p. 291.
3. Davies M. et al. Loihi: A neuromorphic manycore processor with on-chip learning. Ieee Micro, 2018, vol. 38, pp. 82-99.
4. Pei J. et al. Towards artificial general intelligence with hybrid Tianjic chip architecture. Nature, 2019, vol. 572, pp. 106-111.
5. Caron L. C., Mailhot F., Rouat J. FPGA implementation of a spiking neural network for pattern matching. IEEE International Symposium of circuits and systems (ISCAS), 2011, pp. 649-652.
6. Perez-Peña F., Cifredo-Chacon M.A., Quiros-Olozabal A. Digital neuromorphic real-time platform. Neurocomputing, 2020, vol.371, pp. 91-99.
7. Ros E. et al. Real-time computing platform for spiking neurons (RT-spike). IEEE Trans. Neural Networks, 2006, vol.17, pp. 1050-1063.
8. Rosado-Muñoz A., Bataller-Mompeán M., Guerrero-Martínez J. FPGA implementation of spiking neural networks. IFAC Proceedings Volumes, 2012, vol.45, pp. 139-144.
9. Kim J. et al. Hardware-efficient emulation of leaky integrate-and-fire model using template-scaling-based exponential function approximation. IEEE Transactions on circuits and systems I: regular papers, 2020, v.68, pp. 350-362.
10. Ambroise M. et al. Biorealistic spiking neural network on FPGA, 2013, 47th Annual Conference on information sciences and systems (CISS). IEEE, 2013, pp. 1-6.
11. Yaghini Bonabi S. et al. FPGA implementation of a biological neural network based on the Hodgkin-Huxley neuron model. Frontiers in neuroscience, 2014, v.8, pp. 379.
12. Nevarez Y. et al. Accelerator framework of spike-by-spike neural networks for inference and incremental learning in embedded systems. 9th International Conference on modern circuits and systems technologies (MOCAST), IEEE, 2020, pp. 1-5.
13. Rice K.L. et al. FPGA implementation of Izhikevich spiking neural networks for character recognition. International Conference on reconfigurable computing and FPGAs, IEEE, 2009, pp. 451-456.

14. Sun Y. et al. Multi-compartment neuron and population encoding improved spiking neural network for deep distributional reinforcement learning. arXiv preprint arXiv:2301.07275, 2023.
15. Donati E. et al. A spiking implementation of the lamprey's Central Pattern Generator in neuromorphic VLSI. 2014. IEEE Biomedical circuits and systems conference (BioCAS). Proceedings, IEEE, 2014, pp. 512-515.
16. Rostro-Gonzalez H. et al. A CPG system based on spiking neurons for hexapod robot locomotion. Neurocomputing, 2015, v.170, pp. 47-54.
17. Urbanczik R., Senn W. Learning by the dendritic prediction of somatic spiking. Neuron, 2014, vol.81, pp. 521-528.
18. Sacramento J. et al. Dendritic cortical microcircuits approximate the backpropagation algorithm. Advances in neural information processing systems, 2018, vol.31.
19. Бахшиев А.В. Сегментная спайковая модель нейрона CSNM / А.В. Бахшиев, А.А. Демчева // Известия высших учебных заведений. Прикладная нелинейная динамика, 2022. – Т. 30. – №. 3. – С. 299-310.
20. Kopsick J.D. et al. Robust resting-state dynamics in a large-scale spiking neural network model of area CA3 in the mouse hippocampus. Cognitive computation, 2022, pp. 1-21.
21. Astapova L.A. et al. Compartmental spiking neuron model for pattern classification. Journal of physics: conference series, IOP Publishing, 2021, vol. 2094, p. 032032.
22. UCI Machine Learning Repository, available at: https://archive.ics.uci.edu/ml/datasets/iris (accessed: 08/24/2023).
23. Neuro modeler. Available at: https://neuromodeler.ru/ (accessed: 08/24/2023).
24. Bohte S.M., Kok J.N., La Poutré J.A. SpikeProp: backpropagation for networks of spiking neurons. ESANN, 2000, vol.48, pp. 419-424.
25. Wade J.J. et al. SWAT: A spiking neural network training algorithm for classification problems. IEEE Transactions on neural networks, 2010, vol.21, pp. 1817-1830.
26. Dora S. et al. Development of a self-regulating evolving spiking neural network for classification problem. Neurocomputing, 2016, vol.171, pp. 1216-1229.
27. Saranirad V. et al. DoB-SNN: a new neuron assembly-inspired spiking neural network for pattern classification. 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, 2021, pp. 1-6.

Исаков Тим Тимурович. *Аспирант Санкт-Петербургского политехнического университета Петра Великого, математик в Центральном научно-исследовательском и опытно-конструкторском институте робототехники и технической кибернетики (ЦНИИ РТК). Область научных интересов: алгоритм обучения спайковых нейронных сетей и их аппаратной реализацией на ПЛИС. AuthorID: 1117866, SPIN: 8742-1353, ORCID: 0000-0003-4437-5018, 194064, Санкт-Петербург, Тихорецкий пр., 21.*

Бахшиев Александр Валерьевич. *К.т.н., СПбПУ, доцент. Направления исследований: искусственные нейронные сети, нейроморфные системы, компьютерное зрение. AuthorID: 172135, SPIN: 7055-8676, ORCID: 0000-0002-1284-0088, palexab@gmail.com, 195251, Санкт-Петербург, ул. Политехническая, д. 29.*

Корсаков Антон Михайлович. *Старший научный сотрудник ЦНИИ РТК. Направления исследований: искусственные нейронные сети, нейроморфные системы, компьютерное зрение. AuthorID: 172137, SPIN: 3736-3401, anton_korsakov@mail.ru, 194064, Санкт-Петербург, Тихорецкий пр., д. 21.*

UDC 004.27+621.382
DOI:10.25729/ESI.2024.34.2.005

Solution of the classification task using a compartmental spiking neuron model on an FPGA

Tim T. Isakov[1,3], Alexandr V. Bakhshiev[1,2], Anton M. Korsakov[3]

[1]Peter the Great St. Petersburg Polytechnic University,
[2]LLC Computer Vision Systems, Russia, Saint Petersburg,
[3]Russian state scientific center for robotics and technical cybernetics (RTC),
Russia, Saint Petersburg, *t.isakov@rtc.ru*

Abstract. Spike neural networks are a class of neural networks based on plausible neuron models. The spiking nature of such networks, in the presence of specialized accelerators, makes it possible to achieve energy efficiency values that are orders of magnitude higher than those of classical neural networks, which is especially important for integrating neural networks into autonomous systems. However, now such computers are not publicly available, so FPGAs are a good alternative. One of the classes of spike neuron models are segmental models. Segmental models, unlike point ones, allow one to consider the structure of a neuron, which in turn allows one to reproduce more complex dynamics of neural structures. Existing neuromorphic computers allow the implementation of only a limited set of spike models, which is also a reason for using FPGAs. Today, there is no work on hardware implementation of segmented neuron models, so this work is relevant. During the work, the approximation and hardware implementation of CSNM model (Compartmental spiking neuron model) on an FPGA was performed. To evaluate the performance of the resulting implementation, an IRIS data classifier was built. Based on the results, it was concluded that the resulting model has competitive indicators in terms of the amount of FPGA resources used, and the calculation speed is three orders of magnitude higher than on a computer. The accuracy of the resulting implementation is inferior to other works due to the use of a small number of neurons and rough approximation. Further research into approximation methods and incremental learning algorithms will improve accuracy. It is also planned to use RAM for scaling models and optimizing calculations. Another area of future work is implementing on-chip learning, both to speed up model testing and for reinforcement learning research.

Keywords: spiking neural networks, classification, FPGA

Acknowledgements: The work was carried out as the part of the state task of the Russian Ministry of Education and Science for 2024 №075-00697-24-00 "Research and development of a biosimilar sys-tem for controlling the behavior of mobile robots based on energy-efficient soft-ware and hardware neuromorphic tools" (FNRG-2022-0016 1021060307690-3-1.2.1;2.2.2).

References

1. Merolla P. A. et al. A million spiking-neuron integrated circuit with a scalable communication network and interface, Science, 2014, v.345, pp. 668-673.
2. Van Albada S. J. et al. Performance comparison of the digital neuromorphic hardware SpiNNaker and the neural network simulation software NEST for a full-scale cortical microcircuit model, Frontiers in neuroscience, 2018, v.12, p. 291.
3. Davies M. et al. Loihi: A neuromorphic manycore processor with on-chip learning, Ieee Micro, 2018, v.38, pp. 82-99.
4. Pei J. et al. Towards artificial general intelligence with hybrid Tianjic chip architecture, Nature, 2019, v.572, pp. 106-111.
5. Caron L. C., Mailhot F., Rouat J. FPGA implementation of a spiking neural network for pattern matching, 2011 IEEE International Symposium of Circuits and Systems (ISCAS), 2011, pp. 649-652.
6. Perez-Peña F., Cifredo-Chacon M. A., Quiros-Olozabal A. Digital neuromorphic real-time platform, Neurocomputing, 2020, v.371, pp. 91-99.
7. Ros E. et al. Real-time computing platform for spiking neurons (RT-spike), IEEE Trans. Neural Networks, 2006, v.17, pp. 1050-1063.
8. Rosado-Muñoz A., Bataller-Mompeán M., Guerrero-Martínez J. FPGA implementation of spiking neural networks, IFAC Proceedings Volumes, 2012, v.45, pp. 139-144.
9. Kim J. et al. Hardware-efficient emulation of leaky integrate-and-fire model using template-scaling-based exponential function approximation, IEEE Transactions on Circuits and Systems I: Regular Papers, 2020, v.68, pp. 350-362.
10. Ambroise M. et al. Biorealistic spiking neural network on FPGA, 2013, 47th Annual Conference on information sciences and systems (CISS). IEEE, 2013, pp. 1-6.
11. Yaghini Bonabi S. et al. FPGA implementation of a biological neural network based on the Hodgkin-Huxley neuron model, Frontiers in neuroscience, 2014, v.8, pp. 379.
12. Nevarez Y. et al. Accelerator framework of spike-by-spike neural networks for inference and incremental learning in embedded systems, 9th International Conference on Modern Circuits and Systems Technologies (MOCAST), IEEE, 2020, pp. 1-5.
13. Rice K.L. et al. FPGA implementation of Izhikevich spiking neural networks for character recognition. International Conference on reconfigurable computing and FPGAs, IEEE, 2009, pp. 451-456.
14. Sun Y. et al. Multi-compartment neuron and population encoding improved spiking neural network for deep distributional reinforcement learning. arXiv preprint arXiv:2301.07275, 2023.

15. Donati E. et al. A spiking implementation of the lamprey's Central Pattern Generator in neuromorphic VLSI, 2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings, IEEE, 2014, pp. 512-515.
16. Rostro-Gonzalez H. et al. A CPG system based on spiking neurons for hexapod robot locomotion, Neurocomputing, 2015, v.170, pp. 47-54.
17. Urbanczik R., Senn W. Learning by the dendritic prediction of somatic spiking, Neuron, 2014, v.81, pp. 521-528.
18. Sacramento J. et al. Dendritic cortical microcircuits approximate the backpropagation algorithm, Advances in neural information processing systems, 2018, v.31.
19. Bakhshiev A.V., Demcheva A.A. Segmentnaja spaikovaja model neirona CSNM [Compartmental spiking neuron model CSNM]. Izvestiya vysshikh uchebnykh zavedeniy. Prikladnaya nelineynaya dinamika [News of higher educational institutions. Applied nonlinear dynamics], 2022, v.30, no. 3, pp. 299-310.
20. Kopsick J. D. et al. Robust resting-state dynamics in a large-scale spiking neural network model of area CA3 in the mouse hippocampus, Cognitive computation, 2022, pp. 1-21.
21. Astapova L. A. et al. Compartmental spiking neuron model for pattern classification, Journal of Physics: Conference Series, IOP Publishing, 2021, v.2094, p. 032032.
22. UCI Machine Learning Repository, available at: https://archive.ics.uci.edu/ml/datasets/iris (accessed: 08/24/2023).
23. Neuro modeler. Available at: https://neuromodeler.ru/ (accessed: 08/24/2023).
24. Bohte S. M., Kok J. N., La Poutré J. A. SpikeProp: backpropagation for networks of spiking neurons, ESANN, 2000, v.48, pp. 419-424.
25. Wade J. J. et al. SWAT: A spiking neural network training algorithm for classification problems, IEEE Transactions on neural networks, 2010, v.21, pp. 1817-1830.
26. Dora S. et al. Development of a self-regulating evolving spiking neural network for classification problem. Neurocomputing, 2016, v.171, pp. 1216-1229.
27. Saranirad V. et al. DoB-SNN: A New Neuron Assembly-Inspired Spiking Neural Network for Pattern Classification, 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, 2021, pp. 1-6.

Isakov Tim Timurovich. *Graduate student at Peter the Great St. Petersburg polytechnic university, mathematician at the Russian state scientific center for robotics and technical cybernetics (RTC). Research direction: research into learning algorithms for spiking neural networks and their hardware implementation on FPGAs. AuthorID: 1117866, SPIN 8742-1353, ORCID: 0000-0003-4437-5018, t.isakov@rtc.ru, 194064, Saint Petersburg, Tikhoretsky Avenue 21.*

Bakhshiev Alexander Valeryevich. *Ph.D., SPbPU, assistant professor. Research direction: artificial neural networks, neuromorphic systems, computer vision. AuthorID 172135, SPIN 7055-8676, ORCID 0000-0002-1284-0088, palexab@gmail.com, 195251, St. Petersburg, Polytechnicheskaya, 29.*

Korsakov Anton Mikhailovich. *RTC, senior researcher, Research direction: artificial neural networks, neuromorphic systems, computer vision. AuthorID: 172137, SPIN: 3736-3401, anton_korsakov@mail.ru, 194064, St. Petersburg, Tikhoretsky pr., 21.*

Статья поступила в редакцию 08.12.2023; одобрена после рецензирования 30.05.2024; принята к публикации 03.06.2024.
The article was submitted 12/08/2023; approved after reviewing 05/30/2024; accepted for publication 06/03/2024.

УДК 004.852
DOI:10.25729/ESI.2024.34.2.006

Оценка качества интерполяции плотности распределения случайной величины при помощи искусственной нейронной сети

Доррер Михаил Георгиевич

Сибирский государственный университет науки и технологий им. М.Ф. Решетнева,
Россия, Красноярск, *mdorrer@mail.ru*

Аннотация. В работе описаны создание и оценка работоспособности нейросетевой модели функции плотности распределения вероятности случайной величины, заданной набором измерений случайной величины при отсутствии стадии идентификации закона распределения. Потребность в решении данной задачи вызывается ограничениями, вносимыми в точность расчета функции плотности распределения вероятности случайной величины как таблично-гистограммным методом, так и в случае применения подходов к идентификации закона распределения. Задача решалась на языке Python с использованием нейросетевой библиотеки TensorFlow путем создания нейросетевой модели на базе класса Sequential с полносвязными слоями Dense, обученной на данных численного дифференцирования функции распределения случайной величины. Точность прогноза оценивалась с помощью меры расстояния Кульбака-Лейблера для различных соотношений объема экспериментальных данных и количества интервалов интерполяции на синтетических тестовых данных, сгенерированных для 5 законов распределения – Рэлея, Вэйбулла, гамма, экспоненциального и нормального (гауссовского). Для оценки прогностической способности подхода при тестировании интерполятора использовались отсчеты случайной величины, сдвинутые по отношению к используемым при обучении. Предложенное решение показало значительно более высокую точность расчета значений плотности распределения случайной величины по сравнению с гистограммным методом. Разработанный подход будет внедрен в моделирующую часть цифрового двойника бизнес-процесса, основанного на математическом аппарате стохастических GERT-сетей.

Ключевые слова: стохастические модели, плотность распределения случайной величины, нейронные сети, цифровые модели процессов

Цитирование: Доррер М.Г. Оценка качества интерполяции плотности распределения случайной величины при помощи искусственной нейронной сети / М.Г. Доррер // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 59-68. – DOI:10.25729/ESI.2024.34.2.006.

Введение. В настоящее время задачи управления организационными системами постоянно усложняются [1]. Отдельные части организационных систем становятся все более интеллектуальными, число н сложность выполняемых ими функций возрастают. Разные части организационной системы функционируют параллельно и в процессе решения задач постоянно обмениваются между собой информацией. Ведется обмен не только данными, но и во все большей мере звуковой и видеоинформацией, часто в реальном времени. Примерами таких систем являются системы управления бизнес-процессами организации, обеспечивающие взаимодействие пользователей как внутри организации, так и с внешними акторами, и включающие в себя интеллектуальные системы – боты, ассистенты, системы роботизации бизнес-процесса и т.д.

Предлагаемый в данной работе результат был получен в ходе решения задач по созданию цифрового двойника бизнес-процесса на базе системы управления бизнес-процессами [2]. В качестве математического аппарата для моделей аддитивных параметров бизнес-процесса (время выполнения, стоимость, ресурсоемкость) в цифровом двойнике используется модель стохастических GERT-сетей. Для использования этого математического аппарата необходимо было получить решение задачи нахождения непрерывных плотностей распределения данных параметров при прохождении стохастических GERT-сетей с использованием топологического уравнения для оценки вероятностно-временных и вероятностно-ресурсных характеристик организационных систем.

При этом разнородность организационно-технической системы требовала сопряжения нескольких моделей бизнес-процессов, как на одном уровне, так и иерархически. При этом должно было обеспечиваться информационное единство среды, на которой строятся сложные иерархические модели, моделируемые в цифровом двойнике. На ранних этапах реализации это обеспечивалось путем представления распределений времени выполнения отдельных операций (эквивалентных операций, GERT-сетей) их характеристическими функциями в табличном представлении. Такой подход применяется, когда исходная или промежуточная информация получена в виде плотности или функции распределения. Однако применение табличного представления накладывает довольно существенное ограничение на алгоритмы расчета, привязывая их к единой, фиксированной для всей модели интерполяционной сетке.

В данной работе будет рассмотрено решение задачи представления характеристических функций аддитивных параметров операций модели бизнес-процесса в виде искусственной нейронной сети, включая сравнительный анализ данного подхода с базовым, таблично-гистограмным методом расчета функции плотности распределения.

В разделе 1 представлен краткий обзор публикаций, затрагивающих тему цифрового моделирования бизнес-процессов, в разделе 2 – описание постановки задачи и использованных экспериментальных материалов, в разделе 3 показаны применяемые методы решения задачи интерполяции функции плотности распределения вероятности, в разделе 4 описана интерпретация полученных результатов.

1. Связанные работы. Работы по моделированию бизнес-процессов с помощью математических моделей в настоящее время разрабатываются достаточно активно, в публикациях описаны несколько частных решений данной задачи. Так, в работе [3] предложен подход к формальному описанию бизнес-процессов с использованием ряда математических ограничений (которые определяют границы осуществимости бизнес-процесса) и набора целевых функций (которые состоят из различных целей для проектирования бизнес-процесса). Аналогичный подход представлен в работе [4], авторы которой описывают математическую модель, содержащую основные компоненты общего бизнес-процесса. В работе [5] предлагается использовать формальные математические обозначения как способ введения бизнес-правил и проверки логической согласованности схематических моделей. В работе [6] предложено теоретико-множественное формальное описание бизнес-процессов, которое, например, ставит задачи разработки алгоритмов верификации бизнес-процесса [7]. Авторы работы [8] предлагают использовать для анализа бизнес-процессов их описание в форме тензоров специального вида и определенных над ними операций. Первоначально этот подход был разработан для решения более общей задачи анализа процессов в сложных сетевых системах [9].

Для решения ряда частных задач по анализу бизнес-процессов используются стохастические GERT-сети (Graphical Evaluation and Review Technique) [10 - 13]. GERT-сеть позволяет моделировать поток работ в виде ориентированного графа, ребрами которого являются задачи бизнес-процесса [14, 15].

Данная работа является продолжением цикла публикаций, посвященных реализации цифрового двойника бизнес-процесса – ЦДБП (Business Process Digital Twin, BPDT) [16], [17], [18], [2].

Актуальность темы обусловлена проблемами в точности прогноза вероятностных характеристик бизнес-процессов, выявленных в построении функции плотности распределения случайной величины, заданной отчетами случайного процесса, таблично-гистограммном методом [19] и путем идентификации законов распределения [20].

2. Постановка задачи и экспериментальные материалы. Оценка методов расчета плотности распределения случайной величины выполняется с использованием данных, сгенерированных для 5 распределений, для которых известны аналитические выражения плотности распределения. Работа базируется на методах, предложенных в статье [21] с расширением набора моделируемых законов распределения, а также содержит полностью оригинальный блок создания программного объекта – интерполятора функции плотности распределения случайной величины на базе многослойной полносвязной нейронной сети. Используемый набор законов распределения включает закон Релея, гамма-распределение, распределение Вейбулла, экспоненциальное распределение и нормальное (гауссово) распределение.

Исходные данные в виде массивов измерений случайной величины для 5 вышеперечисленных законов распределения генерируются с помощью кода Python, показанного в листинге 1.

Листинг 1. Генерация исходных данных для набора законов распределения

```
# Генерируем плотности распределения в соответствии с 5 законами
pdf_function = {
    'rrand': partial(rel_pdf, 1),
    'grand': partial(gam_pdf, 0.5, 0.5),
    'wrand': partial(weib_pdf, 1, 5),
    'exprand': partial(exp_pdf, 1.5),
    'gaussrand': partial(gauss_pdf, 1.0, 1.0)
}
```

Пример функции **gauss_pdf**, генерирующей синтетические данные отсчетов случайной величины для нормального распределения, приведен в листинге 2.

Листинг 2. Функция генерации исходных данных для нормального закона распределения

```
def gauss_pdf(mu: float, sigma: float, X: list) -> pandas.DataFrame:
    """
    Вычисляет кривую гауссовской плотности распределения вероятности по известной
формуле
    :param n: количество отсчетов
    :param mu: матожидание
    :param sigma: дисперсия
    :param X: координаты по оси абсцисс
    :return: pandas.DataFrame
    """
    pdf_y = []  # Координаты по оси ординат
    for x in X:
        pdf_y.append( (1 / np.sqrt(2 * np.pi * sigma)) * np.exp(-(x - mu)**2 / (2 *
sigma)))
    return pd.DataFrame({'x': X, 'y': pdf_y})
```

Далее выполним сравнение точности расчета функции плотности распределения случайной величины для вышеперечисленных 5 законов между аналитическими значениями и значениями, полученными следующими методами:

- путем интерполяции таблично-гистограммным методом;
- путем интерполяции полносвязной нейронной сетью, обученной на табличных значениях функции плотности распределения случайной величины, рассчитанной как первая производная от функции распределения, которая, в свою очередь, представляет собой вероятность обнаружить значение случайной величины меньше либо равное заданному [22].

3. Методы решения задачи. Сравнение точности исследуемых методов производилось с помощью меры расстояния Кульбака-Лейблера от априорной функции плотности распределения P до получаемой оценки Q, рассчитываемой как $KL = \mathbb{E}\left[P \cdot \ln\left(\frac{P}{Q}\right)\right]$.

Далее производится расчет обучающей выборки для нейросетевого интерполятора функции плотности распределения случайной величины. При этом используется следующий алгоритм [22] (N – количество доступных отсчетов случайной величины):

1) сортировка значений отсчетов случайной величины X по возрастанию, получение упорядоченного набора значений $\hat{X}$;
2) сопоставление с каждым значением в $\hat{X}$ его порядкового номера i, начиная с нуля, тогда с точностью до множителя $i(\hat{X}_i)$ представляет собой оценку функции распределения случайной величины;
3) построение равномерной шкалы из $k + 2$ значений на интервале от X_{min} до X_{max} где k – желаемое число узлов сетки, в которых рассчитывается функция плотности распределения ($k < N$);
4) интерполяция значений номеров переменных из шкалы упорядоченного массива значений переменной в равномерную шкалу, полученную в п. 3;
5) численное дифференцирование интерполированной функции по соседним точкам (для чего сетка и состоит из $k + 2$ значений) с делением каждого значения на N – в результате получается искомая оценка плотности вероятности.

В результате выполнения данного алгоритма формируется переменная **d** – таблица значений плотности распределения случайной величины. В реализованном в рамках настоящего исследования примере она имеет вид DataFrame из библиотеки Pandas, содержащего два столбца вещественных чисел – «**x**» – значения случайной величины и «**y**» – значения плотности распределения для данного значения случайной величины.

Далее на основе полученного датафрейма **d** выполняется формирование и обучение модели нейронной сети, позволяющей получить значение функции плотности распределения не только в узлах сетки, содержащейся в обучающей выборке, но и на всей области определения. Код Python для решения задачи приведен в листинге 3.

Листинг. 3. Код создания и обучения нейронной сети – интерполятора

```
def generate_neural(k: int, rnd_list: list) -> Sequential:
    """
    Получает кривую плотности распределения вероятности
    :param k: количечиво интервалов разбиения гистограммы
    :param rnd_list: случайный процесс
    :return: pandas.DataFrame
    """
    pdf_x = []  # Координаты по оси абсцисс
    pdf_y = []  # Координаты по оси ординат
    n = len(rnd_list)  # количество элементов в рассматриваемой выборке
    h = (max(rnd_list) - min(rnd_list)) / k  # ширина одного интервала
    a = min(rnd_list)  # минимальное значение в рассматриваемой выборке
    rnd_list = sorted(rnd_list)  # сортируем значения
    j = 0  # индекс значения левой границы интервала
    for i in range(0, k):  # проход по интервалам
        count = 0
        while j < n and (a + i * h) <= rnd_list[j] < (a + (i * h) + h):  #
        подсчитываем количество значений в k-м интервале
            count = count + 1
            j += 1
        pdf_x.append(a + i * h + h / 2)  # координата по оси абсцисс полученной
        кривой плотности распределения
        # вероятности
```

```
        pdf_y.append(count / (n * h))  # координата по оси ординат полученной
        кривой плотности распределения
        # вероятности
    d = {'x': pdf_x, 'y': pdf_y}
    x = np.array(d['x'])
    y = np.array(d['y'])
    Regressor = Sequential()
    Regressor.add(Dense(32, input_dim=1, activation='tanh'))
    Regressor.add(Dense(16, activation='tanh'))
    Regressor.add(Dense(1, activation='sigmoid'))
    # Определить функцию ошибки (loss function) и метод оптимизации (optimizer)
    для обучения модели.
    Regressor.compile(loss='mse', optimizer='adam', metrics=['accuracy'])
    # Обучим модель на данных плотности распределения, используя метод fit() в
    Keras.
    Regressor.fit(x, y, epochs=100, verbose = 0)
    # Передаем результаты для оценки
    return Regressor
```

Полученный объект **Regressor** передается в качестве выхода функции формирования объекта-интерполятора и может быть далее использован как встроенный объект, позволяющий получить значение функции плотности распределения случайной величины для любой точки области определения данной функции.

Его применение для построения функции плотности распределения приведено в листинге 4.

Листинг 4. Код формирования функции плотности распределения с помощью нейросети – интерполятора Regressor

```
def pdf_neural(k: int, rnd_list: list, Regressor: Sequential) -> pandas.DataFrame:
    """
    Получает кривую плотности распределения вероятности
    :param k: количечиво интервалов разбиения гистограммы
    :param rnd_list: случайный процесс
    :return: pandas.DataFrame
    """
    pdf_x = []  # Координаты по оси абсцисс
    pdf_y = []  # Координаты по оси ординат
    n = len(rnd_list)  # количество элементов в рассматриваемой выборке
    h = (max(rnd_list) - min(rnd_list)) / k  # ширина одного интервала
    a = min(rnd_list)  # минимальное значение в рассматриваемой выборке
    rnd_list = sorted(rnd_list)  # сортируем значения
    j = 0  # индекс значения левой границы интервала
    for i in range(0, k):  # проход по интервалам
        count = 0
        while j < n and (a + i * h) <= rnd_list[j] < (a + (i * h) + h):  #
        подсчитываем количество значений в k-м интервале
            count = count + 1
            j += 1
        pdf_x.append(a + i * h + h / 2)  # координата по оси абсцисс полученной
        кривой плотности распределения
    pdf_y = Regressor.predict(pdf_x, verbose = 0)
    d = {'x': pdf_x, 'y': pdf_y}
    x = np.array(d['x'])
    y = np.array(d['y'])
    print(x.shape)
    return d
```

Следует отметить, что для оценки прогностической способности модели ее обучение строится на одной выборке точек, а тестирование – на другой, специально сгенерированной для этой цели. Фрагмент кода, генерирующий отсчеты случайной величины для 5 анализируемых законов распределения в базовом и тестовом варианте, приведён в листинге 5.

Листинг 5. Код создания выборок случайных величин для заданных законов

```
# Генерируем случайные серии в соответствии с 5 законами распределения
random_series = {
    'rrand': [rel_rand(N, 1), rel_rand(Ntrain, 1)],  # генерируем случайные числа
        с распределением Релея
    'grand': [gam_rand(N, 0.5, 0.5), gam_rand(Ntrain, 0.5, 0.5)],  # генерируем
        случайные числа с гамма распределением
    'wrand': [weib_rand(N, 1, 5), weib_rand(Ntrain, 1, 5)], # генерируем случайные
        числа с распределением Вейбулла
    'exprand': [exp_rand(N, 1.5), exp_rand(Ntrain, 1.5)],  # генерируем случайные
        числа с экспоненциальным распределением
    'gaussrand': [gauss_rand(N, 1.0, 1.0), gauss_rand(Ntrain, 1.0, 1.0)]  #
        генерируем случайные числа с нормальным распределением
}
```

С учетом этой структуры код, формирующий оценку отклонения рассчитанной аналитически плотности распределения для каждого из 5 анализируемых законов от значений, предсказанных при помощи нейросетевой модели на базе меры Кульбака-Лейблера, выглядит так, как это показано в листинге 6.

Листинг 6. Оценка отклонения прогноза плотности распределения

```
metrics_neural = {}
for key, val in random_series.items():
    for k in k_values:
        Regressor = generate_neural(k, val[1][:Ntrain])
        #Regressor = generate_neural(k, val[:N])
        estimated_pdf = pdf_neural(k, val[0][:N], Regressor)
        theoretical_pdf = pdf_function[key](estimated_pdf['x'])
        metrics_neural.setdefault(key, []).append(
            KL_dist(theoretical_pdf['y'], estimated_pdf['y']))
    'gaussrand': [gauss_rand(N, 1.0, 1.0), gauss_rand(Ntrain, 1.0, 1.0)]  #
        генерируем случайные числа с нормальным распределением
}
```

4. Результаты. Для сравнительной оценки точности полученного метода с базовым, таблично-гистограммным, выполнялось построение меры отклонения Кульбака-Лейблера при разных соотношениях k/N (k – размерность сетки интерполяции, N – количество отсчетов в синтетических данных, в данной работе принято равным 10 000) для каждого из 5 законов распределения. Результат сравнения показан на рис. 1. Законы распределения показаны в легенде к рис. 1: «**rrand**» – Релея, «**grand**» – Гамма, «**wrand**» – Вейбулла, «**exprand**» – экспоненциальное, «**gaussrand**» –нормальное (гауссово). Расчетные данные к рисунку 1 приведены в таблицах 1 и 2.

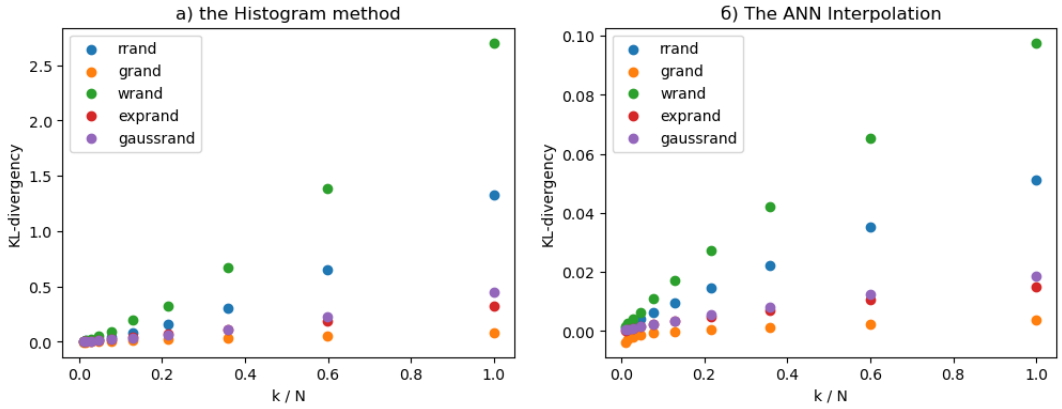


Рис. 1. Сравнительная точность интерполяции для таблично-гистограммного метода и для полученной нейросетевой интерполяции.

Таблица 1. Точность интерполяции гистограммным методом

Величина k/N	**rrand**	**grand**	**wrand**	**exprand**	**gaussrand**
0.0100	0.003333	-0.004176	0.004172	0.001169	0.001051
0.0166	0.005395	-0.002627	0.014493	0.002487	0.002485
0.0278	0.010528	-0.001127	0.027640	0.004309	0.005865
0.0464	0.019553	0.001291	0.065812	0.007118	0.009863
0.0774	0.033756	0.004193	0.106763	0.011950	0.018202
0.1291	0.063296	0.008189	0.186010	0.022943	0.029341
0.2154	0.120291	0.015558	0.398498	0.037571	0.051529
0.3593	0.226427	0.027862	0.706398	0.066140	0.107347
0.5994	0.483891	0.047202	1.495670	0.108961	0.198159
1.0000	1.012645	0.076647	3.077090	0.179611	0.404091

Таблица 2. Точность интерполяции при помощи нейросети

Величина k/N	**rrand**	**grand**	**wrand**	**exprand**	**gaussrand**
0.0100	0.216864	0.113760	0.442127	0.396204	0.065176
0.0166	0.318754	0.130621	0.576789	0.398202	0.103963
0.0278	0.327285	0.163094	0.657532	0.427052	0.131896
0.0464	0.557939	0.170919	0.826366	0.387221	0.142835
0.0774	0.389054	0.213921	0.884478	0.379168	0.157188
0.1291	0.493654	0.245398	0.910070	0.477201	0.174697
0.2154	0.468682	0.235149	0.931008	0.499094	0.178577
0.3593	0.454258	0.235967	0.965550	0.500662	0.188330
0.5994	0.531077	0.263226	0.985074	0.509339	0.179670
1.0000	0.413812	0.257140	0.997786	0.495032	0.185333

Как видно, в обоих случаях графики показывают монотонно возрастающие зависимости метрики отклонения от соотношения k/N, однако, нейросетевой интерполятор, обученный на данных численного дифференцирования функции распределения случайной величины, показывает на порядок меньшие значения метрики отклонения Кульбака-Лейблера между кривой оценки плотности распределения и теоретической кривой, чем базовый (гистограммный) метод, основанный на построении гистограммы.

5. Вывод. Предложенный подход к нейросетевой интерполяции функции плотности распределения случайной величины по заданному набору наблюдений отличается более высокой точностью, по сравнению с базовым таблично-гистограммным методом. Преимущество предложенного подхода особенно велико при больших значениях k/N, что позволяет говорить о преимуществах предложенного метода в ситуации малого объема экспериментальных данных, что особенно ценно для задачи анализа и прогнозирования бизнес-процессов, где частота выполнения операций редко превышает сотни, максимум – тысячи раз за период функционирования системы.

Таким образом, поставленная задача – построение нейросетевой интерполяции плотности распределения случайной величины – была успешно решена. Успех обеспечило совместно применение таких методов и алгоритмов, как:

1. Подготовка обучающих данных для нейросетевого интерполятора путем расчета значений функции плотности распределения случайной величины как первой производной от функции распределения.
2. Реализация интерполятора в виде многослойной нейронной сети (Sequental), позволяющей встраивать данную функцию в любой код, требующий получения значений функции плотности распределения случайной величины в заданных точках области распределения.

Дальнейшим развитием данной работы будет встраивание разработанной модели интерполятора в моделирующую часть цифрового двойника бизнес-процессов [17] и проверка его прогностической способности на реальных протоколах выполнения бизнес-процессов.

Список источников

1. Ahmad T., Van Looy A. Business process management and digital innovations: a systematic literature review. Sustainability, 2020, vol. 12, no. 17, 6827 p.
2. Dorrer M. Implementation of the digital twin of educational business processes on the ELMA BPMS platform. Proceedings of the II International scientific conference on advances in science, engineering and digital education: (ASEDU-II 2021) AIP Conference Proceedings, 2022, 040088 p.
3. Hofacker I., Vetschera R. Algorithmical approaches to business process design. Comput. Oper. Res., 2001, vol. 28, no. 13, pp. 1253-1275.
4. Powell S.G., Schwaninger M., Trimble C. Measurement and control of business processes. Syst. Dyn. Rev., 2001, vol. 17, no. 1, pp. 63-91.
5. Valiris G., Glykas M. Business analysis metrics for business process redesign. Bus. Process Manag. J., 2004, vol. 10, no. 4, pp. 445-480.
6. Калянов Г.Н. Теория и практика реорганизации бизнес-процессов / Г.Н. Калянов. – Москва: СИНТЕГ, 2000. – 203 с.
7. Доррер М.Г. Оптимизация расчёта инвариантов сети Петри в рамках задачи формирования сценариев интеграционного тестирования / М.Г. Доррер, В.В. Курохтин // Моделирование и анализ информационных систем, 2012. – Т. 19. – № 4. – С. 128-143.
8. Петров М.И. Использование тензорной методологии анализа систем для описания бизнес-процессов предприятия. Средства автоматизации / М.И. Петров, И.Г. Треногий, Е.А. Беловатый // Сибирский аэрокосмический журнал, 2008. – Т. 3. – С. 157–161.
9. Kron G. Tensor analysis of networks. J. Wiley & Sons, 1939.
10. Barjis J., Dietz J.L.G. Business process modeling and analysis using GERT networks. Enterprise Information systems. Dordrecht: Springer Netherlands, 2000, pp. 71-80.
11. Zyryanov A.A., Dorrer M.G. The algoritnm of business process model translation into the GERT-network model. Bull. KrasGAU, 2012, no. 12, pp. 13-18.
12. Ambika S. et al. Application of gert analysis in management. Int. J. Latest Eng. Manag. Res, 2018, vol. 03, no. 02, pp. 01-04.
13. Dorrer M., Dorrer A., Zyr,anov A. Numerical modeling of business processes using the apparatus of GERT Networks. Society 5.0, Cyberspace for advanced human-centered society, 2021, pp. 47-55.
14. Phillips D.T., Garcia-Diaz A. Fundamentals of network analysis. Networks. Englewood Cliffs, NJ, Prentice Hall, 1981, 474 p.
15. Pritsker A.A.B. GERT: graphical evaluation and review technique. Santa Monica, CA: RAND Corporation, 1966. 150 p.
16. Доррер М.Г. Реализация цифрового двойника бизнес-процессов на базе системы ELMA / М.Г. Доррер // ИТНОУ информационные технологии в науке, образовании и управлении, 2021. – Т. 1 (17).
17. Dorrer M.G. The digital twin of the business process model. J. Phys. Conf. Ser., 2020, vol. 1679, 032096 p.
18. Dorrer M.G. The prototype of the organizational maturity model's digital twin of an educational institution. J. Phys. Conf. Ser., 2020, vol. 1691, 012121 p.
19. Шибанов А.П. Обобщенные GERT-сети для моделирования протоколов, алгоритмов и программ телекоммуникационных систем. / А.П.Шибанов // Рязанская государственная радиотехническая академия, 2003. – 307 с.
20. Зырянов А.А. Методика вероятностного прогнозирования состояния организационно-технологических систем при помощи формализмов GERT-сетей. / А.А. Зырянов // Сиб. аэрокосм. акад. им. акад. М.Ф. Решетнева, 2013. – 146 с.
21. Милащенко Е.А. Получаем кривую плотности распределения вероятности случайного процесса. Хабр, 2021. – URL: https://habr.com/ru/articles/585232/.
22. Прядченко В. Получаем кривую плотности распределения вероятности… быстрее и точнее. Хабр, 2021. – URL: https://habr.com/ru/articles/587372/.

Доррер Михаил Георгиевич. *К.т.н., СибГУ им. М.Ф. Решетнева, доцент. Направления исследований: искусственные нейронные сети, моделирование бизнес-процессов, имитационное моделирование. AuthorID: 114859, SPIN: 9983-0377, ORCID: 0000-0002-8598-8331, dorrer_mg@sibsau.ru, 660037, Россия, г. Красноярск, пр. Мира, д.82*

UDC 004.852
DOI:10.25729/ESI.2024.34.2.006

Evaluation of the quality of interpolation of the random variable distribution density using an artificial neural network

Mikhail G. Dorrer
Reshetnev Siberian State University of Science and Technology,
Russia, Krasnoyarsk, *dorrer_mg@sibsau.ru*

Abstract. The paper describes the creation and evaluation of the performance of a neural network model of the probability distribution density function of a random variable, given by a set of measurements of a random variable in the absence of the identification stage of the distribution law. The need to solve this problem is caused by limitations introduced into the accuracy of calculating the probability distribution density function of a random variable both by the tabular-histogram method and in the case of applying approaches to the identification of the distribution law. The problem was solved in Python using the TensorFlow neural network library by creating a neural network model based on the Sequential class with fully connected Dense layers, trained on data from numerical differentiation of the random variable distribution function. The accuracy of the forecast was estimated using the Kullback-Leibler distance measure for various ratios of the volume of experimental data and the number of interpolation intervals on synthetic test data generated for 5 laws of distribution - Rayleigh, Weibull, gamma, exponential and normal (Gaussian). To assess the predictive ability of the approach when testing the interpolator, random variable samples shifted relative to those used in training were used. The proposed solution showed a significantly higher accuracy in calculating the values of the distribution density of a random variable compared to the histogram method. The developed approach will be implemented in the modeling part of the digital twin of a business process based on the mathematical apparatus of stochastic GERT networks.

Keywords: stochastic models, random variable distribution density, neural networks, digital process models

References

1. Ahmad T., Van Looy A. Business process management and digital innovations: a systematic literature review. Sustainability, 2020, vol. 12, no. 17, 6827 p.
2. Dorrer M. Implementation of the digital twin of educational business processes on the ELMA BPMS platform. Proceedings of the II International scientific conference on advances in science, engineering and digital education: (ASEDU-II 2021) AIP Conference Proceedings, 2022, 040088 p.
3. Hofacker I., Vetschera R. Algorithmical approaches to business process design. Comput. Oper. Res., 2001, vol. 28, no. 13, pp. 1253-1275.
4. Powell S.G., Schwaninger M., Trimble C. Measurement and control of business processes. Syst. Dyn. Rev., 2001, vol. 17, no. 1, pp. 63-91.
5. Valiris G., Glykas M. Business analysis metrics for business process redesign. Bus. Process Manag. J., 2004, vol. 10, no. 4, pp. 445-480.
6. Kalyanov G.N. Teoriya i praktika reorganizatsii biznes-protsessov [Theory and practice of business process reorganization]. Moscow: SINTEG, 2000. 203 p.
7. Dorrer M.G., Kurohtin V.V. Optimizatsiya rascheta invariantov seti Petri v ramkakh zadachi formirovaniya skenariev integratsionnogo testirovaniya [Optimization of the calculation of Petri net invariants within the task of forming scenarios of integration testing]. Modelirovanie i analiz informatsionnykh sistem [Modeling and analysis of information systems], 2012, vol. 19, no. 4, pp. 128-143.
8. Petrov M.I., Trenogiy I.G., Belovatyy E.A. Ispol'zovanie tenzornoy metodologii analiza sistem dlya opisaniya biznes-protsessov predpriyatiya. Sredstva avtomatizatsii [Using tensor methodology for system analysis to describe enterprise business processes. Automation tools]. Sibirskiy aerokosmicheskiy zhurnal [Siberian aerospace journal], 2008, vol. 3, pp. 157-161.
9. Kron G. Tensor analysis of networks. J. Wiley & Sons, 1939.
10. Barjis J., Dietz J.L.G. Business process modeling and analysis using GERT networks. Enterprise Information systems. Dordrecht: Springer Netherlands, 2000, pp. 71-80.
11. Zyryanov A.A., Dorrer M.G. The algoritnm of business process model translation into the GERT-network model. Bull. KrasGAU, 2012, no. 12, pp. 13-18.
12. Ambika S. et al. Application of gert analysis in management. Int. J. Latest Eng. Manag. Res, 2018, vol. 03, no. 02, pp. 01-04.

13. Dorrer M., Dorrer A., Zyryanov A. Numerical modeling of business processes using the apparatus of GERT Networks. Society 5.0, Cyberspace for advanced human-centered society, 2021, pp. 47-55.
14. Phillips D.T., Garcia-Diaz A. Fundamentals of network analysis. Networks. Englewood Cliffs, NJ: Prentice Hall, 1981, 474 p.
15. Pritsker A.A.B. GERT: graphical evaluation and review technique. Santa Monica, CA: RAND Corporation, 1966. 150 p.
16. Dorrer M.G. Realizatsiya tsifrovogo dvoynika biznes-protsessov na baze sistemy ELMA [Implementation of the digital twin of business processes on the ELMA system]. ITNOU informatsionnye tekhnologii v nauke, obrazovanii i upravlenii [ITNOU information technologies in science, education and management], 2021, vol. 1 (17).
17. Dorrer M.G. The digital twin of the business process model. J. Phys. Conf. Ser., 2020, vol. 1679, 032096 p.
18. Dorrer M.G. The prototype of the organizational maturity model's digital twin of an educational institution. J. Phys. Conf. Ser., 2020, vol. 1691, 012121 p.
19. Shibanov A.P. Obobshchennye GERT-seti dlya modelirovaniya protokolov, algoritmov i programm telekommunikatsionnykh sistem [Generalized GERT networks for modeling protocols, algorithms and programs of telecommunication systems]. Ryazanskaya gosudarstvennaya radiotekhnicheskaya akademiya [Ryazan State radio engineering academy], 2003, 307 p.
20. Zyryanov A.A. Metodika veroyatnostnogo prognozirovaniya sostoyaniya organizatsionno-tekhnologicheskikh sistem pri pomoshchi formalizmov GERT-setey [Methodology for probabilistic forecasting of the state of organizational and technological systems using GERT network formalisms]. Sibirskiy aerokosmicheskiy akademiy im. akad. M.F. Reshetneva [Siberian aerospace academy named after M.F. Reshetnev], 2013, 146 p.
21. Milashchenko E.A. Poluchaem krivuyu plotnosti raspredeleniya veroyatnosti sluchaynogo protsessa [Getting the probability density curve of a random process]. Habr, 2021, available at: https://habr.com/ru/articles/585232/.
22. Pryadchenko V. Poluchaem krivuyu plotnosti raspredeleniya veroyatnosti... bystree i tochnee [Getting the probability density curve... faster and more accurate]. Habr, 2021, available at: https://habr.com/ru/articles/587372/.

Dorrer Mikhail Georgievich. *Ph.D., Reshetnev university, docent. Research direction: artificial neural networks, business process modeling, simulation modeling. AuthorID*: *114859, SPIN*: *9983-0377, ORCID: 0000-0002-8598-8331, dorrer_mg@sibsau.ru, 660037, Russian Federation, Krasnoyarsk, Krasnoyarsky Rabochy Ave.*

Статья поступила в редакцию 08.12.2023; одобрена после рецензирования 30.05.2024; принята к публикации 03.06.2024.

The article was submitted 12/08/2023; approved after reviewing 05/30/2024; accepted for publication 06/03/2024.

УДК 004.89
DOI:10.25729/ESI.2024.34.2.007

Использование диаграмм переходов состояний для автоматизированного создания баз знаний

Дородных Никита Олегович, Юрин Александр Юрьевич

Институт динамики систем и теории управления имени В.М. Матросова СО РАН,
Россия, Иркутск, *nikidorny@icc.ru*

Аннотация. Построение баз знаний в форме продукций, онтологий или графов знаний продолжает оставаться достаточно трудоемкой задачей в рамках разработки различных предметно-ориентированных интеллектуальных систем. В данной статье рассмотрены подход и программное средство автоматизации создания баз знаний на основе анализа и преобразования концептуальных моделей в виде диаграмм переходов состояний. Подход основан на выделении структурных элементов диаграмм и их трансформации в конструкции целевого языка представления знаний. Приведено описание основных этапов подхода, анализируемых конструкций рассматриваемого формата диаграмм переходов состояний, а также реализация подхода в форме веб-ориентированной программной системы – Knowledge Modeling System (KMS). Представлен иллюстративный пример преобразования диаграмм переходов состояний для формирования плана анализа отказа технической системы.

Ключевые слова: инженерия знаний, получение знаний, база знаний, онтология, диаграмма переходов состояний, трансформация моделей, генерация кода, продукции

Цитирование: Дородных Н.О. Использование диаграмм переходов состояний для автоматизированного создания баз знаний / Н.О. Дородных, А.Ю. Юрин // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 69-81. – DOI:10.25729/ESI.2024.34.2.007.

Введение. Создание интеллектуальных систем является перспективным направлением в сфере информационных технологий. Системы подобного типа способны обрабатывать большие объемы данных, анализировать информацию и принимать решения на основе полученных результатов, что позволяет значительно ускорить процессы в различных отраслях экономики, повысить эффективность работы предприятий и улучшить качество жизни людей. Кроме того, интеллектуальные системы могут быть использованы для решения сложных и слабо-формализованных задач в области медицины [1-3], энергетики [4, 5], техники [6-8], природной и техногенной безопасности [9-11]. Как правило, в основе подобного рода систем лежит База Знаний (БЗ), содержащая формализованные знания предметной области (например, в форме продукций, онтологий или графов знаний), полученные в результате практической деятельности и профессионального опыта, позволяющие специалистам ставить и решать задачи в этой области [12]. Однако, задача создания эффективных методов и средств разработки БЗ интеллектуальных систем не решена в полном объеме до сих пор, поэтому исследования, направленные на развитие методов обработки информации для создания элементов систем, основанных на знаниях, при решении практических задач в различных предметных областях являются актуальными.

На сегодняшний день для построения БЗ используют различные специализированные редакторы, которые позволяют представлять формализованное описание понятий предметной области и структур БЗ на определенном языке представления знаний (ЯПЗ) [13]. Однако такие системы обладают низкой интеграционной способностью со средствами визуального моделирования и модулями интерпретации знаний, в лучшем случаев поддерживая один определенный ЯПЗ. Более того, данные редакторы направлены на хорошо подготовленных специалистов (например, инженеров по знаниям, программистов) и не ориентированы на экспертов предметной области, которые являются основным источником знаний и опыта. Это порождает проблему переноса предметных неформализованных знаний (компетенций) от эксперта через инженера по знаниям в БЗ. Для преодоления данной проблемы разрабатываются разные методики и средства. В частности, очень часто на стадиях извлечения

и структурирования знаний активно используются различные концептуальные (информационные) модели (например, концепт-карты, диаграммы Исикавы, деревья событий или отказов, семантические модели), имеющие общесистемную направленность и ориентированные на систематизацию знаний или поддержку принятия решений [14]. Эти модели являются удобным и понятным для эксперта способом представления знаний. К таким моделям можно также отнести диаграммы переходов состояний (*state transition diagram* или *statechart diagram*) [15], используемые для моделирования поведения системы в зависимости от ее состояний и внешних воздействий. Каждая подобная модель представляет собой граф, где *узлы* – это состояния системы, а *дуги* (*переходы*) – это события или действия, которые приводят к изменению состояния системы. На диаграмме также могут быть указаны условия, выполняемые при каждом переходе. Диаграмма Переходов Состояний (ДПС) используется для проектирования и анализа сложных систем, таких, как программное обеспечение, аппаратные устройства, бизнес-процессы и т.д. Она помогает понять логику работы системы и выявить возможные ошибки в ее работе. Также ДПС может использоваться для документирования уже существующей системы или процесса. Это позволяет легко описывать и передавать знания о работе системы другим специалистам. В целом, ДПС является очень полезным инструментом для любого проекта или задачи, где необходимо четко определить последовательность действий и состояний системы.

Для построения ДПС существуют различные программные средства, выполняющие роль визуальных редакторов и осуществляющие расчеты на основе обработки данных диаграмм. Редакторы обычно представлены как в виде настольных приложений, так и веб-сервисов, в частности:

- Flexberry Designer [16];
- Visual Paradigm Online [17];
- Enterprise Architect [18];
- MATLAB & Simulink – State Diagram [19];
- EASE: State diagram editor [20];
- Draw.io [21];
- Borland CaliberRM [22].

Однако эти системы, хоть и позволяют разрабатывать визуальные диаграммы разной степени абстракции, соответствующие знаниям эксперта, но не предназначены для разработки БЗ и поэтому не могут обеспечить полноту процесса разработки от моделей предметной области до формализованного кода БЗ на определенном ЯПЗ. Эта особенность затрудняет практическое использование построенных диаграмм для создания БЗ при разработке интеллектуальных систем.

Таким образом, целью данной работы является повышение эффективности разработки БЗ на основе анализа и преобразования ДПС. Для этого предлагаются новый подход и программная система – Knowledge Modeling System (KMS) [23], основное назначение которой – поддержка моделирования знаний экспертов в различных предметных визуальных нотациях, в том числе в виде ДПС, и синтез кодов БЗ на основе построенных диаграмм. В качестве целевых формализмов представления знаний выбраны продукции, в частности, ЯПЗ CLIPS (C Language Integrated Production System) [24], таблицы решений (*decision table*) [25] и онтологии, в частности, ЯПЗ OWL (Web Ontology Language) [26]. Данный подход использовался для прототипирования БЗ для решения задач в области техногенной и природной безопасности.

1. Постановка задачи. Задача анализа ДПС и их преобразования в БЗ может быть сведена к задаче трансформации моделей [27]. Трансформация моделей является одной из основных составляющих модельно-ориентированного подхода к разработке программного

обеспечения (*Model-Driven Engineering*) [28]. В общем случае, под трансформацией моделей понимается процесс автоматической генерации целевой модели по исходной модели, в соответствии с некоторым набором правил трансформации. При этом под правилом трансформации подразумевается описание того, как одна или более конструкций на исходном языке моделирования может быть преобразована (отображена) в одну или более конструкций на целевом языке моделирования [29].

В рамках предлагаемого подхода постановку задачи можно формализовать следующим образом:

$$T: CM^{STD} \rightarrow KB, \quad (1)$$

где T – это оператор преобразования исходной концептуальной модели в целевую БЗ; CM^{STD} – это исходная концептуальная модель в виде ДПС (*STD*); KB – это целевая БЗ, при этом: $KB = \langle Code^{CLIPS}, Code^{DT}, Code^{OWL} \rangle$.

Для решения поставленной задачи предлагается специализированный подход, реализованный в виде веб-ориентированной программной системы (KMS), которая обеспечивает поддержку визуального моделирования ДПС и их трансформацию в код БЗ на целевом ЯПЗ. Далее подробнее рассмотрим данный подход и систему.

2. Предлагаемый подход.

2.1. Структура диаграммы переходов состояний. ДПС являются одним из мощных инструментов моделирования и документирования поведения системы. Они представляют собой графическое изображение последовательности состояний системы (*узлов* или *вершин*) и переходов между ними (*стрелки* или *дуги*) при определенных событиях (*условиях*). Таким образом, основная идея ДПС заключается в том, что система имеет некоторые возможные состояния, и при определенных условиях она может перейти из одного состояния в другое. Каждый переход описывается событием, которое вызывает изменение состояния системы. При этом у состояний и переходов могут быть определенные свойства (*характеристики*). Также на данных диаграммах в зависимости от особенностей реализации (например, в UML) изображают начало и конец переходов состояний. Абстрактный пример ДПС, содержащий все основные элементы, представлен на рисунке 1.

Основные особенности ДПС:

- позволяют легко представить поведение системы в виде последовательности состояний и переходов между ними;
- могут использоваться для моделирования как простых, так и сложных систем;
- помогают выявить возможные ошибки в работе системы на этапе ее проектирования;
- используются для документирования работы уже существующих систем, чтобы облегчить понимание их работы и упростить процесс отладки;
- могут быть использованы для создания тестовых сценариев и проверки работоспособности системы или оценки риска.

ДПС широко применяются в различных областях, таких, как: разработка и тестирование программного обеспечения, проектирование и разработка аппаратных (электронных) устройств, автоматизация бизнес-процессов, моделирование работы сложных технических систем и др.

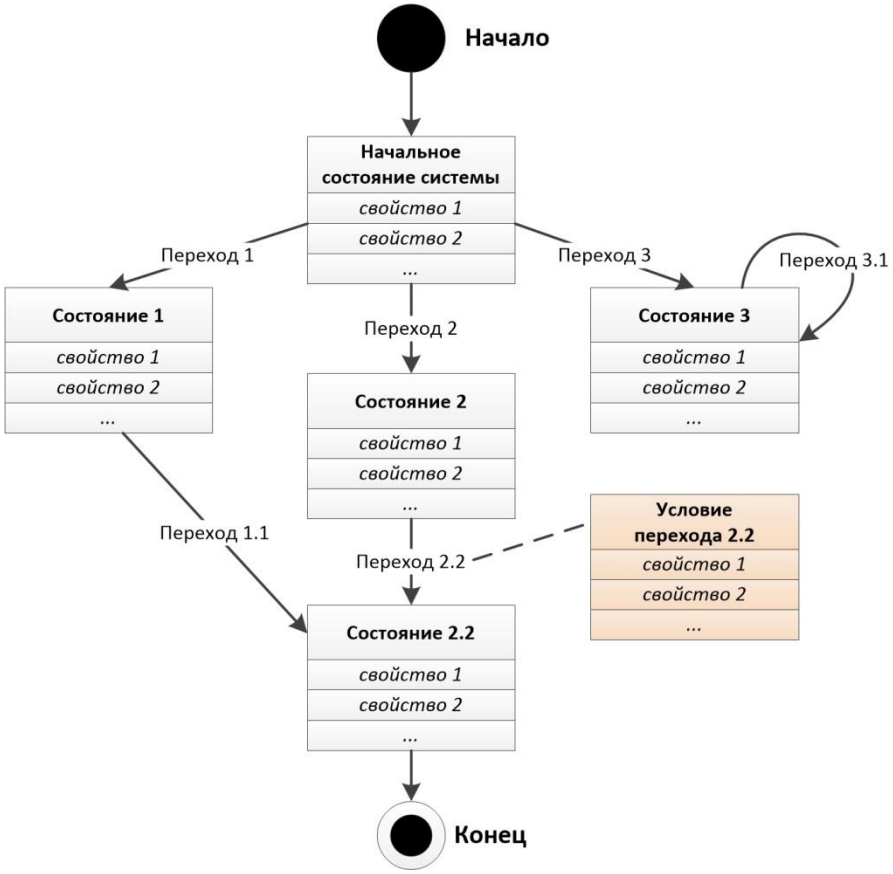


Рис. 1. Абстрактный пример ДПС

Метамодель (*abstract syntax*) ДПС, определяющая в абстрактной форме основные концепты, из которых состоит ДПС, представлена на рисунке 2. Построенная метамодель соответствует мета-метамодели Ecore [30] и в дальнейшем используется в качестве исходной метамодели при разработке правил трансформации, описывающих соответствия между элементами данной метамодели и целевой метамодели БЗ.

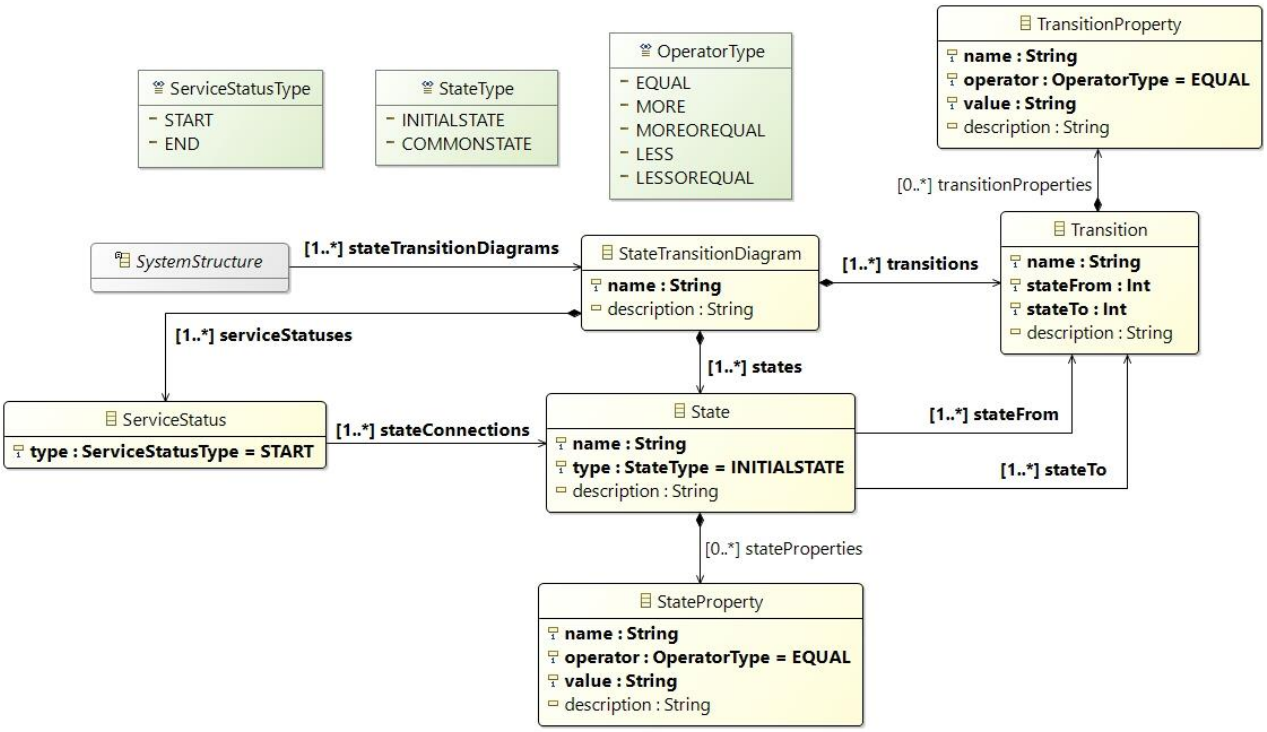


Рис. 2. Метамодель ДПС

На настоящее время не существует единого общепринятого формата или стандарта текстовой нотации для представления ДПС. Учитывая это, а также исходя из специфики ДПС, принято решение о разработке собственного формата (*спецификации*) сериализации ДПС с использованием языка XML, который является универсальным и наиболее распространенным способом для интеграции программных систем и обеспечения обмена информацией между приложениями.

Описание разработанного XML-формата (*concrete syntax*) ДПС, определяющее основные XML-конструкции, представлено в Таблице 1.

Таблица 1. XML-формат представления ДПС

Элементы XML-структуры ДПС	**Описание**
Diagram	Общее описание ДПС, содержащее информацию об исследуемой системе.
State	Информация о состоянии. Также содержит тип состояния с возможными значениями: начальное состояние (*initial state*) и обычное состояние (*common state*).
StateProperty	Информация о свойстве (характеристике) состояния. Также содержит оператор и возможное значение свойства.
Transition	Информация о переходе одного состояния в другое.
TransitionProperty	Информация о свойстве (условии) перехода. Также содержит оператор и возможное значение условия перехода.

3.2. Основные этапы подхода. Для решения поставленной задачи (1) специализируем обобщенный алгоритм трансформации концептуальных моделей в код БЗ, описанный в [31]. Таким образом, основной задачей является преобразование (отображение) элементов исходной ДПС, представленной в формате XML, в конструкции целевого представления БЗ, что может быть представлено в виде последовательности действий (рис. 3).

На этапе 1 при помощи визуального редактора – State Transition Diagram Editor (STDE) [32], входящего в состав системы KMS, пользователь (эксперт предметной области) строит ДПС, описывающую последовательность переходов состояний некоторой системы. На этапе 2 построенная диаграмма представляется (сериализуется) в формате XML с использованием разработанной спецификации. На этапе 3 процесса анализа XML-структуры исходной ДПС выделяются элементы диаграммы и их отношения. Далее (этап 4) на их основе автоматически формируется (генерируется) либо модель продукций, либо модель онтологии (пользователь выбирает необходимую). Данные модели выступают в качестве универсального абстрактного средства представления знаний, не зависящего от целевого ЯПЗ (например, CLIPS, Jess, Drools, SWRL, OWL, RDF) или источника моделей.

При помощи специальной графической нотации – Rule Visual Modeling Language (RVML) [33] предоставляется возможность визуализации, модификации (проверки) полученных продукций (этап 5.1). При этом на этапе 6.1 происходит генерация либо целевой таблицы решений в формате CSV, либо кода БЗ в формате CLIPS на основе сформированной модели продукций.

Возможность модификации онтологической модели осуществляется средствами системы – Knowledge Base Development System (KBDS) [34] (этап 5.2). А на этапе 6.2 происходит генерация кода онтологической БЗ в формате OWL2 DL.

Таким образом, уточним оператор преобразования концептуальной модели из (1):

$$T = \langle T_{CM-RM}, T_{RM-RKB}, T_{CM-OM}, T_{OM-OKB} \rangle,$$
$$T_{CM-RM}: CM_{XML}^{STD} \to RM,\ T_{RM-RKB}: RM \to RKB,\ RKB = \langle Code^{CLIPS}, Code^{DT} \rangle,$$
$$T_{CM-OM}: CM_{XML}^{STD} \to OM,\ T_{OM-OKB}: OM \to Code^{OWL},$$

где T_{CM-RM} – оператор преобразования ДПС в модель продукций; T_{RM-RKB} – оператор преобразования модели продукций в код БЗ на ЯПЗ CLIPS или таблицу решений в формате CSV; CM_{XML}^{STD} – представление ДПС в XML-формате; RM – представление полученных знаний в виде модели продукций; $Code^{CLIPS}$ – код БЗ на ЯПЗ CLIPS; $Code^{DT}$ – таблица решений в формате CSV; T_{CM-OM} – оператор преобразования ДПС в модель онтологии; T_{OM-OKB} – оператор преобразования модели онтологии в код БЗ на ЯПЗ OWL2 DL; OM – представление полученных знаний в виде модели онтологии; $Code^{OWL}$ – код БЗ на ЯПЗ OWL2 DL.

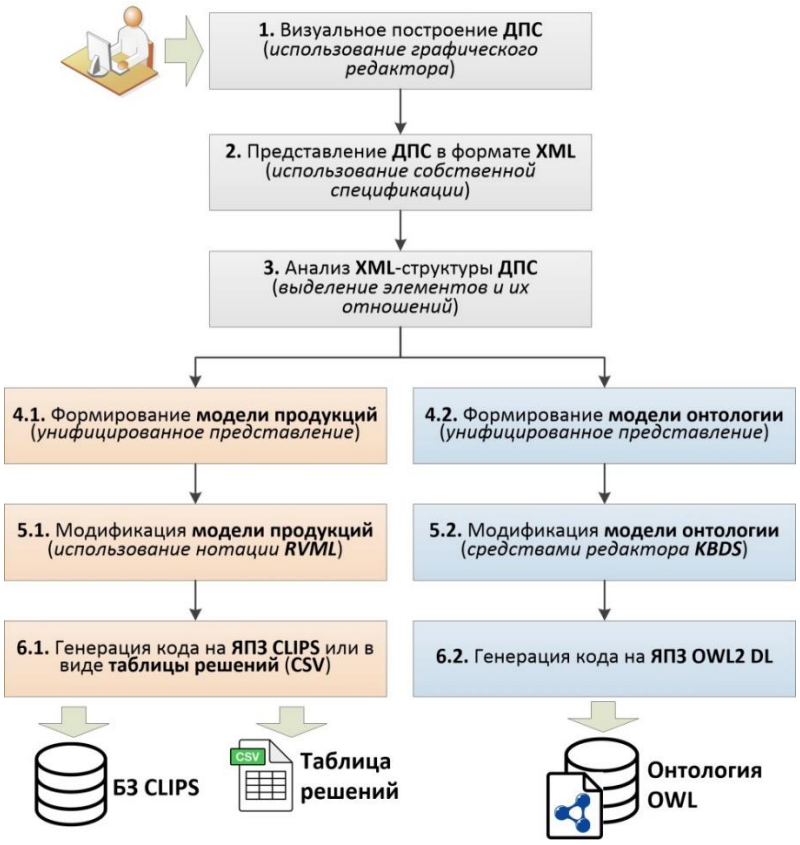


Рис. 3. Основные этапы предлагаемого подхода

3. Программная реализация. Разработанный подход реализован в виде веб-ориентированной программной системы – KMS [26], разработанной на языке PHP с использованием фреймворка Yii2 и паттерна проектирования «Model-View-Controller». Данное средство обладает клиент-серверной архитектурой и ориентировано на непрограммирующих пользователей (например, предметных экспертов, аналитиков данных).

В состав системы KMS входят два визуальных редактора:

– *Extended Event Tree Editor (EETE)* – это графический редактор для визуального моделирования (проектирования) семантических древовидных структур в виде классических или расширенных деревьев событий. Данные диаграммы могут быть также преобразованы в коды БЗ [35];
– *State Transition Diagram Editor (STDE)* – это графический редактор для визуального моделирования (проектирования) ДПС [32].

Основные функции KMS:

– создание, редактирование, просмотр и удаление пользователей;
– аутентификация (вход и выход) и авторизация пользователей (определение прав доступа);
– создание, редактирование, просмотр и удаление проектов, в рамках которых ведется работа с различными диаграммами;

– создание, редактирование, просмотр и удаление ДПС, в рамках определенного проекта;
– создание, редактирование, просмотр и удаление диаграмм деревьев событий (как классических, так и расширенных), в рамках определенного проекта;
– импорт и экспорт диаграмм в виде сериализованных файлов в формате XML;
– автоматическое преобразование диаграмм в модель продукций и онтологий;
– генерация (экспорт) построенных моделей в файлы форматов ЯПЗ CLIPS и OWL2 DL и электронные таблицы формата CSV.

4. Пример практического применения. Рассмотрим пример автоматизированного формирования БЗ на примере фрагмента ДПС. В [7] решалась задача создания программы анализа отказа технической системы для подсистемы-планировщика. В частности, на основе модели динамики технического состояния [7], алгоритма анализа отказа и с учетом структуры программы и алгоритма планирования с помощью специального редактора STDE [35] была построена визуальная модель БЗ в форме ДПС (Рис. 4).

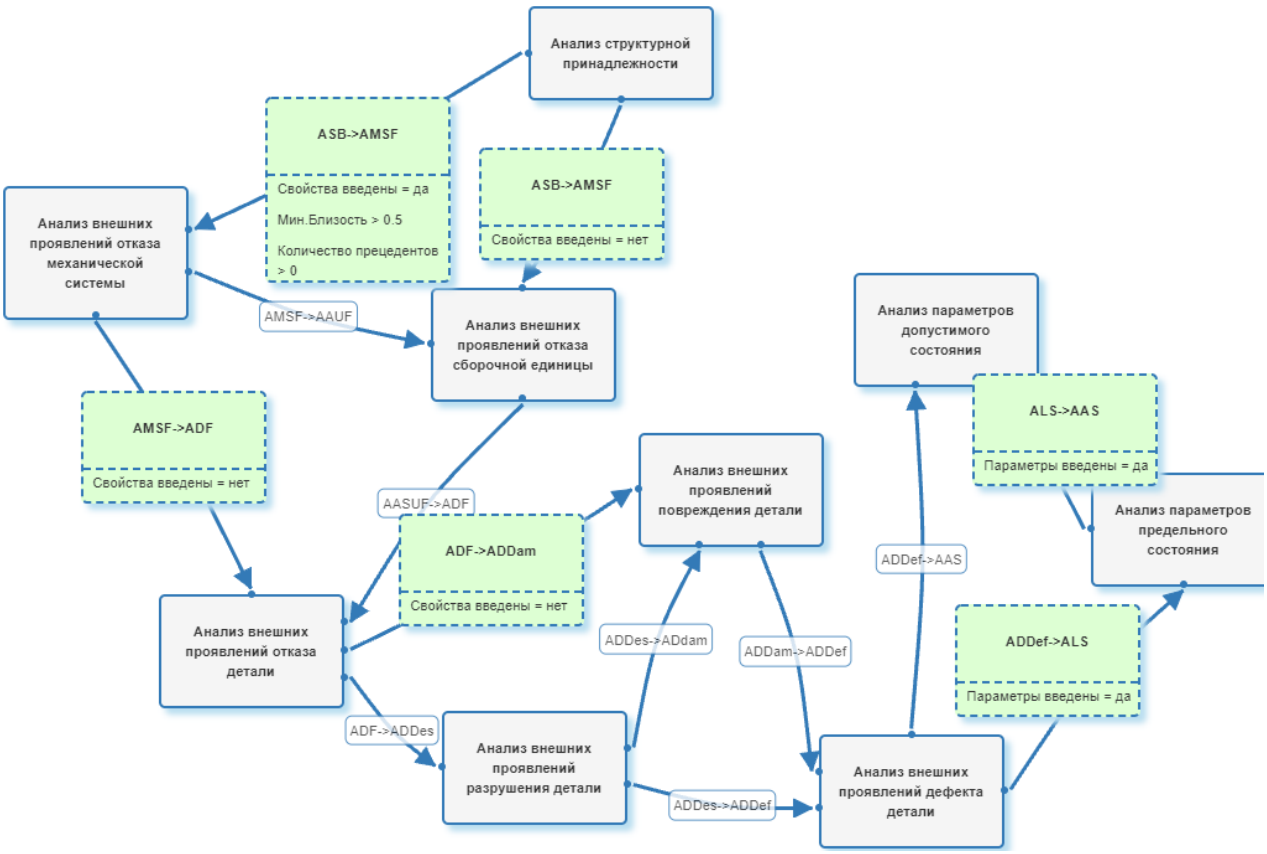


Рис. 4. Пример фрагмента ДПС, описывающей алгоритм анализа отказов

Графическому представлению ДПС (рис. 4) соответствует фрагмент XML-документа (листинг 1). При этом выделены ключевые конструкции (табл. 1), на основе которых осуществлялось извлечение необходимых элементов ДПС.

Листинг 1. Фрагмент XML-кода ДПС, описывающего алгоритм анализа отказов

```
<?xml version="1.0" encoding="UTF-8"?>
<Diagram id="3" name="Д 6.3.5 База знаний планировщика анализа отказа"
description="Диаграмма для базы знаний планировщика анализа отказов">
  <State id="26" name="Анализ структурной принадлежности" type="Initial state"
description=""/>
  <State id="27" name="Анализ внешних проявлений отказа механической системы"
type="State" description=""/>
…
  <Transition id="28" name="ASB-&gt;AMSF" state-from="26" state-to="27"
description="">
```

```
        <TransitionProperty id="30" name="Свойства введены" operator="=" value="да"
description=""/>
        <TransitionProperty id="32" name="Мин.Близость" operator="&gt;" value="0.5"
description=""/>
        <TransitionProperty id="40" name="Количество прецедентов" operator="&gt;"
value="0" description=""/>
       </Transition>
     …
```

На основе извлеченных элементов ДПС сгенерирована модель продукций и онтологии. Далее приводится фрагмент полученной модели продукций в форме диаграммы RVML (рис. 5).

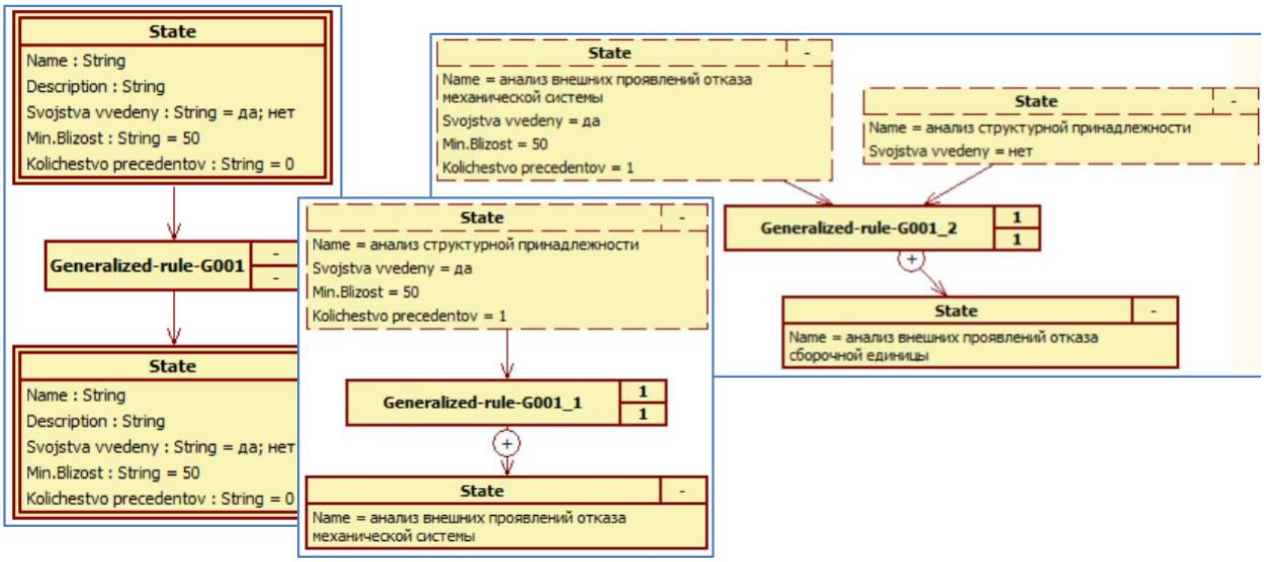


Рис. 5. Пример фрагмента полученной модели продукций в виде диаграммы RVML: шаблон правила и пример двух конкретных правил

Полученная модель продукций была преобразована в код БЗ на ЯПЗ CLIPS с уточнением знаков операторов, определяющих ограничения на значения слотов, и добавлением дополнительных условий. Пример одного из полученных правил приводится в листинге 2. На рисунке 6 также приводится пример фрагмента полученной БЗ в виде таблицы решений в формате CSV.

Листинг 2. Фрагмент CLIPS-кода правила из алгоритма анализа отказов

```
(defrule Generalized-rule-G001-2 "Description of the rule: Generalized-rule-
G001 2"
(declare (salience 1))
 (State ;State
 (Name "АНАЛИЗ ВНЕШНИХ ПРОЯВЛЕНИЙ ОТКАЗА МЕХАНИЧЕСКОЙ СИСТЕМЫ")
 (Svojstva-vvedeny "ДА")
 (Min.Blizost >50)
 (Kolichestvo-precedentov >0)
)
(State ;State
 (Name "АНАЛИЗ СТРУКТУРНОЙ ПРИНАДЛЕЖНОСТИ")
 (Svojstva-vvedeny "НЕТ")
)
(State ;State
 (Name "АНАЛИЗ ВНЕШНИХ ПРОЯВЛЕНИЙ ОТКАЗА МЕХАНИЧЕСКОЙ СИСТЕМЫ")
 (Svojstva-vvedeny "ДА")
 (Min.Blizost <50)
)
=>
(assert
(State ;State
 (Name "АНАЛИЗ ВНЕШНИХ ПРОЯВЛЕНИЙ ОТКАЗА СБОРОЧНОЙ ЕДИНИЦЫ")
```

```
 (GUI-form-name "AAUF")
 ))
)
```

	State	Stat	State	State::Name	#State::Name
2	да	0.5	0	Анализ структурной принадлежности	Анализ внешних проявлений отказа механической системы
3	нет			Анализ структурной принадлежности	Анализ внешних проявлений отказа сборочной единицы
4	да	0.5	0	Анализ внешних проявлений отказа механической системы	Анализ внешних проявлений отказа сборочной единицы
5	нет			Анализ внешних проявлений отказа механической системы	Анализ внешних проявлений отказа детали
6	да	0.5	0	Анализ внешних проявлений отказа сборочной единицы	Анализ внешних проявлений отказа детали
7	да	0.5	0	Анализ внешних проявлений отказа детали	Анализ внешних проявлений разрушения детали
8	нет			Анализ внешних проявлений отказа детали	Анализ внешних проявлений повреждения детали
9	да	0.5	0	Анализ внешних проявлений разрушения детали	Анализ внешних проявлений повреждения детали
10	нет			Анализ внешних проявлений разрушения детали	Анализ внешних проявлений дефекта детали
11	да	0.5	0	Анализ внешних проявлений повреждения детали	Анализ внешних проявлений дефекта детали
12	да			Анализ внешних проявлений дефекта детали	Анализ параметров предельного состояния
13	нет			Анализ внешних проявлений дефекта детали	Анализ параметров допустимого состояния

Рис. 6. Фрагмент полученной таблицы решений в формате CSV

В свою очередь, полученная модель онтологии была преобразована в код онтологической БЗ в формате OWL2 DL (Листинг 3). При этом все состояния интерпретировались как *классы*, а переходы как *объектные-свойства* (*ObjectProperty*). Свойства (*характеристики*) переходов выражались через *свойства-значения* (*DatatypeProperty*), которые задаются для специального служебного класса перехода.

Листинг 3. Фрагмент OWL-кода БЗ, описывающего алгоритм анализа отказов

```
<owl:Class rdf:ID="Состояние"/>
<owl:Class rdf:ID="Переход"/>
<owl:Class rdf:ID="АнализСтруктурнойПринадлежности">
      <rdfs:subClassOf rdf:resource="#Состояние"/>
</owl:Class>
<owl:Class rdf:ID="АнализВнешнихПроявленийОтказаМеханической">
      <rdfs:subClassOf rdf:resource="#Состояние"/>
</owl:Class>
<owl:DatatypeProperty rdf:ID="минБлизость">
      <rdfs:domain rdf:resource="#Переход"/>
      <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#double"/>
</owl:DatatypeProperty>
<owl:ObjectProperty rdf:ID="имеетПереход">
      <rdfs:domain rdf:resource="#Состояние"/>
      <rdfs:range rdf:resource="#Переход"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="ASB->AMSF">
      <rdfs:subPropertyOf rdf:resource="#имеетПереход"/>
      <rdfs:domain rdf:resource="#АнализСтруктурнойПринадлежности"/>
      <rdfs:range rdf:resource="#АнализВнешнихПроявленийОтказаМеханической"/>
</owl:ObjectProperty>
…
```

Заключение. Создание предметно-ориентированных интеллектуальных систем является одним из ключевых направлений развития современных информационных технологий, при этом автоматизация разработки БЗ для этих систем является одной из актуальных и перспективных областей в сфере искусственного интеллекта, особенно, когда речь идет об использовании уже накопленной информации, представленной, в частности, в форме концептуальных моделей.

В статье рассмотрен новый подход к автоматизации разработки БЗ на основе ДПС, основанный на анализе структурных элементов ДПС и их преобразовании в конструкции целевого продукционного или онтологического ЯПЗ. Такой подход позволяет, с одной стороны, избежать ошибок программирования на этапе формализации знаний за счет автоматической кодогенерации, а также сократить время, затрачиваемое на разработку БЗ, с другой – непосредственно вовлечь экспертов предметной области в процесс разработки прототипов БЗ, позволяя им создавать программный код, оперируя понятными для них

предметно-ориентированными моделями. Предлагаемый подход реализован в виде веб-ориентированной программной системы KMS, которая может быть использована в качестве инструмента для моделирования предметных знаний и получения сгенерированного кода прототипов БЗ, которые в дальнейшем могут быть доработаны в специализированных средствах и использованы в различных интеллектуальных системах.

Благодарности. Работа выполнена в рамках госзадания Минобрнауки России по проекту «Методы и технологии облачной сервис-ориентированной цифровой платформы сбора, хранения и обработки больших объёмов разноформатных междисциплинарных данных и знаний, основанные на применении искусственного интеллекта, модельно-управляемого подхода и машинного обучения» (№ гос. регистрации: 121030500071-2).

Список источников

1. Грибова В.В. Семантические модели для оценки влияния комплекса факторов на развитие заболеваний / В.В. Грибова, Д.Б. Окунь, Е.А. Шалфеева // Онтология проектирования, 2021. – Т.11. – №4(42). – С.464-477.
2. Ясницкий Л.Н. Нейроэкспертная система диагностики, прогнозирования и управления рисками сердечнососудистых заболеваний / Л.Н. Ясницкий, Ф.М. Черепанов // Прикладная математика и вопросы управления, 2018. – №3. – С.107-126.
3. Alam F., Giglou H.B., Malik K.M. Automated clinical knowledge graph generation framework for evidence based medicine. Expert systems with applications, 2023, vol .233, 120964.
4. Воропай Н.И. ИТ-инфраструктура для построения интеллектуальных систем управления развитием и функционированием систем энергетики на основе цифровых двойников и цифровых образов / Н.И. Воропай, Л.В. Массель, И.Н. Колосок, А.Г. Массель // Известия Российской академии наук. Энергетика, 2021. – №1. – С.3-13.
5. Zhang L., Su H., Zio E., Jiang L., Fan L., Zhang J. A graph structure feature-based framework for the pattern recognition of the operational states of integrated energy systems. Expert systems with applications, 2023, vol.213, 119039.
6. Берман А.Ф. Технология создания основанных на знаниях систем для исследования динамики состояния технических объектов / А.Ф. Берман, Н.О. Дородных, О.А. Николайчук, А.Ю. Юрин // Вычислительные технологии, 2023. – Т.28. – №4. – С.94-108.
7. Берман А.Ф. Интеллектуальная информационная система анализа отказов / А.Ф. Берман, О.А. Николайчук, А.Ю. Юрин // Проблемы машиностроения и надежности машин, 2012. – №4. – С.88-96.
8. Zhang J., Pu X., Zhao R., Li J., Nie Z. Implicit contradictions identification and solution process model for complex technical systems. Computers & industrial engineering, 2023, vol.177, 108822.
9. Ноженкова Л.Ф. Создание комплексной системы безопасности региона на основе системной интеграции технологий / Л.Ф. Ноженкова, В.В. Ничепорчук, А.И. Ноженков // Информатизация и связь, 2013. – №2. – С. 122-124.
10. Nicheporchuk V., Favorskaya M.N., Gryazin I. Framework for intelligent wildlife monitoring. Smart innovation, Systems and technologies, 2020, vol.193, pp. 167-177.
11. Nikolaychuk O.A., Pestova J.V., Yurin A.Yu. Wildfire susceptibility mapping in Baikal natural territory using random forest. Forests, 2024, vol.15, no.1, 170.
12. Гаврилова Т.А. Инженерия знаний. Модели и методы / Т.А. Гаврилова, Д.В. Кудрявцев, Д.И. Муромцев. – СПб.: Лань, 2016. – 324 с.
13. Джарратано Дж. Экспертные системы: принципы разработки и программирования, 4-е издание: пер. с англ. / Дж. Джарратано, Г. Райли. – М.: Вильямс, 2007. – 1152 с.
14. Гаврилова Т.А. Визуально-аналитическое мышление и интеллект-карты в онтологическом инжиниринге / Т.А. Гаврилова, Э.В. Страхович // Онтология проектирования, 2020. – Т.10. – №1(35). – С.87-99.
15. Хопкрофт Дж. Введение в теорию автоматов, языков и вычислений, 2-е издание: пер. с англ. / Дж. Хопкрофт, Р. Мотвани, Дж. Ульман. – М.: Вильямс, 2008. – 528 с.
16. Flexberry Designer. Available at: https://flexberry.net/ru/developers-flexberry-designer.html (accessed: 01/21/2024).
17. Visual paradigm online. Available at: https://online.visual-paradigm.com/ (accessed: 01/21/2024).
18. Enterprise Architect. Available at: https://sparxsystems.com/ (accessed: 01/21/2024).
19. MATLAB & Simulink – State Diagram. Available at: https://www.mathworks.com/discovery/state-diagram.html (accessed: 01/21/2024).

20. EASE: State diagram editor. Available at: https://www.hdlworks.com/products/ease/state_diagram.html (accessed: 01/21/2024).
21. Draw.io. Available at: https://app.diagrams.net/ (accessed: 01/21/2024).
22. Borland CaliberRM. Available at: https://borland-caliber.software.informer.com/11.4/ (accessed: 01/21/2024).
23. Knowledge Modeling System (KMS). Available at: http://kms.knowledge-core.ru/ (accessed: 01/21/2024).
24. CLIPS: A tool for building expert systems. Available at: https://www.clipsrules.net/ (accessed: 01/21/2024).
25. Еремеев А.П. Развитие аппарата таблиц решений в контексте создания гибридных интеллектуальных систем / А.П. Еремеев // Гибридные и синергетические интеллектуальные системы: матер. III Всерос. Поспеловской конф. с междунар. участием, 2016. – С.121-132.
26. OWL 2 web ontology language document overview (Second Edition). 2012, available at: https://www.w3.org/TR/owl2-overview (accessed: 01/21/2024).
27. Mens T., Gorp P.V. A taxonomy of model transformations. Electronic notes in theoretical computer science, 2006, vol.152, pp. 125-142.
28. da Silva A.R. Model-driven engineering: A survey supported by the unified conceptual model. Computer languages, systems & structures, 2015, vol.43, pp. 139-155.
29. Czarnecki K., Helsen S. Feature-based survey of model transformation approaches. IBM Systems Journal, 2006, vol.45, no.3, pp. 621-645.
30. Ecore structure description (Metamodelling Language), available at: https://download.eclipse.org/modeling/emf/emf/javadoc/2.11/org/eclipse/emf/ecore/package-summary.html (accessed: 01/21/2024).
31. Бычков И.В. Подход к разработке программных компонентов для формирования баз знаний на основе концептуальных моделей / И.В. Бычков, Н.О. Дородных, А.Ю. Юрин // Вычислительные технологии, 2016. – Т. 21. – №4. – С.16-36.
32. Дородных Н.О. State transition diagram editor / Н.О. Дородных, Д.Н. Шпаченко, А.Ю. Юрин // Свидетельство о государственной регистрации программы для ЭВМ. М. Рег. № 2022664023 от 22.07.22.
33. Rule visual modeling language (RVML. Available at: http://knowledge-core.ru/index.php?p=rvml&lan=ru (accessed: 01/21/2024).
34. Knowledge base development system (KBDS). Available at: http://www.kbds.knowledge-core.ru/ (accessed: 01/21/2024).
35. Дородных Н.О. Автоматизированное создание продукционных баз знаний на основе деревьев событий / Н.О. Дородных, О.А. Николайчук, А.Ю. Юрин // Информационные и математические технологии в науке и управлении, 2017. – №2(6). – С.30-41.

Дородных Никита Олегович. *К.т.н., старший научный сотрудник Института динамики систем и теории управления им. В.М. Матросова СО РАН (ИДСТУ СО РАН). Основные направления исследований: автоматизация создания интеллектуальных систем и баз знаний, получение знаний из документов, таблиц, концептуальных моделей, семантическая интерпретация таблиц. ORCID: 0000-0001-7794-4462, Author ID (RSCI): 979843, Author ID (Scopus): 57202323578, Researcher ID: E-8870-2014, tualatin32@mail.ru.*

Юрин Александр Юрьевич. *Д.т.н., заведующий лабораторией Информационных технологий исследования природной и техногенной безопасности ИДСТУ СО РАН, профессор Института информационных технологий и анализа данных ИрНИТУ. Основные направления исследований: разработка интеллектуальных систем и баз знаний, использование прецедентного подхода и семантических технологий при проектировании интеллектуальных диагностических систем. ORCID: 0000-0001-9089-5730, AuthorID (RSCI): 174845, AuthorID (Scopus): 16311168300, Researcher ID: A-4355-2014. iskander@icc.ru.*

UDC 004.89
DOI:10.25729/ESI.2024.34.2.007

Using state transition diagrams for automated creation of knowledge bases

Nikita O. Dorodnykh, Alexander Yu. Yurin

Matrosov institute for system dynamics and control theory SB RAS,
Russia, Irkutsk, *nikidorny@icc.ru*

Abstract. Building knowledge bases in the form of rules, ontologies, or knowledge graphs continues to be a rather time-consuming task in the development of various domain-specific intelligent systems. This article discusses an approach and software for automating the creation of knowledge bases using the analysis and transformation of conceptual models in the form of state transition diagrams. The approach is based on the identification of structural elements of diagrams and their mapping into constructions of the target knowledge representation language. The main stages of the approach are described, as are the analyzed constructions of the considered format of state transition diagrams, as well as the implementation of the approach in the form of web-oriented software, namely, Knowledge Modeling System (KMS). An illustrative example of converting state transition diagrams to form a failure analysis plan is presented.

Keywords: knowledge engineering, knowledge acquisition, knowledge base, ontology, state transition diagram, model transformation, code generation, rules

Acknowledgements: The reported study was supported by the Ministry of Education and Science of the Russian Federation (Project no. 121030500071-2 "Methods and technologies of a cloud-based service-oriented platform for collecting, storing and processing large volumes of multi-format interdisciplinary data and knowledge based upon the use of artificial intelligence, model-driven approach and machine learning").

References

1. Gribova V.V., Okun D.B., Shalfeeva E.A. Semanticheskie modeli dlya ocenki vliyaniya kompleksa faktorov na razvitie zabolevanij [Semantic models for assessing the influence of a combination of factors on the development of diseases]. Ontologiya proyektirovaniya [Ontology of designing], 2021, vol.11, no.4, pp. 464-477.
2. Yasnickij L.N., Cherepanov F.M. Nejroekspertnaya sistema diagnostiki, prognozirovaniya i upravleniya riskami serdechnososudistyh zabolevanij [Neuroexpert system for diagnostics, prediction and risk management of cardiovascular diseases]. Prikladnaya matematika i voprosy upravleniya [Applied mathematics and management issues], 2018, no.3, pp. 107-126.
3. Alam F., Giglou H.B., Malik K.M. Automated clinical knowledge graph generation framework for evidence based medicine. Expert systems with applications, 2023, vol.233, 120964.
4. Voropai N.I., Massel L.V., Kolosok I.N., Massel A.G. IT-infrastruktura dlya postroeniya intellektual'nyh sistem upravleniya razvitiem i funkcionirovaniem sistem energetiki na osnove cifrovyh dvojnikov i cifrovyh obrazov [IT-infrastructure for construction of intelligent management systems of development and functioning of energy systems based on digital twins and digital images]. Izvestiya Rossiyskoy akademii nauk. Energetika [Proceedings of the Russian academy of sciences. Power engineering], 2021, no.1, pp. 3-13.
5. Zhang L., Su H., Zio E., Jiang L., Fan L., Zhang J. A graph structure feature-based framework for the pattern recognition of the operational states of integrated energy systems. Expert systems with applications, 2023, vol.213, 119039
6. Berman A.F., Dorodnykh N.O., Nikolaychuk O.A., Yurin A.Yu. Tekhnologiya sozdaniya osnovannyh na znaniyah sistem dlya issledovaniya dinamiki sostoyaniya tekhnicheskih obektov [Technology for creating knowledge-based systems for studying the dynamics of the state of technical objects]. Vychislitel'nyye tekhnologii [Computational technologies], 2023, vol.28, no.4, pp. 94-108.
7. Berman A.F., Nikolaychuk O.A., Yurin A.Yu. Intellektual'naya informatsionnaya sistema analiza otkazov [Intellectual data system for analyzing failures]. Problemy mashinostroyeniya i nadezhnosti mashin [Journal of machinery manufacture and reliability], 2012, vol. 41 (4), pp. 337-343.
8. Zhang J., Pu X., Zhao R., Li J., Nie Z. Implicit contradictions identification and solution process model for complex technical systems. Computers & Industrial Engineering, 2023, vol.177, 108822.
9. Nozhenkova L.F., Nicheporchuk V.V., Nozhenkova A.I. Sozdanie kompleksnoj sistemy bez-opasnosti regiona na osnove sistemnoj integracii tehnologij [Creating the comprehensive regional safety system based on system integration of technologies]. Informatizatsiya i svyaz' [Informatization and communication], 2013, no.2, pp. 122-124.
10. Nicheporchuk V., Favorskaya M.N., Gryazin I. Framework for intelligent wildlife monitoring. Smart innovation, Systems and technologies, 2020, vol.193, pp. 167-177.
11. Nikolaychuk O.A., Pestova J.V., Yurin A.Yu. Wildfire susceptibility mapping in baikal natural territory using random forest. Forests, 2024, vol.15, no.1, 170.
12. Gavrilova T.A., Kudryavtsev D.V., Muromtsev D.I. Inzhenerija znanij. Modeli i metody [Knowledge Engineering. Models and methods]. SPb., Lan, 2016, 324 p.
13. Giarratano J.C., Riley G.D. Ekspertnyye sistemy: printsipy razrabotki i programmirovaniya, 4-ye izdaniye: per. s angl. [Expert systems: Principles and Programming. 4th ed. trans. from English]. M., Williams, 2007. – 1152 p.
14. Gavrilova T.A., Strakhovich E.V. Vizual'no-analiticheskoe myshlenie i intellekt-karty v ontologicheskom inzhiniringe [Visual analytical thinking and mind maps for ontology engineering]. Ontologiya proyektirovaniya [Ontology of designing], 2020, vol.10, no.1, pp. 87-99.
15. Hopcroft J.E., Motwani R., Ullman J.D. Vvedeniye v teoriyu avtomatov, yazykov i vychisleniy, 2-ye izdaniye: per. s angl. [Introduction to automata theory, languages, and computation. 2rd ed.]. M., Williams, 2008, 528 p.

16. Flexberry Designer. Available at: https://flexberry.net/ru/developers-flexberry-designer.html (accessed: 01/21/2024).
17. Visual paradigm online. Available at: https://online.visual-paradigm.com/ (accessed: 01/21/2024).
18. Enterprise Architect. Available at: https://sparxsystems.com/ (accessed: 01/21/2024).
19. MATLAB & Simulink – State Diagram. Available at: https://www.mathworks.com/discovery/state-diagram.html (accessed: 01/21/2024).
20. EASE: State diagram editor. Available at: https://www.hdlworks.com/products/ease/state_diagram.html (accessed: 01/21/2024).
21. Draw.io. Available at: https://app.diagrams.net/ (accessed: 01/21/2024).
22. Borland CaliberRM. Available at: https://borland-caliber.software.informer.com/11.4/ (accessed: 01/21/2024).
23. Knowledge Modeling System (KMS). Available at: http://kms.knowledge-core.ru/ (accessed: 01/21/2024).
24. CLIPS: A tool for building expert systems. Available at: https://www.clipsrules.net/ (accessed: 01/21/2024).
25. Eremeev A.P. Razvitie apparata tablic reshenij v kontekste sozdaniya gibridnyh intellektual'nyh sistem [The development of decision tables apparatus in the context of creating hybrid intelligent systems]. Gibridnyye i sinergeticheskiye intellektual'nyye sistemy: mater. III Vseros. Pospelovskoy konf. s mezhdunar. uchastiyem [Proc. 3rd All-Russ. Pospelov's conf. on hybrid and synergistic intelligent systems], 2016, pp. 121-132.
26. OWL 2 web ontology language document overview (Second Edition). 2012, available at: https://www.w3.org/TR/owl2-overview (accessed: 01/21/2024).
27. Mens T., Gorp P.V. A taxonomy of model transformations. Electronic notes in theoretical computer science, 2006, vol.152, pp. 125-142.
28. da Silva A.R. Model-driven engineering: A survey supported by the unified conceptual model. Computer languages, systems & structures, 2015, vol.43, pp. 139-155.
29. Czarnecki K., Helsen S. Feature-based survey of model transformation approaches. IBM Systems Journal, 2006, vol.45, no.3, pp. 621-645.
30. Ecore structure description (Metamodelling Language), available at: https://download.eclipse.org /modeling/emf/emf/javadoc/2.11/org/eclipse/emf/ecore/package-summary.html (accessed: 01/21/2024).
31. Bychkov I.V., Dorodnykh N.O., Yurin A.Yu. Podhod k razrabotke programmnyh komponentov dlja formirovanija baz znanij na osnove konceptual'nyh modelej [Approach to the development of software components for generation of knowledge bases based on conceptual models]. Vychislitel'nyye tekhnologii [Computational technologies], 2016, vol.21, no.4, pp.16-36.
32. Dorodnykh N.O., Shpachenko D.N., Yurin A.Yu. State transition diagram editor. Certificate of state registration of a computer program. M. Reg. No. 2022664023, dated 07/22/22.
33. Rule visual modeling language (RVML). Available at: http://knowledge-core.ru/index.php?p=rvml&lan=ru (accessed: 01/21/2024).
34. Knowledge base development system (KBDS). Available at: http://www.kbds.knowledge-core.ru/ (accessed: 01/21/2024).
35. Dorodnykh N.O., Nikolaychuk O.A., Yurin A.Yu. Avtomatizirovannoe sozdanie produkcionnyh baz znanij na osnove derev'ev sobytij [Automated creation of rule-based knowledge bases on the basis of event trees]. Informatsionnyye i matematicheskiye tekhnologii v nauke i upravlenii [Information and mathematical technologies in science and management], 2017, vol.2, no.6, pp. 30-41.

Dorodnykh Nikita Olegovych. *Ph.D., senior associate researcher at Matrosov institute for system dynamics and control theory SB RAS (ISDCT SB RAS). Main research domains: computer-aided development of intelligent systems and knowledge bases, knowledge acquisition based on documents, tables and conceptual models, semantic table interpretation. ORCID: 0000-0001-7794-4462, Author ID (RSCI): 979843, Author ID (Scopus): 57202323578, Researcher ID (WoS): E-8870-2014, tualatin32@mail.ru.*

Yurin Alexander Yurievich. *Ph.D., head of a laboratory "Information and telecommunication technologies for investigation of natural and technogenic safety" at ISDCT SB RAS and associate professor of INRTU. Main research domains: development of intelligent systems and knowledge bases, application of the case-based reasoning and semantic technologies in the design of diagnostic intelligent systems, maintenance of reliability and safety of complex technical systems. ORCID: 0000-0001-9089-5730, Author ID (RSCI): 174845, Author ID (Scopus): 16311168300, Researcher ID (WoS): A-4355-2014, iskander@icc.ru.*

Статья поступила в редакцию 09.02.2024; одобрена после рецензирования 08.04.2024; принята к публикации 06.06.2024.

The article was submitted 02/09/2024; approved after reviewing 04/08/2024; accepted for publication 06/06/2024.

Математическое моделирование и его применение в научных исследованиях

УДК 004.942
DOI:10.25729/ESI.2024.34.2.008

Транспортная задача с ограничением на время перевозок

Бахвалов Сергей Владимирович, Маланова Татьяна Валерьевна, Янхаев Данил Вадимович

Иркутский национальный исследовательский технический университет,
Россия, Иркутск, *malanova_tanya@mail.ru*

Аннотация. В статье представлена модифицированная транспортная задача, в которой учитывается перевозка грузов от поставщиков к потребителям за несколько рейсов, причем весь товар должен быть доставлен в установленный срок. Авторы отдают предпочтение задачам линейного программирования в силу существования различных решателей, позволяющих найти решение существующими методами. В связи с этим была разработана математическая модель транспортной задачи, как задачи целочисленного линейного программирования, и предложено решение задачи в среде программирования Python с использованием библиотеки PuLP. Для наглядности рассмотрен простейший пример.

Ключевые слова: математическое моделирование, оптимизация, линейное программирование, транспортная задача, библиотека PuLP на Python

Цитирование: Бахвалов С.В. Транспортная задача с ограничением на время перевозок / С.В. Бахвалов, Т.В. Маланова, Д.В. Янхаев // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 82-89. – DOI: 10.25729/ESI.2024.34.2.008.

Введение. В современной экономике пункты производства товаров и их потребители зачастую находятся на больших расстояниях, вследствие чего затраты на перевозку продукции являются существенными. Очевидно, что потребители, приобретающие товары с наценкой, включающие данные издержки, отдают предпочтение конкурирующему товару подобного типа с меньшей ценой. Кроме того, время на доставку товаров часто должно быть жестко ограниченным в силу того, что товары могут быть скоропортящимися, товары в пункте потребления могут закончиться. Таким образом, для уменьшения общей себестоимости продукции необходимо формирование такого плана перевозки грузов, чтобы затраты на перевозку были минимальными, и груз был доставлен вовремя.

Как известно, транспортная задача (Т-задача) является одной из самых распространенных задач исследования операций [1-3]. Их решение может позволить повысить эффективность и экономию ресурсов в области логистики и перевозок, способствуя улучшению качества обслуживания и удовлетворению потребностей клиентов. Модель Т-задачи может приобрети дополнительные условия с ограничениями (в дальнейшем модифицированная Т-задача), которые обусловлены спецификой сферы деятельности предприятия и его географического местоположения [4-6]. На сегодняшний день известно множество ее модификаций: транспортная задача по критерию стоимости [1, 2], по критерию времени [2, 7], с промежуточными пунктами [8], многопродуктовая Т-задача [9] и т.д. Также в силу того, что на формирование плана перевозок может влиять множество факторов, в литературе часто рассматриваются многокритериальные транспортные задачи [10-14].

В статье рассмотрена следующая модификация Т-задачи. В пунктах отправления сосредоточен груз в некоторых количествах. Этот груз следует доставить в пункты назначения в ограниченном количестве, определенным спросом, используя различный транспорт.

Рейс из одного пункта отправления в определенный пункт назначения характеризуется затратами времени и стоимости. Товар может перевозиться за несколько рейсов. Каждый рейс имеет заданный объем вместимости, определенный видом транспорта. Только поставщик за-

нимается перевозкой грузов. Наиболее часто при перевозке больших объемов груза автотранспортом оплата производится в ценах за рейс (а не за объем или массу перевозимого груза). Необходимо составить план перевозки груза так, чтобы затраты на перевозку были минимальными, а также, чтобы перевозка была осуществлена в установленные сроки.

Так как оплата за перевозку груза осуществляется в ценах за рейс, то необходимо привести математическую модель, зависящую от количества рейсов. Анализ литературы показал, что, несмотря большое разнообразие Т-задач, задачи в такой постановке не рассматривались.

1. Постановка задачи. Необходимо построить математическую модель задачи поиска оптимального плана перевозки товаров от поставщиков до потребителей, в которой функция затрат определяется по количеству рейсов, и учтено ограничение времени на перевозку.

Как известно, задачи линейного программирования (ЛП) являются наиболее изученными в области задач оптимизации, и для их решения существуют различные готовые библиотеки в различных средах программирования, таким образом, разрабатываемая математическая модель должна также оставаться в классе задач ЛП, чтобы была возможность использовать уже существующие возможности с необходимыми доработками.

2. Описание и подготовка исходных данных. Введем следующие обозначения:

A_i – поставщики продукции с их объемами a_i $(i = \overline{1, m})$;

B_j – потребители с величинами спроса b_j $(j = \overline{1, n})$;

$C = \|c_{ij}\|_{m \times n}$ – стоимостные затраты на перевозку из i-го пункта производства в i-й пункт;

$T = \|t_{ij}\|_{m \times n}$ – временные затраты на перевозку из i-го пункта производства в i-й пункт;

s – время, за которое необходимо произвести перевозку всех грузов.

Для того, чтобы определить значения параметров c_{ij} и t_{ij}, необходимо выполнить следующие действия.

Так как доставить груз от i-го пункта производства к j-му пункту возможно различным способом (самолетом, поездом, автотранспортом, водным транспортом), то сначала необходимо получить массив объектов $P = \|p_{ij}\|_{m \times n}$. Здесь под p_{ij} понимаем объект, обозначающий совокупность возможных путей, по которым можно добраться от i-го пункта производства к j-му пункту потребления. Объект имеет атрибут – вектор возможных путей, где каждый путь – также объект с атрибутами:

- цена перевозки c_{ij}^p;
- время доставки t_{ij}^p;
- статус – выбран / не выбран данный путь для планирования.

Предполагается, что все возможные пути совокупности p_{ij} можно найти либо вручную, либо с использованием геоинформационных систем, и пользователь имеет возможность вносить изменения в массив P.

Среди всех возможных путей в совокупности p_{ij} выбираем тот, который соответствует минимальным затратам и открыт на период планирования, или тот путь, который выбрал сам пользователь. В результате в матрицу C вносим соответствующее значение цены перевозки, а в матрицу T – значения времени перевозки.

Таким образом, коммуникации, связывающие склады и пункты потребления выбираются, исходя из условия: $c_{ij} = \min_k c_{ij}^p$ или вручную лицом, принимающим решение (ЛПР).

3. Формирование модели. Пусть количество перевезенной продукции i -го пункта в j-ый определяется матрицей $X = \|x_{ij}\|_{m \times n}$.

Необходимо определить такой план перевозки X, чтобы стоимость его перевозки была наименьшей, все заявки были удовлетворены, все запасы были вывезены, а время перевозки было не больше заданного значения s.

Временные затраты определяются следующим образом. Учитывая, что перевозка будет производиться рейсами, обозначим количество рейсов, которое нужно выполнить при перевозке за r_{ij}. Чтобы оставаться в рамках задачи линейного программирования, сделаем следующие допущения. Если на складах имеются несколько машин, то пусть все машины выезжают в пункты потребления одновременно, считаем это за один совокупный рейс из i -го пункта в j-ый. Кроме того, поставщик перевозит груз в каждый пункт потребления последовательно (не одновременно). Тогда:

$$r_{ij} = \left\lceil \frac{x_{ij}}{g_{ij}} \right\rceil, i = \overline{1,m}, j = \overline{1,n}. \quad (1)$$

Здесь $G = \|g_{ij}\|_{m\times n}$ – общая грузоподъемность совокупного транспорта на i-м складе, осуществляющего перевозку из i-го пункта производства в j-й пункт; r_{ij} –целое число; знак $\lceil \dots \rceil$ обозначает округление до целого числа в большую сторону.

Параметры g_{ij} и s задаются пользователем, либо хранятся в соответствующей базе данных.

Общие затраты на перевозку всех грузов составляют $\sum_{i=1}^{m}\sum_{j=1}^{n}(c_{ij}r_{ij})$, и эти затраты необходимо минимизировать.

Для ЛПР план должен выражаться не в рейсах, а в объемах перевозимого груза, поэтому, для возможности обратного перехода от количества рейсов r_{ij} к объему перевозок x_{ij}, необходимо учитывать погрешность округления z_{ij} по формуле (1), тогда справедливо соотношение:

$$r_{ij} = \frac{x_{ij}+z_{ij}}{g_{ij}}. \quad (2)$$

Отсюда,

$$z_{ij} = g_{ij}r_{ij} - x_{ij}, \quad (3)$$

и объем перевозимого груза $x_{ij} = g_{ij}r_{ij} - z_{ij}$ должен быть неотрицательным.

Суммарное время перевозки от каждого i-го пункта до всех пунктов потребления составит $\sum_{j=1}^{n} r_{ij}t_{ij}$.

Следует заметить, что допущение о том, что из i-го пункта все рейсы могут быть только в один пункт, и пока все не будет вывезено, в другой пункт перевозка не будет осуществляться, вносит дополнительную погрешность в модель, но на составление плана перевозки это допущение не должно существенно повлиять. В данном случае учитывается лишь максимально затраченное время из возможного.

В результате математическая модель задачи будет иметь вид:

$$F(r) = \sum_{i=1}^{m}\sum_{j=1}^{n}(c_{ij}r_{ij}) \to \min \quad (4)$$

при ограничениях

$$\begin{cases} \sum_{j=1}^{n} r_{ij}t_{ij} \le s, i = \overline{1,m} \\ \sum_{i=1}^{m}(g_{ij}r_{ij} - z_{ij}) = b_j,\ j = \overline{1,n} \\ \sum_{j=1}^{n}(g_{ij}r_{ij} - z_{ij}) = a_i,\ i = \overline{1,m} \\ g_{ij}r_{ij} - z_{ij} > 0,\ i = \overline{1,m}, j = \overline{1,n} \\ r_{ij} \ge 0,\ i = \overline{1,m}, j = \overline{1,n} \\ z_{ij} > 0,\ i = \overline{1,m}, j = \overline{1,n} \end{cases} \quad (5)$$

4. Решение поставленной задачи среде программирования Python с использованием библиотеки PuLP. Задача (4), (5) является модифицированной транспортной с добавлением

ограничения по времени. Ввод указанных ограничений по времени не дает нам возможности решать полученную задачу методами, приспособленными для решения Т-задач, например, методом потенциалов, так как число ограничений больше, чем $m + n - 1$. Таким образом, количество базисных клеток в соответствующей таблице для Т-задачи не обязательно должно быть равным $m + n - 1$, что не дает возможности нужным образом работать с базисными клетками: вычислять потенциалы, формировать оценочную матрицу, строить циклы.

В связи с этим в данной работе предлагается использовать общие методы задачи целочисленного линейного программирования. Для этого воспользуемся библиотекой PuLP на языке программирования Python в среде разработки Jupyter Notebook. PuLP – программный интерфейс для решения задач линейного программирования на Python, предназначенный для определения задачи и вызова решателей. В качестве решателя по умолчанию используется COIN-OR Branch and Cut Solver (CBC) [15].

Для решения задачи (4), (5) была разработана функция, которая принимает на вход в качестве аргументов матрицы A, B, G, T и интерпретирует входные данные Т-задачи к валидной постановке модели библиотеки PuLP, и дальнейшего её решения с использованием решателя CBC. Общая схема работы функции представлена на рисунке 1.

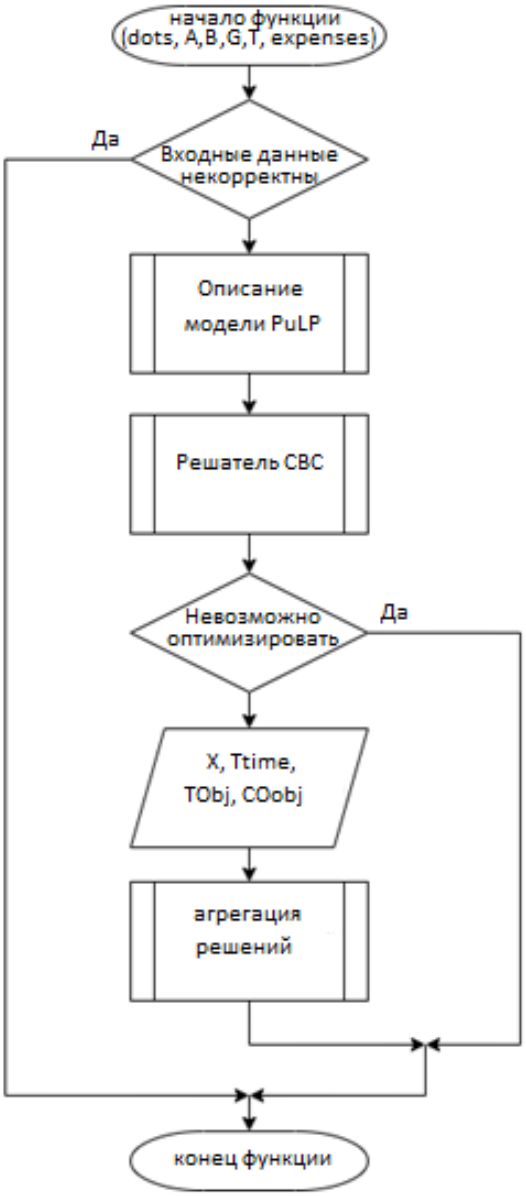


Рис. 1. Схема работы функции

Процесс описания модели в PuLP является важным этапом, ведь от него зависит, что получим в качестве решения. Код на языке Python процесса описания модели PuLP должен включать:

- инициализацию модели; на первом этапе вызывается метод создания модели, аргументами которого являются наименование модели и направление оптимизации (максимизация или минимизация);
- следующим этапом следует инициализация переменных решения; создаются словари R и Z, и в цикле по каждому поставщику A и потребителю B заполняются переменными;
- создание целевой функции; на данном этапе с помощью аргументов функции и созданных переменных решения описывается целевая функция, далее данное описание подается как аргумент в метод созданной модели PuLP;
- создание ограничений; заключительным этапом происходит описание ограничений в виде строкового типа данных, для каждого вида ограничения в цикле, в зависимости от постановки задачи, задаем наименование ограничения и его математическую структуру.

Далее необходимо решить сформированную задачу ЛП выбранным решателем.

В общем случае рассматриваемая задача имеет решение, если система ограничений (5) совместна. По теореме о необходимом и достаточном условии совместности системы линейных неравенств рассматриваемая система ограничений (5) (без учета ограничений на целочисленность) совместна, если ее ранг равен числу ограничений [16]. Кроме того, следует отметить, что количество пунктов поставщиков и потребителей существенно влияет на время решения задачи, так как решатели основаны на таких методах, как метод отсекающих плоскостей и метод ветвей и границ, которые используют многократное применение стандартных методов задачи ЛП.

5. Пример. Пусть необходимо вывезти груз из пунктов A_1, A_2, A_3, в которых хранятся запасы в следующих объемах соответственно: 3000, 4900, 3800. Груз должен быть доставлен в пункты B_1, B_2, B_3, B_4 в следующих объемах соответственно: 3200, 2900, 3000, 2600.

Затраты на перевозку груза одним рейсом из каждого склада каждому потребителю заданы матрицей C:

$$C = \begin{pmatrix} 28 & 28 & 26 & 80 \\ 120 & 21 & 31 & 39 \\ 80 & 27 & 120 & 34 \end{pmatrix}.$$

Время, затраченное на один рейс, определено матрицей T:

$$T = \begin{pmatrix} 6 & 4 & 5 & 24 \\ 24 & 6 & 4 & 5 \\ 24 & 7 & 24 & 6 \end{pmatrix}.$$

Совокупная грузоподъемность транспорта, которая определена для перевозки из i-го пункта в j-й задана в матрице G:

$$G = \begin{pmatrix} 150 & 210 & 120 & 800 \\ 200 & 120 & 180 & 180 \\ 800 & 180 & 200 & 120 \end{pmatrix}.$$

Следует отметить, что, если поставщик будет использовать один и тот же транспорт для перевозки всех грузов всем своим потребителям, то в соответствующей для него строке матрицы G будут одинаковые значения g_{ij}.

Время, за которое нужно вывезти груз, определено переменной $s = 126$.

В результате решения задачи (4) – (5) получаем следующий план перевозок:

$$X = \begin{pmatrix} 0 & 2080 & 120 & 800 \\ 0 & 220 & 2880 & 1800 \\ 3200 & 600 & 0 & 0 \end{pmatrix}.$$

По формуле (1) число рейсов представлено следующей матрицей:

$$\mathrm{r} = \begin{pmatrix} 0 & 10 & 1 & 1 \\ 0 & 2 & 16 & 10 \\ 4 & 4 & 0 & 0 \end{pmatrix}$$

Тогда суммарное время из каждого пункта определяется следующим образом:

$$s_1 = 0 \cdot 6 + 10 \cdot 4 + 1 \cdot 5 + 1 \cdot 24 = 60$$
$$s_2 = 0 \cdot 24 + 2 \cdot 6 + 16 \cdot 4 + 10 \cdot 5 = 126$$
$$s_3 = 3200 \cdot 24 + 7 \cdot 4 + 0 \cdot 24 + 0 \cdot 6 = 124$$

В результате видим, что время перевозки не превышает числа $s = 126$, за которое нужно перевезти вест груз.

Цена перевозки составляет 1742 единицы.

Если пренебречь ограничением на время перевозки, то есть решить обычную транспортную задачу, то оптимальный план позволит перевезти груз с затратами в 1640 единиц. Но время перевозки продукции от каждого склада составит соответственно $\mathrm{s}_1 = 84$, $\mathrm{s}_2 = 163$, $\mathrm{s}_3 = 123$. Как видим, этот план позволяет вывезти товар со всех складов всем потребителям за минимальную стоимость, но время перевозки составляет более 126.

Отсюда следует, что внесение дополнительного ограничения в транспортную задачу сужает область допустимых значений снизу, что приводит к увеличению минимального значения на этой области.

Заключение. Таким образом, предложенная модель транспортной задачи позволяет находить оптимальный план перевозки грузов, когда груз доставляется за несколько рейсов, а время на перевозку ограничено. Данная модель может быть применена для задач, имеющих практическое значение. Полученная модификация транспортной задачи относится к классу задач дискретного линейного программирования, поэтому результат можно получить известными методами решения задач целочисленного программирования с использованием известных решателей.

Список источников

1. Зайченко Ю.П. Исследование операций: учеб. для вузов / Ю. П. Зайченко. – 3-е изд., перераб. и доп. – Киев: Выща шк., 1988. –552 с.
2. Таха Х.А. Введение в исследование операций: пер. с англ. / Х.А. Таха. – М.: Вильямс, 2016. – 912 с.
3. Тюрин А.Ю. Модели транспортного обслуживания в цепях поставок пищевой промышленности / А.Ю. Тюрин // Вестн. КузГТУ, 2011. – №4. – С.89-92.
4. Петрунин С.В. Метод решения открытых транспортных задач / С.В. Петрунин, Г.В. Кренева // Научный вестник Московского государственного технического университета гражданской авиации, 2014. – № 202. С. 58-60.
5. Петрунин С.В. О решении транспортных задач большой размерности / С.В. Петрунин // Научный Вестник МГТУ ГА, 2008. – № 131. – С. 183-185.
6. Петрунин С.В. Решение транспортных задач ПС-методом при ограничениях на переменные / С.В. Петрунин // Научный вестник МГТУ ГА, 2014. – №202. – С. 53-57
7. Нечитайло Н.М. Транспортная задача по критерию минимума суммарного времени и модификация метода Балинского для её решения / Н.М. Нечитайло, С.В. Мартемьянов, В.Л. Панасов // Инженерный вестник Дона, 2016. – № 4 (43). – 64 с.
8. Кривопалов В.Ю. Решение открытой транспортной задачи с промежуточными пунктами / В.Ю. Кривопалов // Сборник научных трудов Международной научно-технической конференции, 2015. – Т. 2. – С. 86-91.
9. Монтлевич В.М. О некоторых постановках многопродуктовых транспортных задач / В.М. Монтлевич, И.А. Бородинова // Вестник СамГУ, 2008. – №7 (66). – С. 86-93.
10. Yu L., Zeleny M. The set of all nondominated solutions in linear cases and multycriteria simplex method. J. of Math. Anal. and applic,1975, vol. 45, no.2, pp. 430-468.
11. Золотарюк А.В. Математическая модель многокритериальной оптимизации транспортных перевозок / А.В. Золотарюк // Инновационные технологии в науке и образовании, 2015. – № 1. – С. 317-320.
12. Кошкин Б.П. О многокритериальной транспортной задаче / Б.П. Кошкин, С.И. Носков, В.А. Оленцевич, А.И. Рязанцев // Фундаментальные исследования, 2017. – №7. – С. 35-38.

13. Носков С.И. Двухкритериальная транспортная задача / С.И. Носков, А.И. Рязанцев // T-Comm: Телекоммуникации и транспорт, 2019. – Т. 13. – № 2. – С. 59-63.
14. Осыкина Ю.А. Многокритериальная транспортная задача с разрывной целевой функцией / Ю.А. Осыкина, Г.Д. Чернышова // Вестник Воронежского государственного университета. Серия: системный анализ и информационные технологии, 2008. – № 2. – С. 10-12.
15. Vlasov A., Stanovskih A. Additional constraints in the vehicle routing problem. Universum, 2022, 4(97), available at: https://7universum.com/ru/tech/archive/item/13513.
16. Черников С.Н., Системы линейных неравенств / С.Н. Черников // Успехи математических наук, 1953. – Т. 8. – № 2(54). – С. 7-73.

Бахвалов Сергей Владимирович. *Доцент, к.т.н., Иркутский национальный исследовательский технический университет, SPIN:1842-7155, AuthorID: 169991, ORCID:0009-0009-83961926, bsv@istu.edu, 664074, г. Иркутск, ул. Лермонтова 83.*

Маланова Татьяна Валерьевна. *Доцент, к.т.н., Иркутский национальный исследовательский технический университет, AuthorID: 530825, ORCID: 0000-0002-1586-2530, malanova_tanya@mail.ru, 664074, г. Иркутск, ул. Лермонтова 83.*

Янхаев Данил Вадимович. *Магистрант, Иркутский национальный исследовательский технический университет, yanhaev2000@gmail.com, 664074, г. Иркутск, ул. Лермонтова 83.*

UDC 004.942
DOI: 10.25729/ESI.2024.34.2.008

Transport problem with transportation time constraint

Sergey V. Bakhvalov, Tatyana V. Malanova, Danil V. Yanhaev

National research Irkutsk state technical university,
Russia, Irkutsk, *malanova_tanya@mail.ru*

Abstract. The article presents a modified transport problem, which takes into account the transportation of goods from suppliers to consumers over several flights, and all goods must be delivered on time. The authors give preference to linear programming problems due to the existence of various solvers that allow one to find a solution using existing methods. In this regard, a mathematical model of the transport problem was developed as an integer linear programming problem, and a solution to the problem was proposed in the Python programming environment using the PuLP library. For clarity, a simple example is considered.

Keywords: mathematical modeling, optimization, linear programming, transport problem, PuLP library in Python

References

1. Zaychenko Yu.P. Issledovanie operatsiy: ucheb. dlya vuzov [Operations research]. 2nd ed., Kiev, Vyshcha shk., 1988, 552 p.
2. Takha H. A. Vvedeniye v issledovaniye operatsiy, per. s angl. [Operations research: an introduction, trans. from English]. M., Williams, 2007, 838 p.
3. Tyurin A.Yu. Vvedeniye v issledovaniye operatsiy [Utilization of advanced methods of food industry goods transportation routing]. Vestn. KuzGTU [Bulletin of Kuzbass state technical university], 2011, no.4, pp.89-92.
4. Petrunin S.V., Kreneva G.V. Metod resheniya otkrytykh transportnykh zadach [Solution of open transportation problems] Nauchnyy vestnik Moskovskogo gosudarstvennogo tekhnicheskogo universiteta grazhdanskoy aviatsii [Scientific bulletin of the Moscow state technical university of civil aviation], 2014, no. 202, pp. 58-60z
5. Petrunin S.V. O reshenii transportnykh zadach bol'shoy razmernosti [Solution of transportation problems of large volume]. Nauchnyy vestnik MGTU GA [Scientific bulletin of the MSTU CA], 2008, no. 131, pp. 183-185.
6. Petrunin S.V. Resheniye transportnykh zadach PS-metodom pri ogranicheniyakh na peremennyye [Solution of transportation problems with bounded-variables using PC-method]. Nauchnyy vestnik MGTU GA [Scientific bulletin of the MSTU CA], 2014, no.202, pp. 53-57.

7. Nechitaylo N.M., Martem'yanov S.V., Panasov V.L. Transportnaya zadacha po kriteriyu minimuma summarnogo vremeni i modifikatsiya metoda balinskogo dlya ee resheniya [Transport problem based on the criterion of minimum total time and modification of the Balinsky method for solving it]. Inzhenernyy vestnik Dona [Engineering bulletin of the Don], 2016, no. 4 (43), pp. 64-73.
8. Krivopalov V.Yu. Reshenie otkrytoy transportnoy zadachi s promezhutochnymi punktami [Solution of an open transport problem with intermediate points]. Sbornik nauchnykh trudov Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii [Collection of scientific papers of the international scientific and technical conference], 2015, vol. 2, pp. 86-91.
9. Montlevich V.M., Borodinova I.A. O nekotorykh postanovkakh mnogoproduktovykh transportnykh zadach [About some tasks of multicommodity transporting problems]. Vestnik SamGU [Vestnik of Samara State University], 2008, no.7 (66), pp. 86-93.
10. Yu L., Zeleny M. The set of all nondominated solutions in linear cases and multycriteria simplex method. J. of Math. Anal. and applic,1975, vol. 45, no.2, pp. 430-468.
11. Zolotaryuk A.V. Matematicheskaya model' mnogokriterial'noy optimizatsii transportnykh perevozok [Mathematical model of multi-criteria optimization of transportation organization]. Innovatsionnye tekhnologii v nauke i obrazovanii [Innovative technologies in science and education], 2015, no. 1, pp. 317-320.
12. Koshkin B.P., Noskov S.I., Olentsevich V.A., Ryazantsev A.I. O mnogokriterial'noy transportnoy zadache [About multi-criteria transport problem]. Fundamental'nye issledovaniya [Basic Research], 2017, no.7, pp. 35-38.
13. Noskov S.I., Ryazantsev A.I. Dvukhkriterial'naya transportnaya zadacha [Two-criteria transport problem]. T-Comm, Telekommunikatsii i transport [Telecommunications and transport], 2019, vol. 13, no. 2, pp. 59-63.
14. Osykina Yu.A., Chernyshova G.D. Mnogokriterial'naya transportnaya zadacha s razryvnoy tselevoy funktsiey [Multicriteria transport problem with discontinuous objective function]. Vestnik Voronezhskogo gosudarstvennogo universiteta, Seriya sistemnyy analiz i informatsionnye tekhnologii [Bulletin of Voronezh State University. Series systems analysis and information technology], 2008, no. 2, pp. 10-12.
15. Vlasov A., Stanovskih A. Additional constraints in the vehicle routing problem. Universum, 2022, 4(97), available at: https://7universum.com/ru/tech/archive/item/13513.
16. Chernikov S.N. Sistemy linejnyh neravenstv [Systems of linear inequalities]. Uspehi matematicheskih nauk [Advances in mathematical sciences], 1953, vol. 8. no. 2(54), pp. 7-73.

***Sergey Vladimirovich Bakhvalov**. Cand. Sci. (Technics), associate professor at software engineering center, institute of information technology and data analysis, Irkutsk national research technical university, SPIN:1842-7155; AuthorID: 169991, ORCID:0009-0009-83961926, bsv@istu.edu, 664074, Irkutsk,83 Lermontov Str.*

***Tatyana Valeryevna Malanova**. Cand. Sci. (Technics), associate professor at software engineering center, institute of information technology and data analysis, Irkutsk national research technical university, AuthorID: 530825, ORCID: 0000-0002-1586-2530, malanova_tanya@mail.ru, 664074, Irkutsk,83 Lermontov Str.*

***Danil Vavimovich Yanhaev.** Master's student, Irkutsk national research technical university, yanhaev2000@gmail.com, 664074, Irkutsk,83 Lermontov Str.*

Статья поступила в редакцию 15.12.2023; одобрена после рецензирования 06.06.2024; принята к публикации 06.06.2024.

The article was submitted 12/15/2023; approved after reviewing 06/06/2024; accepted for publication 06/06/2024.

УДК 519.6
DOI:10.25729/ESI.2024.34.2.009

Алгоритм сглаживания пути на триангулированной поверхности

Серебровская Екатерина Александровна, Марчевский Илья Константинович, Ерофеева Мария Александровна

Московский государственный технический университет имени Н.Э. Баумана,
Россия, Москва, *e.a.serebrovskaya@gmail.com*

Аннотация. Предложен и реализован в Wolfram Mathematica алгоритм оптимизации длины пути на триангулированной поверхности. Первые два шага «легковесны», но предполагают вариацию траектории, проходящей по ребрам, лишь в пределах примыкающих к ним треугольников. Последующие шаги позволяют за несколько итераций прийти к кратчайшему в математическом смысле пути. Сходимость алгоритма не доказана строго, но обеспечивается в большом количестве рассмотренных примеров.

Ключевые слова: кратчайший путь, триангулированная поверхность, условная оптимизация, Wolfram Mathematica

Цитирование: Серебровская Е.А. Алгоритм сглаживания пути на триангулированной поверхности / Е.А. Серебровская, И.К. Марчевский, М.А. Ерофеева // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 90-98. – DOI:10.25729/ESI.2024.34.2.009.

Введение. В процессе разработки новой модификации бессеточного лагранжева метода вихревых петель [1] для моделирования пространственного обтекания тел потоком несжимаемой среды и вычисления действующих на них аэрогидродинамических нагрузок возникла проблема коррекции пространственной кривой сложной формы, которая определяет форму каждой петли. Подобные петли задают форму вихревых структур, тем самым определяя распределение завихренности в области течения (в вихревых методах именно завихренность является первичной расчетной величиной). Такой подход позволяет, в частности, обеспечить точное выполнение условия бездивергентности.

Вихревые петли генерируются на обтекаемой поверхности, а затем движутся в области течения; для моделирования этих процессов разработано семейство вычислительных алгоритмов, в том числе полуэмпирических. В силу неизбежных погрешностей, связанных, в том числе, с заданием триангулированной формы обтекаемой поверхности вместо гладкой, конечного шага расчета по времени и многими другими допущениями, некоторые участки петель в ходе расчета их движения в потоке «проникают» в обтекаемое тело и поэтому требуется «вынести» их на поверхность тела, обеспечив достаточную гладкость вновь формируемой траектории. Вместо последней задачи можно решить задачу поиска кратчайшего пути на триангулированной поверхности.

Задача поиска кратчайшего пути, лежащего на некоторой поверхности, весьма проста по постановке и может быть отнесена к классическим задачам дифференциальной геометрии кривых и поверхностей – речь идет о построении геодезической. В простом и популярном изложении этот вопрос рассмотрен в [2]. Однако в различных вычислительных алгоритмах поверхность лишь в редких случаях бывает заданной аналитически и обладает необходимой степенью гладкости; как правило, на практике приходится иметь дело с ее упрощенным представлением в виде поверхности некоторого многогранника с плоскими гранями. Далее будем рассматривать наиболее общий случай триангулированной поверхности, поскольку грани с более чем тремя вершинами можно разбить на треугольники. Для такой «кусочно-плоской» (или, по-другому, «граненой») поверхности неприменим аппарат дифференциального исчисления, и для решения поставленной задачи следует применять иные подходы.

1. Кратчайший путь на графе и на триангулированной поверхности. Близкая, на первый взгляд, задача о поиске кратчайшего пути на взвешенном графе давно и успешно решена: наиболее известны алгоритмы Дейкстры [3] (1959 г.), Флойда – Уоршалла (1962 г.), Белмана

– Форда (1969 г.). Не вдаваясь в детали, отметим, что они могут быть применены к графу, ребрами которого выступают ребра рассматриваемой триангулированной поверхности – ее «каркас»: в результате будет найден кратчайший из путей, проходящих по ребрам. Отметим, что если начальная и конечная точки не совпадают с вершинами, то треугольные грани, которым они принадлежат, следует очевидным образом разбить на три треугольника и свести задачу к более простой. Найденный в результате путь, помимо того, что не является кратчайшим на поверхности, обладает особенностью, которая может быть важной для некоторых приложений: он не является гладким (рис. 1).

Естественно, речь идет не о гладкости в математическом смысле: никакой из путей на триангулированной поверхности не может быть гладким, если только он не проходит целиком по граням, лежащим в одной плоскости, или не включает в себя специальным образом построенные сопряжения; в последнем случае он едва ли может быть кратчайшим. Нестрогое, но понятное и наглядное определение «гладкого» пути на триангулированной поверхности можно дать следующим образом: если сделать развертку для тех граней поверхности, по которым проходит путь, то на развертке углы между звеньями полученной кривой должны быть развернутыми, за исключением случаев прохождения пути через вершину. Близкий к оптимальному путь, углы на развертке которого близки к развернутым, будем называть «достаточно гладким».

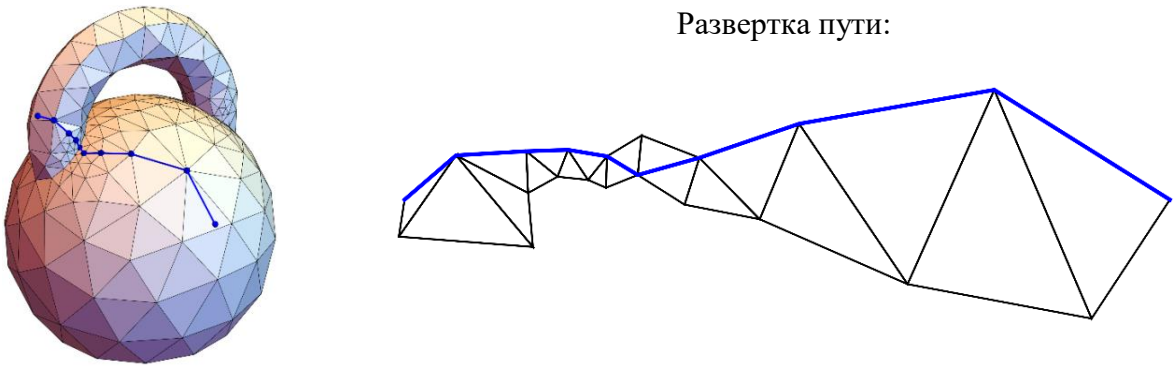


Рис. 1. Путь, найденный при помощи алгоритма Дейкстры, и его развертка

Также следует заметить, что функция длины пути может иметь как глобальный, так и локальные минимумы, соответственно, необходимо различать глобально- и локально-кратчайшие пути. Очевидно, что если путь на развертке образует прямую линию, то он будет, во-первых, локально-кратчайшим, а во-вторых – наиболее гладким из возможных в указанном выше смысле. Далее необходимо сделать три замечания.

1. Локально-кратчайших путей между двумя фиксированными точками может быть несколько, они могут при этом значительно различаться по длине, однако каждый из них на развертке будет представлять собой прямую линию (рис. 2); будем рассматривать процедуру поиска локально-кратчайшего пути, используя в качестве начального приближения результат работы алгоритма Дейкстры.
2. Локально-кратчайший путь на развертке может не быть прямой линией, а оставаться ломаной: последнее характерно для тех случаев, когда исходная поверхность имеет седловые точки, в этом случае искомый путь может проходить через вершину седла (и, соответственно, через вершины примыкающих к седловой точке треугольников на развертке), рис. 3.
3. Задачу поиска локально-кратчайшего (и тем более глобально-кратчайшего) пути на практике во многих случаях, как представляется, можно заменить более «слабой» задачей поиска достаточно гладкого в вышеуказанном смысле пути.

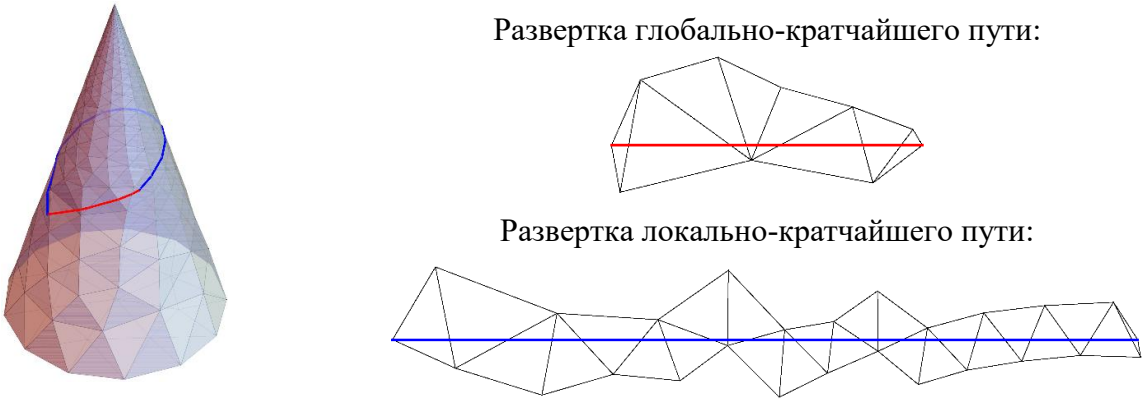


Рис. 2. Два локально-кратчайших пути на конусе и их развертки

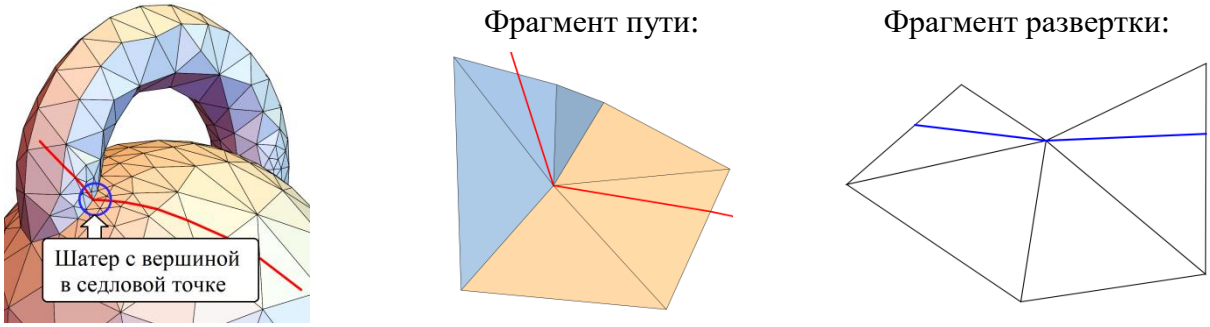


Рис. 3. Кратчайший путь в окрестности седловой точки

Несмотря на кажущуюся простоту задачи поиска кратчайшего пути на триангулированной поверхности, универсальные методы ее эффективного решения до сих пор не разработаны. Публикации на эту тему стали активно появляться еще в 1970-е годы; над смежными проблемами работали такие известные специалисты по алгоритмам вычислительной геометрии, как Д. Киркпатрик, Ф. Препарата, Д. Ли и др. Задача значительно упрощается, если рассматриваемая поверхность выпуклая; краткий обзор соответствующих алгоритмов приведен в [4], однако на практике требуются более универсальные алгоритмы, применимые к невыпуклым телам сложной формы (например, см. рис. 3). Подобные алгоритмы стали появляться в 1980-1990-е годы, выделим среди них работу [5]; обзор таких алгоритмов и их сравнительный анализ по состоянию на 2009 г. приведены в [6]. Вплоть до сегодняшнего дня регулярно появляются новые алгоритмы, среди которых отметим [7, 8]; в абсолютном большинстве из них в той или иной мере использована технология построения разверток, многие алгоритмы предполагают построение некоторого дерева и поиск по нему. Тем не менее, до сих пор остаются актуальными методы «грубой силы», предполагающие сильное дробление поверхности на ячейки очень простой формы, на которых решить задачу, по крайней мере, локально, весьма легко; к таким алгоритмам относится [8].

Отметим, что в хорошо известной библиотеке вычислительной геометрии CGAL [9] имеется модуль Triangulated Surface Mesh Shortest Paths, позволяющий решить задачу поиска глобально-кратчайшего пути на триангулированной поверхности. Реализованный в нем алгоритм основан на работе [10], в которой, в свою очередь, разработана усовершенствованная версия достаточно «старого» алгоритма [5].

Сравнительная простота использования библиотеки (вопросы ее предварительной компиляции и настройки опускаем) может в значительной мере нивелироваться не слишком высокой производительностью: с ее помощью наиболее эффективно могут быть решены задачи поиска путей от выбранной точки на поверхности до множества других точек, причем даже

значительное увеличение числа точек назначения приводит к весьма малому росту вычислительной сложности. Последнее связано со структурой алгоритма: сначала производится занимающая значительное время «предобработка» всей поверхности для заданной точки-источника, а затем выполняется прокладка пути в точку назначения, что является намного более легковесной операцией. Если же требуется решать задачу для множества различных пар точек «начало – конец маршрута», и при этом, как было отмечено выше, оптимальность пути не является критичной, а достаточно лишь обеспечить сравнительную его гладкость, указанный алгоритм весьма неэффективен.

2. Алгоритм решения задачи. Предлагаемый алгоритм состоит из нескольких шагов, на каждом из которых происходит сглаживание пути и уменьшение его длины. Приведем их краткое описание, опуская некоторые «технические» детали реализации, которые представляются довольно очевидными, но их описание заняло бы слишком много места и едва ли способствовало пониманию сути алгоритма.

1. Используя алгоритм Дейкстры, построить кратчайший путь на графе – каркасной модели поверхности; он будет иметь вид $\{A_1, A_2, \dots, A_N\}$, где A_i – номера (индексы) вершин многогранника.

2. Разбить найденный путь на тройки $\{A_1, A_2, A_3\}$, $\{A_3, A_4, A_5\}$ и т.д.; для каждой такой тройки рассмотреть треугольники, имеющие общей вершиной точку с четным индексом A_{2k} (их совокупность будем называть словом «шатер»). Для каждого шатра рассмотреть две его части, ограниченные ребрами $A_{2k-1}A_{2k}$ и $A_{2k}A_{2k+1}$; построить их развертки. Если угол при вершине A'_{2k} (штрих обозначает образ вершины на развертке) меньше развернутого, то кратчайший путь на этом «полушатре» – прямая, соединяющая точки A'_{2k-1} и A'_{2k+1}, которая не будет проходить через A'_{2k}. Если углы на развертках обоих полушатров меньше развернутого, то из двух путей следует выбрать тот, что короче. Если, наоборот, на обеих развертках углы при A'_{2k} больше развернутого, то имеем дело с седловой точкой (рис. 3); в этом случае любой из путей по шатру между начальной и конечной точками будет длиннее исходного пути, проходящего через вершину.

3. Заменить вершины в пути, имеющие четные индексы, наборами точек, соответствующих точкам пересечения полученного пути с ребрами шатров (рис. 4); путь в этом случае будет иметь вид $\{A_1, \{C_1^2,\ C_2^2, \dots, C_{m_2}^2\}, A_3, \{C_1^4,\ C_2^4, \dots, C_{m_4}^4\}, \dots, A_N\}$, при этом положение каждой «промежуточной» точки C_m^{2k} удобно задавать, указывая вершины ребра, на котором она лежит, и весовой коэффициент: $C_m^{2k} \leftrightarrow \{A_{2k},\ A_m^{2k}, \alpha_m^{2k}\}$, где A_m^{2k} – та из вершин поверхности, что лежит на соответствующем месте в основании шатра с вершиной A_{2k}; $\alpha_m^{2k} \in [0, 1]$ – весовой коэффициент, нулевое и единичное значения веса говорят о совпадении точки пути с вершиной A_{2k} и A_m^{2k} соответственно.

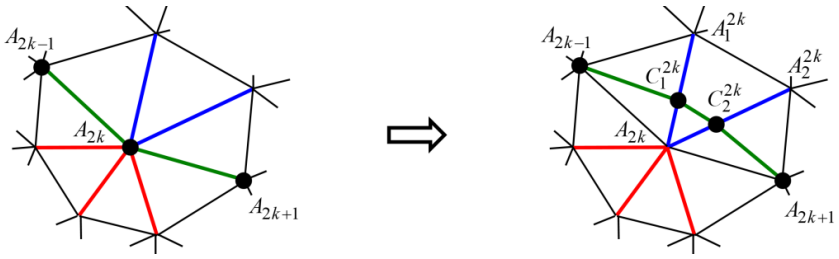


Рис. 4. Оптимизация на шатре с четным индексом вершины

4. Проделать аналогичную операцию, но рассматривая шатры с вершинами в точках с нечетными индексами A_{2k+1}; в этом случае «разрез» на полушатры будет производиться не по ребрам шатра, а по линиям $C_{m_{2k}}^{2k}A_{2k+1}$ и $A_{2k+1}C_1^{2k+2}$, проходящим в общем случае по граням шатра (рис. 5). В итоге полученный путь будет образован точками $\{A_1, \{C_1^2,\ C_2^2, \dots, C_{m_2}^2\}, \{C_1^3,$

$C_2^3, \dots, C_{m_3}^3\}, \{C_1^4,\ C_2^4, \dots, C_{m_4}^4\}, \dots, A_N\}$, а каждая из точек C_m^k – задана тройкой указанного в предыдущем пункте вида.

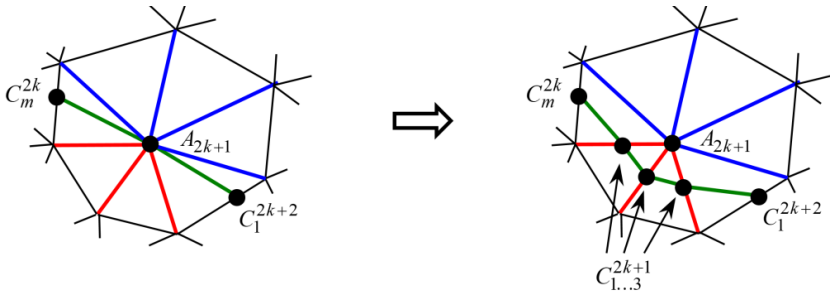


Рис. 5. Оптимизация на шатре с нечетным индексом вершины

Результат выполнения предыдущих операций назовем «слабой оптимизацией» пути: в результате путь сглаживается и укорачивается, однако остается проходящим по тем из треугольных граней, которые примыкают (по ребрам) к исходному пути. В простых случаях найденный таким образом путь может оказаться достаточно гладким, однако в общем случае требуется дальнейшая оптимизация.

4. «Вытянуть» путь. Для этого нужно построить функцию длины пути на соответствующей развертке, считая веса α_i^j переменными (их количество равно числу внутренних точек на пути), и найти ее условный минимум в гиперкубе $0 \le \alpha_i^j \le 1$. Несмотря на простоту постановки, эта задача нетривиальна в тех ситуациях, когда некоторые из параметров α_i^j выходят на ограничения: градиент функции длины пути в этом случае имеет неопределенность вида [0/0], а компоненты матрицы Гессе становятся бесконечными. Причина состоит в том, что выход весового параметра α_i^j на ограничение означает приближение соответствующей точки на оптимизируемом пути к одной из вершин. Однако в этом случае и несколько других весовых коэффициентов обязаны принять нулевое или единичное значение, что будет соответствовать сближению вплоть до совпадения нескольких смежных точек C_i^j и, следовательно, равенству нулю длины участка пути между этими точками.

Отметим также, что даже если проблема выхода параметров α_i^j на ограничения не возникает, или она каким-либо аналитическим либо алгоритмическим способом решена, задача безусловной оптимизации пути является весьма непростой для практического решения: если $\Delta\alpha_i^j$ – отклонения параметров от оптимальных значений, то отклонение целевой функции длины пути на развертке от минимального значения имеет вид:

$$\Delta\Psi = O\left(\sum\nolimits_{i,j} q_{i,j}\left(\Delta\alpha_i^j\right)^2\right),$$

где $q_{i,j}$ – некоторые коэффициенты порядка единицы.

При сравнительно небольшом числе звеньев ломаной ее численная минимизация не вызывает существенных сложностей даже при использовании простых методов типа наискорейшего спуска или сопряженных градиентов, однако с увеличением количества звеньев скорость сходимости градиентных методов становится весьма низкой. Разработка эффективных и робастных алгоритмов решения данной задачи представляет собой вопрос для отдельного исследования.

В то же время представляется, что данную задачу можно решить, воспользовавшись простой и наглядной механической аналогией, о которой упомянуто в [2]: путь следует представить в виде нити, которая проходит по развертке и натягивается в начальной и конечной точках. Результирующей формой пути будет кусочно-прямолинейная, «цепляющая» в своих угловых точках некоторые из вершин треугольников развертки (см. далее рис. 7).

5. «Смена развертки». Для точек ломаной, соответствующих нулевым и единичным весовым коэффициентам, т.е. совпадающих с вершинами ячеек, следует рассмотреть шатер с вершиной в этой точке и заменить в развертке треугольники из входящего в нее полушатра треугольниками из второго полушатра (рис. 6). При этом разумно сделать шаг слабой оптимизации, развернув «новый» полушатер и проложив по нему путь, чтобы он образовал отрезок прямой.

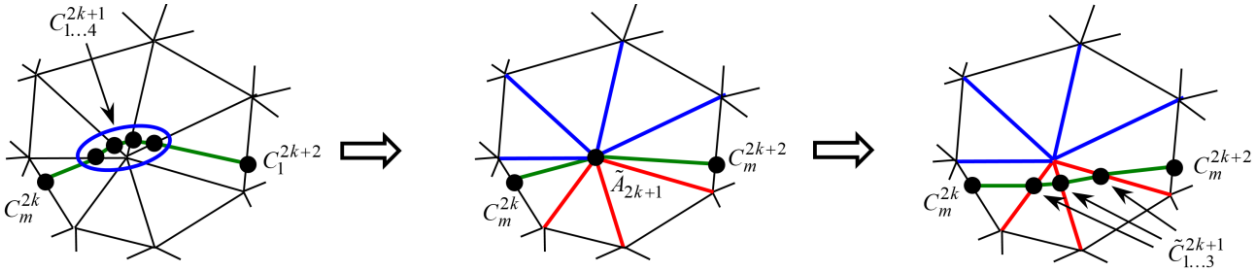


Рис. 6. Схема смены развертки

Отметим, что во избежание зацикливания алгоритма целесообразно за один шаг оптимизации, соответствующий данному пункту, изменять развертку лишь в одной «угловой» вершине предыдущего пути. После этого следует повторить п. 5 и вытянуть путь на новой развертке (рис. 7). Алгоритм завершается при получении на развертке прямолинейного пути без изломов, либо пути с изломами в вершинах, соответствующих седловым точкам исходной триангулированной поверхности, когда найти более короткий путь невозможно.

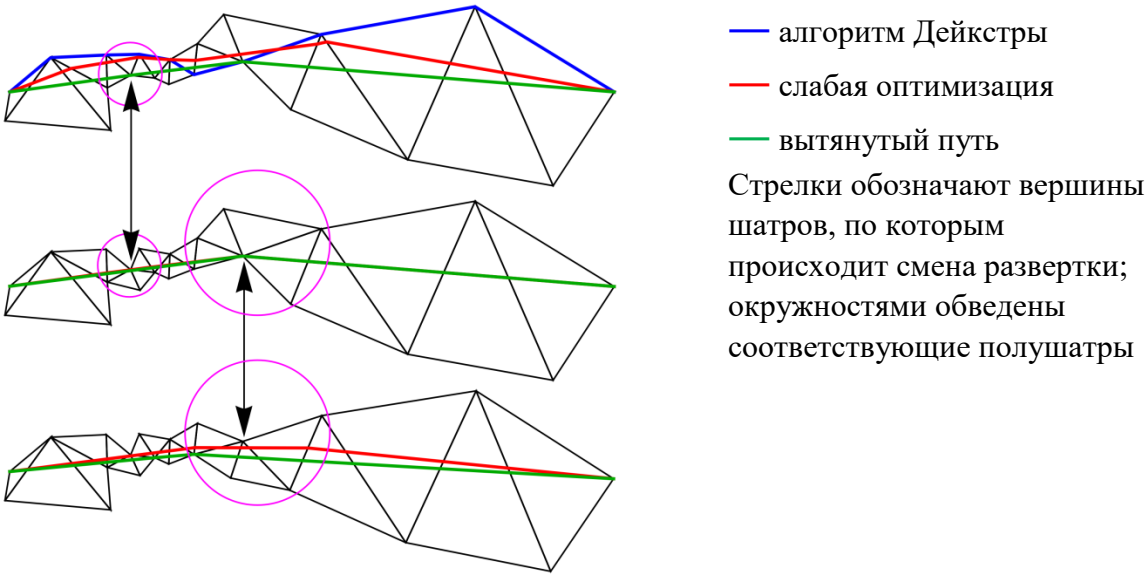


Рис. 7. Несколько первых шагов оптимизации пути

3. Результаты, полученные с помощью реализованного алгоритма. Рассмотрим примеры применения предложенного алгоритма для модели сферы с равномерной поверхностной сеткой, грубой модели тела сложной формы и весьма подробной модели винтового движителя со сгущением сетки на лопастях (рис. 8).

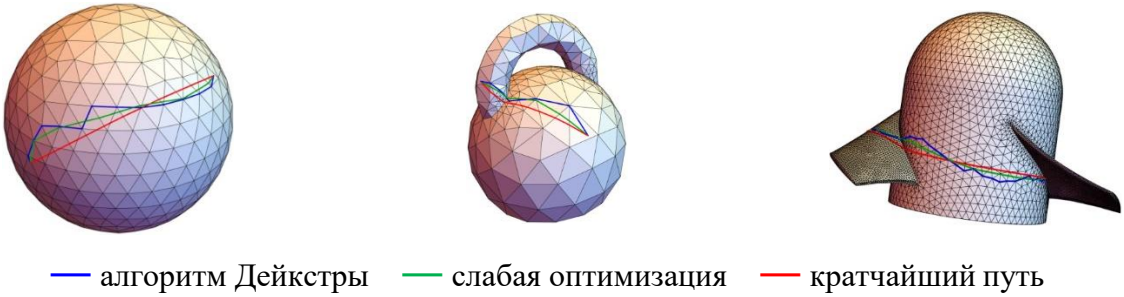


Рис. 8. Результаты оптимизации пути

Путь, найденный с помощью алгоритма Дейкстры, включает для данных тел 13, 9 и 44 сегмента соответственно. Кратчайший путь состоит из 25, 19 и 102 сегментов; для его отыскания после двух шагов слабой оптимизации выполнены 4, 7 и 99 итераций вытягивания пути. График изменения длины пути показан на рис. 9.

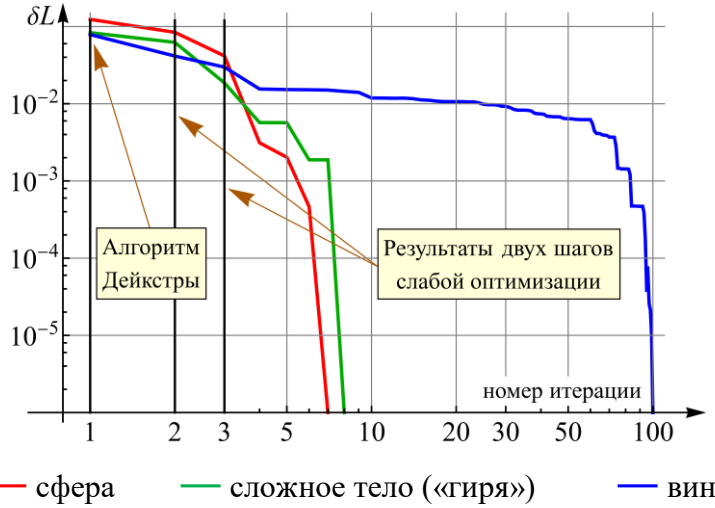


Рис. 9. Изменение длины в процессе оптимизации пути

Величина δL означает отношение превышения длины пути над минимальной длиной к этой минимальной длине. Видно, что путь, найденный с использованием алгоритма Дейкстры, в рассмотренных примерах отличается по длине от оптимального примерно на 10 %; после двух шагов слабой оптимизации отличие составляет 2…4 %. Для модели винта отличие на величину около 1% сохраняется в течение следующих 60 итераций и только затем резко снижается.

График изменения суммарного угла на развертке в зависимости от количества выполненных шагов алгоритма (итераций) представлен на рис. 10.

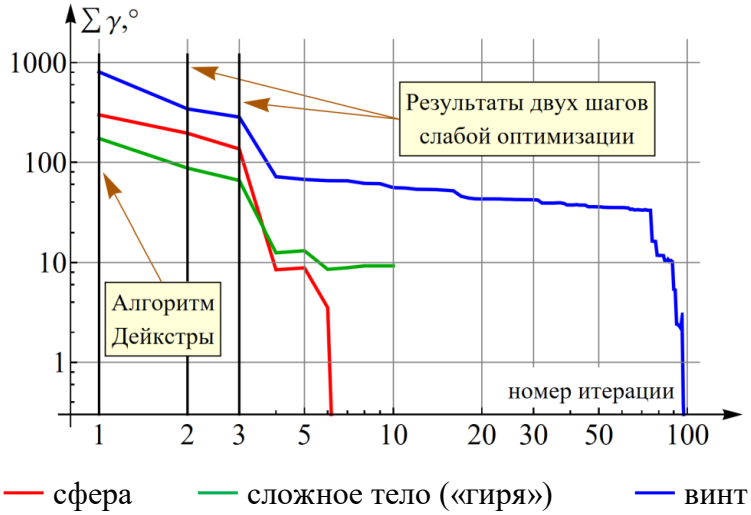


Рис. 10. Изменение суммарного угла (в градусах) на развертке в процессе оптимизации пути

Величина $\sum\gamma$ – это сумма углов (в градусах) между сегментами пути на той развертке, по которой на данном шаге производится вытягивание пути. Видно, что после двух шагов слабой оптимизации суммарный угол уменьшается всего в 2…3 раза, тогда как первый шаг вытягивания пути обеспечивает еще не менее, чем 5-кратное уменьшение суммарного угла (для модели сферы эта величина превышает 20 раз). Далее величина суммарного угла изменяется обычно немонотонно, в ряде случаев при смене развертки наблюдается небольшое ее увеличение. Это не связано с ухудшением самого пути, а объясняется тем фактом, что для одного и того же пути может существовать несколько разверток. Последнее имеет место в тех случаях,

когда путь проходит через «седловую» вершину: в этом случае смена развертки в ней и новая процедура вытягивания приводят к тому же самому пути, но рассматриваемому на другой развертке – т.е. длина пути остается прежней, а угол может и увеличиться (в вычислительном алгоритме реализовано недопущение зацикливания в таких ситуациях). Именно эта ситуация и наблюдается во всех рассмотренных примерах; кажущаяся неоптимальность угла на модели гири и винта связана именно с наличием на кратчайшем пути «седловой» вершины.

Заключение. Предложенный алгоритм поиска кратчайшего пути на триангулированной поверхности в рассмотренных модельных задачах позволил построить оптимальные в математическом смысле пути. Алгоритм итерационный; выполнение нескольких первых шагов позволяет существенно улучшить путь. Полная оптимизация пути, состоящего из большого количества сегментов, особенно для неравномерных поверхностных сеток, может требовать выполнения большого числа итераций, что негативно сказывается на времени его исполнения, поэтому задача оптимизации алгоритма остается актуальной. Кроме того, представленный алгоритм реализован в системе компьютерной алгебры Wolfram Mathematica, тогда как для его практического применения требуется создание его эффективной реализации или модификации на языке C++.

Список источников

1. Dergachev S.A., Marchevsky I.K., Shcheglov G.A. Flow simulation around 3D bodies by using Lagrangian vortex loops method with boundary condition satisfaction with respect to tangential velocity components. Aerospace science and technology, 2019, vol. 94, DOI: 10.1016/j.ast.2019.105374
2. Люстерник Л.А. Кратчайшие линии. Вариационные задачи. – М.: ГИТТЛ, 1955. – Вып. 19. – 104 с
3. Dijkstra E.W. A note on two problems in connexion with graphs. Numerische Mathematik, 1959, vol. 1, pp. 269-271, DOI:10.1007/BF01386390.
4. Hwang Y.H., Chang R.C., Tu H.Y. Finding all shortest path edge sequences on a convex polyhedron. Lecture notes in computer science. Workshop on algorithms and data structures, 1989, vol. 382, pp. 251-266, DOI:10.1007/3-540-51542-9_23.
5. Chen J., Han Y. Shortest paths on a polyhedron. Proc. of 6th ACM symp. on comput. Geometry. Berkley, 1990, pp. 360-369, DOI:10.1145/98524.98601.
6. Wenk C., Cook A.F. Shortest path problems on a polyhedral surface. Computational geometry. Dagstuhl seminar proceedings, 2009, vol. 9111, pp. 1-30, DOI: 10.4230/DagSemProc.09111.5.
7. An P.T., Van Hoai T., Ba Thinh V. The funnel tree algorithm for finding shortest paths on polyhedral surfaces. Optimization, 2023, DOI:10.1080/02331934.2023.2241496
8. An P.T., Le N.T., Trang L.H., Wong R.C.-W. Finding shortest gentle paths on polyhedral terrains by the method of multiple shooting. Journal of computational science, 2023, vol. 67, DOI:10.1016/j.jocs.2022.101935
9. CGAL: The Computational geometry algorithms library. Available at: https://www.cgal.org (accessed: 04/10/2024)
10. Xin S.-Q., Wang G.-J. Improving Chen and Han's algorithm on the discrete geodesic problem. ACM trans. Graph. 2009, vol. 28, no. 4, DOI: 10.1145/1559755.1559761.

Серебровская Екатерина Александровна. *Студент кафедры Математическое моделирование МГТУ им. Н.Э. Баумана, e.a.serebrovskaya@gmail.com, 105005, Москва, 2-я ул., д. Бауманская 5, к. 1.*

Марчевский Илья Константинович. *Доцент, д.ф.-м.н., МГТУ им. Н.Э. Баумана, профессор кафедры Прикладная математика, AuthorID: 183544, SPIN: 8988-9507, ORCID: 0000-0003-4899-4828, iliamarchevsky@mail.ru, 105005, Москва, 2-я Бауманская ул., д. 5, к. 1.*

Ерофеева Мария Александровна. *Студент кафедры Прикладная математика МГТУ им. Н.Э. Баумана, mariya.a.erofeeva@gmail.com, 105005, Москва, 2-я Бауманская ул., д. 5, к. 1.*

UDC 519.6
DOI:10.25729/ESI.2024.34.2.009

Path smoothing algorithm for the triangulated surface

Ekaterina A. Serebrovskaya, Ilia K. Marchevsky, Mariya A. Erofeeva

Bauman Moscow State Technical University,
Russia, Moscow, *e.a.serebrovskaya@gmail.com*

Abstract. An algorithm for optimizing the path length on a triangulated surface has been proposed and implemented in Wolfram Mathematica. The first two steps are “lightweight” and involve varying the path along the edges only within the triangles adjacent to them. Subsequent steps allow arriving at a shortest path in the mathematical sense in a few iterations. The convergence of the algorithm has not been rigorously proven, but has been ensured in a large number of examples considered.

Keywords: shortest path, triangulated surface, conditional optimization, Wolfram Mathematica

References

1. Dergachev S.A., Marchevsky I.K., Shcheglov G.A. Flow simulation around 3D bodies by using Lagrangian vortex loops method with boundary condition satisfaction with respect to tangential velocity components. Aerospace science and technology, 2019, vol. 94, DOI: 10.1016/j.ast.2019.105374
2. Lyusternik L.A. Kratchayshiye linii. Variatsionnyye zadachi. [Shortest Paths. Variational Problems]. M.: GITTL, 1964. 102 p.
3. Dijkstra E.W. A note on two problems in connexion with graphs. Numerische Mathematik, 1959, vol. 1, pp. 269-271, DOI:10.1007/BF01386390.
4. Hwang Y.H., Chang R.C., Tu H.Y. Finding all shortest path edge sequences on a convex polyhedron. Lecture notes in computer science. Workshop on algorithms and data structures, 1989, vol. 382, pp. 251-266, DOI:10.1007/3-540-51542-9_23.
5. Chen J., Han Y. Shortest paths on a polyhedron. Proc. of 6th ACM symp. on comput. Geometry. Berkley, 1990, pp. 360-369, DOI:10.1145/98524.98601.
6. Wenk C., Cook A.F. Shortest path problems on a polyhedral surface. Computational geometry. Dagstuhl seminar proceedings, 2009, vol. 9111, pp. 1-30, DOI: 10.4230/DagSemProc.09111.5.
7. An P.T., Van Hoai T., Ba Thinh V. The funnel tree algorithm for finding shortest paths on polyhedral surfaces. Optimization, 2023, DOI:10.1080/02331934.2023.2241496
8. An P.T., Le N.T., Trang L.H., Wong R.C.-W. Finding shortest gentle paths on polyhedral terrains by the method of multiple shooting. Journal of computational science, 2023, vol. 67, DOI:10.1016/j.jocs.2022.101935
9. CGAL: The Computational geometry algorithms library. Available at: https://www.cgal.org (accessed: 04/10/2024)
10. Xin S.-Q., Wang G.-J. Improving Chen and Han’s algorithm on the discrete geodesic problem. ACM trans. Graph. 2009, vol. 28, no. 4, DOI: 10.1145/1559755.1559761.

Serebrovskaya Ekaterina Alexandrovna. *Student of the department of Mathematical modeling, Bauman Moscow state technical university, e.a.serebrovskaya@gmail.com, 105005, Moscow, 2nd Baumanskaya St., 5, building 1.*

Marchevsky Ilia Konstantinovich. *Ph.D., D.Sc. (phys.-math.), associate professor, professor of the department of applied mathematics, Bauman Moscow state technical university, AuthorID: 183544, SPIN: 8988-9507, ORCID: 0000-0003-4899-4828, iliamarchevsky@mail.ru, 105005, Moscow, 2nd Baumanskaya St., 5, building 1.*

Erofeeva Mariya Alexandrovna. *Student of the department of applied mathematics, Bauman Moscow state technical university, mariya.a.erofeeva@gmail.com, 105005, Moscow, 2nd Baumanskaya St., 5, building 1.*

Статья поступила в редакцию 15.04.2024; одобрена после рецензирования 06.06.2024; принята к публикации 06.06.2024.

The article was submitted 04/15/2024; approved after reviewing 06/06/2024; accepted for publication 06/06/2024.

Математические, информационные и интеллектуальные технологии в энергетике

УДК: 621.311
DOI:10.25729/ESI.2024.34.2.010

Выбор компонентов гибридного энергокомплекса для удаленных районов с использованием методов многокритериальной оценки

Погодаева Инесса Николаевна[1], Шакиров Владислав Альбертович[2]

[1]Братский государственный университет, Россия, Братск, *m.inessa13@yandex.ru*
[2]Иркутский национальный исследовательский технический университет, Россия, Иркутск

Аннотация. Проблема энергоснабжения районов, удаленных от энергосистемы, сохраняет актуальность и находит свое решение преимущественно в модернизации дизельных электростанций и построении гибридных энергокомплексов с возобновляемыми источниками энергии. Жизненный цикл гибридных энергокомплексов составляет несколько десятков лет, поэтому при проектировании должны быть учтены все основные цели – экономическая и техническая эффективность, минимизация воздействий на окружающую среду, надежность. Многокритериальность проблемы требует привлечения лиц, принимающих решения, для выражения предпочтений относительно важности показателей эффективности различных вариантов энергокомплекса. Стохастический характер возобновляемой генерации, многообразие режимных ограничений приводят к необходимости использования специального программного обеспечения для моделирования режимов с часовым разрешением. В статье рассматривается подход к многокритериальному выбору гибридного энергокомплекса из множества альтернатив, формируемых в программе HOMER PRO, с использованием трех методов: TOPSIS и PROMETHEE I, II. Методы имеют отличия в процедурах оценки альтернатив и этим обеспечивают повышение обоснованности выбора. Численный пример рассмотрен для поселка Усть-Соболевка в Приморском крае. Варианты гибридного энергокомплекса формируются с использованием технологий дизельной генерации, ветроэнергетических установок, фотоэлектрических преобразователей, микро-гидроэлектростанций и накопителей энергии. В исследовании рассматриваются три сценария развития района с различными оценками важности критериев.

Ключевые слова: гибридный энергокомплекс, многокритериальный анализ, электростанция, моделирование, возобновляемые источники энергии, TOPSIS, PROMETHEE

Цитирование: Погодаева И.Н. Выбор компонентов гибридного энергокомплекса для удаленных районов с использованием методов многокритериальной оценки / И.Н. Погодаева, В.А. Шакиров // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 99-111. – DOI:10.25729/ESI.2024.34.2.010.

Введение. Развитие технологий использования возобновляемых источников энергии (ВИЭ) на протяжении десятилетий не теряет актуальности во всем мире. Большая часть возобновляемой энергии, которую в перспективе можно использовать, приходится на солнечную – более 62%. За последние 10 лет суммарная установленная мощность электростанций, работающих на возобновляемых источниках энергии, выросла в 2,15 раза[1] [1]. Многие страны нацеливают свою энергетическую политику таким образом, чтобы доля ВИЭ в общем энергетическом балансе с каждым годом возрастала. Достижение углеродной нейтральности в энергетической сфере является одним из важных приоритетов.

В России в последние годы развиваются механизмы поддержки возобновляемой энергетики, вводимой для функционирования как на оптовом, так и на розничных рынках электроэнергии. Благодаря программе реализации договоров о предоставлении мощности построены 70 солнечных электростанций (СЭС), 25 ветряных электростанций (ВЭС), 5 малых гидроэлектростанций [2]. В 2023 году введены в эксплуатацию 340,3 МВт объектов ВИЭ, в том числе: на оптовом рынке: ВЭС – 252 МВт; малых гидроэлектростанций – 40,9 МВт; на

[1] МЭА (2023 г.), World Energy Outlook 2023, МЭА, Париж,URL: https://www.iea.org/reports/world-energy-outlook-2023

розничных рынках электроэнергии: СЭС – 44,1 МВт; микрогидроэлектростанций ГЭС (мГЭС) – 3,3 МВт [2].

В России создаются условия для внедрения передовых технологий на основе ВИЭ в удаленных от энергосистемы районах. Привлечение инвестиций осуществляется путем заключения энергосервисных контрактов, в рамках которых проводится модернизация дизельных электростанций и создание гибридных энергокомплексов (ГЭК) с установками на основе возобновляемых источников энергии. Это позволяет существенно снижать расход дорогостоящего дизельного топлива, уменьшать воздействие на окружающую среду.

При выборе компонентов ГЭК требуется рассмотрение множества альтернативных технологий с учетом их технико-экономических характеристик и природно-климатических условий территории. Выбор компонентов ГЭК усложняется необходимостью рассмотрения нескольких критериев эффективности, как на стадии планирования, так и на стадии функционирования энергокомплекса.

В научной литературе теме ГЭК посвящены многочисленные исследования. В работе [3] дан обзор основных подходов к оптимизации параметров ГЭК. В работе [4] рассмотрены подходы к оптимизации состава ГЭК с углубленным рассмотрением эвристических подходов и их комбинаций. В исследовании [5] проводится сравнение трех эвристических алгоритмов оптимизации для выбора состава ГЭК. Оптимальные параметры системы выбираются по трем критериям: нормированная стоимость производства энергии (LCOE), недоотпуск электроэнергии потребителям, величина произведенной избыточной энергии. В работе [6] предлагается комплексное использование ВИЭ для островов Греции с целью покрытия потребностей населения в водоснабжении. Оптимизация проводится для двух сценариев функционирования системы ГЭК на основе авторегрессии второго порядка и модели авторегрессии скользящего среднего. В исследовании [7] представлен подход к управлению функционированием ГЭК с применением искусственной нейронной сети и эвристического алгоритма каракатиц. В работе [8] представлены анализ и моделирование ГЭК для сельскохозяйственного комплекса во Вьетнаме с целью покрытия потребности в энергии для полива, освещения, механизации и других сельскохозяйственных процессов. Моделирование выполнено в программном комплексе HOMER PRO. Авторы исследования [9] предлагают многокритериальный подход к проектированию ГЭК с учетом трех критериев: вероятность дефицита мощности, величина избыточно произведенной энергии и разница между генерируемой мощностью и емкостью накопителей энергии. Результаты исследования показывают, что реализация многокритериальной оптимизации приводит к более практичным и надежным результатам. Исследование [10] рассматривает три ГЭК с разными источниками энергии. Для предотвращения излишнего использования батарей и увеличения срока службы оборудования рассматриваются различные стратегии управления энегрокомплексом. Для проведения многокритериального анализа используются девять показателей оценки технологий, экономики и окружающей среды. Результаты исследования показывают, что системные модели, стратегия управления питанием и комплексные показатели эффективности могут служить основой для моделирования ГЭК. В исследовании [11] предложена методика выбора оборудования ГЭК на основе многокритериального анализа. В исследовании использовались пять показателей для оценки различных вариантов решения, включая стоимость ресурсов, объем инвестиций, сокращение выбросов парниковых газов, долю ВИЭ и влияние на используемые земли.

Одним из ключевых исследований по выбору состава ГЭК выступает работа [12] с широким обзором исследований и анализом более 500 статей о ГЭК за период с 1995 по 2020 год. В статье выполнено категорирование исследований по географическому расположению, конфигурации и компонентам, используемым методам.

На основе анализа зарубежных источников можно сделать вывод о целесообразности применения многокритериального подхода к выбору состава оборудования ГЭК. Такой подход позволяет учитывать и сравнивать различные варианты по экономическим, техническим, экологическим и социальным критериям.

В статье будет предложена методика выбора компонентов ГЭК с использованием программы HOMER PRO и трех методов многокритериального анализа – TOPSIS, PROMETHEE 1 и 2. Применение нескольких методов позволяет выделить наиболее эффективные конфигурации ГЭК, показывающие эффективность независимо от способа формализации задачи и выражения предпочтений лица, принимающего решения. Применение методики рассматривается на примере создания ГЭК в удаленном районе Приморского края.

Методика проведения исследования. Выбор компонентов ГЭК на основе многокритериального анализа включает в себя несколько этапов (рисунок 1).

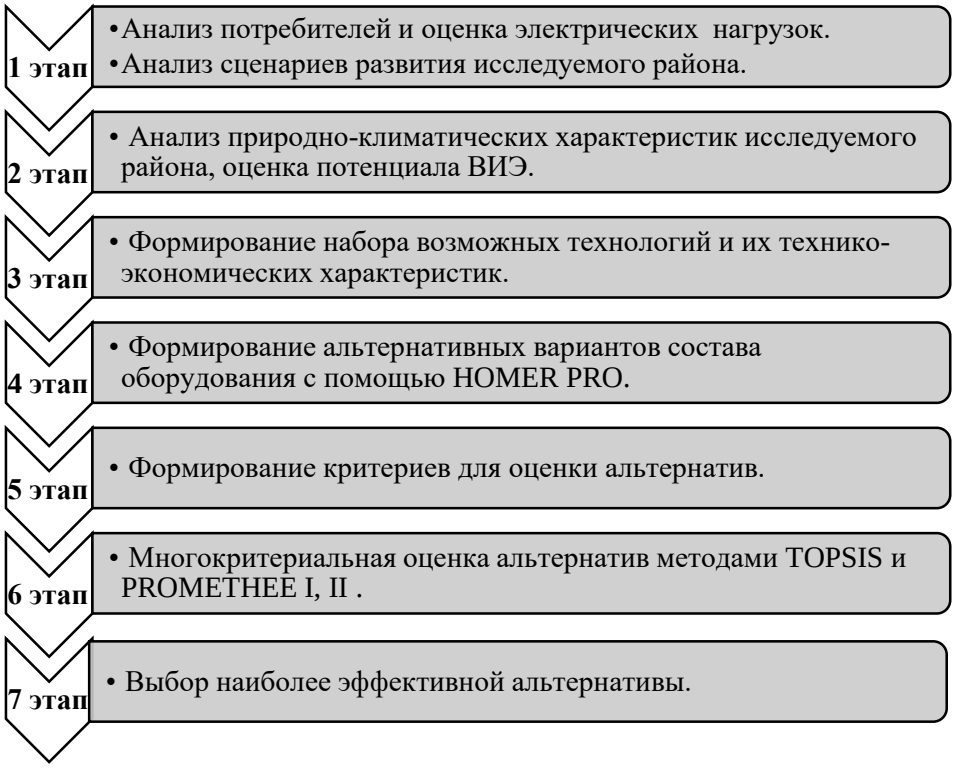


Рис. 1. Этапы методики многокритериального выбора компонентов ГЭК

На первом этапе выполняется анализ потребителей, режимов их работы, требований к надежности электроснабжения. Проводится оценка годового графика электрических нагрузок с часовым разрешением.

На втором этапе проводится анализ природно-климатических характеристик исследуемого района. Выполняется оценка гелиоэнергетического, ветроэнергетического, гидроэнергетического потенциала, а также топливных ресурсов. В результате оценки может быть сделан вывод о наборе рассматриваемых технологий в составе ГЭК.

На третьем этапе формируется набор возможных технологий и их технико-экономических характеристик. К таким характеристикам относят удельные капитальные вложения, затраты на обслуживание и эксплуатацию оборудования, затраты на замену оборудования, удельный расход топлива дизельных генераторов и другие.

На четвертом этапе создаются альтернативные варианты состава оборудования с помощью HOMER PRO. HOMER PRO – это программный продукт, который предназначен для анализа и оптимизации ГЭК. HOMER Pro осуществляет однокритериальную оптимизацию по показателю суммарных дисконтированных затрат за период жизненного цикла проекта. В

связи с этим в ряде исследований, например, [8] были предложены подходы к многокритериальному выбору конфигурации ГЭК из множества, формируемого HOMER Pro.

На пятом этапе формируются критерии для оценки альтернатив. Обзор исследований показал, что основные рассматриваемые критерии для выбора компонентов ГЭК: LCOE, объем выбросов CO_2, доля ВИЭ в общей генерации электроэнергии. Одним из основных критериев практически во всех исследованиях является LCOE.

LCOE (Levelized Cost of Energy) позволяет сравнивать стоимость производства энергии от различных источников и определять наиболее экономически эффективные варианты для реализации. Этот показатель может учитывать такие факторы, как срок службы оборудования, ставки по кредитам, операционные расходы и другие параметры, что делает его полезным инструментом для принятия решений в энергетической отрасли. Данный показатель рассчитывается по следующей формуле [13]:

$$LCOE = \frac{\sum_{t=1}^{T}(IC_t + C_{ft} + C_{omt})(1+n)^{1-t}}{\sum_{t=1}^{T} W_t(1+dr)^{1-t}}, \quad (1)$$

где IC_t – капитальные вложения в t-ый год, млн руб.; C_{ft} – затраты на топливо с учетом доставки в t-ый год, млн руб.; C_{omt} – затраты на обслуживание и ремонт основного оборудования, млн руб.; W_t – производство электроэнергии за t-ый год, млн. кВт·ч, n – жизненный цикл системы; dr – ставка дисконтирования.

На шестом этапе (рис. 1) проводится многокритериальная оценка альтернатив методами TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) и PROMETHEE (Preference Ranking Organisation METHod for Enrichment Evaluations) I и II модификаций.

Перед использованием этих методов необходимо определить веса для каждого критерия. Веса отражают относительную важность каждого критерия в принятии решения. Веса можно определить с помощью экспертных оценок или с помощью математических методов, таких как AHP (Analytic Hierarchy Process) [14].

Математически метод TOPSIS реализуется по следующим этапам [15]:

Нормализация оценок по критериям:

$$n_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \quad (2)$$

где x_{ij} – оценка альтернативы i по критерию j; m – количество альтернатив.

Определение взвешенной матрицы нормированных решений с учетом весов критериев, элементы которой определяются как:

$$u_{ij} = w_j \times n_{ij}, i = 1 \dots m; j = 1 \dots n \quad (3)$$

где w_j – вес j-го критерия; n_{ij} – нормализованное значение оценки альтернативы i по критерию j; n – количество критериев.

Определение идеального позитивного (ИПР) и негативного (ИНР) решений:

$$A^+ = ((\max_i u_i / j \in L), (\min_i u_i / j \in C)) \quad (4)$$

$$A^- = ((\min_i u_i / j \in L), (\max_i u_i / j \in C)) \quad (5)$$

где L соответствует критериям выгоды, C – критериям затрат.

Выполняются расчеты:

1) расчет расстояний альтернатив до ИПР и ИНР:

$$d_i^+ = \sqrt{\sum_{j=1}^{n}(u_{ij} - A_j^+)^2,\ j = 1,2, \dots, m} \quad (6)$$

$$d_i^- = \sqrt{\sum_{j=1}^{n}(u_{ij} - A_j^-)^2,\ j = 1,2, \dots, m}\ , \quad (7)$$

2) определение интегрального показателя (коэффициента близости) (8) для каждой сравниваемой альтернативы:

$$R_i = \frac{d_i^-}{d_i^- + d_i^+} \quad (8)$$

PROMETHEE I и PROMETHEE II – это две основные вариации метода PROMETHEE.

PROMETHEE I создает частичное упорядочение, что означает, что не всегда удается определить однозначно, какая альтернатива предпочтительнее, поскольку могут быть ситуации несравнимости.

PROMETHEE II является методом полного ранжирования альтернатив. Особенностью PROMETHEE II является устранение неопределенностей, которые возникали в PROMETHEE I, предоставляя полное ранжирование альтернатив.

Основные этапы PROMETHEE можно изложить так [16]:

1) определение критериев и их весов;
2) выбор функции предпочтения для каждого критерия;
3) расчет относительных индексов предпочтения для всех пар альтернатив по каждому критерию по одной из формул:

$$\pi(a,b) = \frac{1}{n}\sum_{j=1}^{n} f_j(a,b), \quad (9)$$

$$\pi(a,b) = \frac{1}{n}\sum_{j=1}^{n} w_j f_j(a,b), \quad (10)$$

4) вычисление совокупного индекса предпочтения для каждой пары альтернатив с использованием весов критериев;
5) вычисление положительного (11) и отрицательного (12) потоков предпочтения для каждой альтернативы:

$$\varphi^+(a) = \sum_{b\in A} \pi(a,b), \quad (11)$$

$$\varphi^-(a) = \sum_{b\in A} \pi(b,a), \quad (12)$$

величина $\varphi^+(a)$ показывает, как альтернатива a соотносится с уступающими ей альтернативами, $\varphi^-(a)$ – как превосходящие a альтернативы соотносятся с ней;

6) расчет чистого потока предпочтения для PROMETHEE II и формирование полного ранжирования альтернатив.

$$\varphi(a) = \varphi^+(a) - \varphi^-(a). \quad (13)$$

На заключительном этапе проводится выбор наиболее эффективной альтернативы. Процедура сводится к выбору такой конфигурации ГЭК, которая имеет высокую оценку или ранг по результатам трех методов.

Результаты и обсуждение. Применение методики рассматривается на примере разработки ГЭК для электроснабжения поселения Усть-Соболевка, расположенного в Приморском крае. Село Усть-Соболевка стоит на левом берегу реки Соболевка примерно в 1 км до впадения её в бухту Соболевка Японского моря. В связи с удаленностью села от энергосистемы для обеспечения энергией населения Усть-Соболевки используются дизельные генераторы.

Следует отметить, что населенный пункт имеет высокий потенциал для развития экотуризма, поэтому при рассмотрении задачи создания ГЭК необходимо учитывать не только технические и экономические факторы, но и экологические и социальные аспекты (рисунок 1 – этап 1, этап 2).

На основе анализа перспективного электропотребления были рассчитаны электрические нагрузки потребителей села Усть-Соболевка. На рис. 2 показаны суточные графики нагрузок для каждого месяца, сформированные в HOMER PRO.

HOMER PRO позволяет сформировать данные показателей потенциала ВИЭ на основе специальных климатических баз данных. На рис. 3-6 показаны данные солнечной радиации, скорости ветра, температуры воздуха.

Технико-экономические характеристики (рис. 1, этап 3) рассматриваемого при оптимизации ГЭК оборудования представлены в таблице 1.

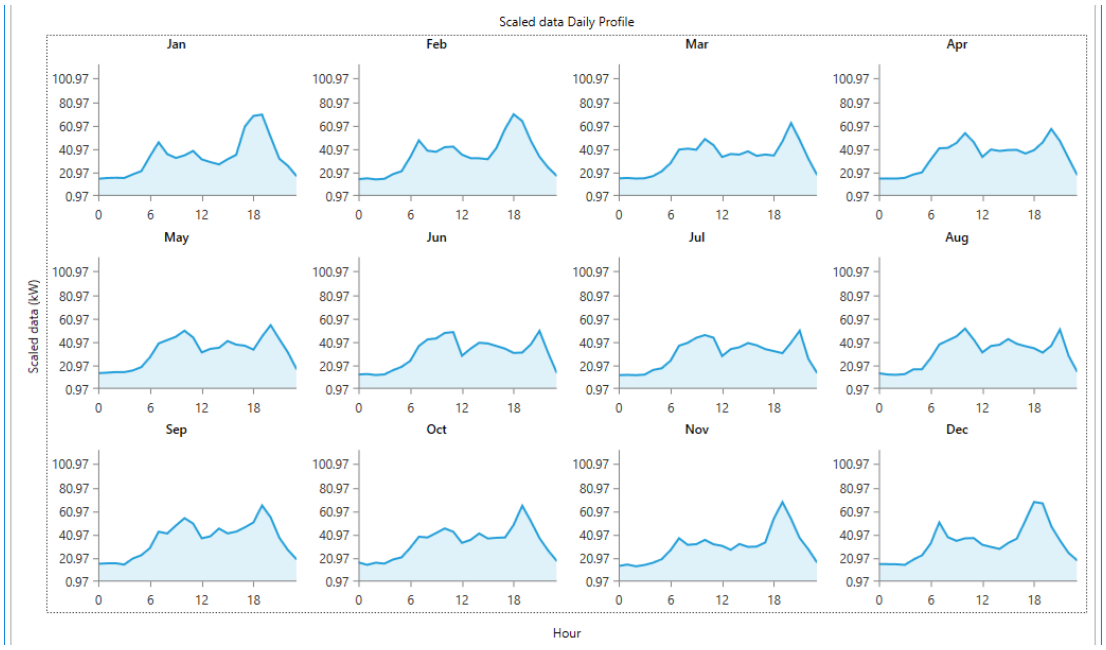


Рис. 2. Суточные графики электрических нагрузок в HOMER PRO

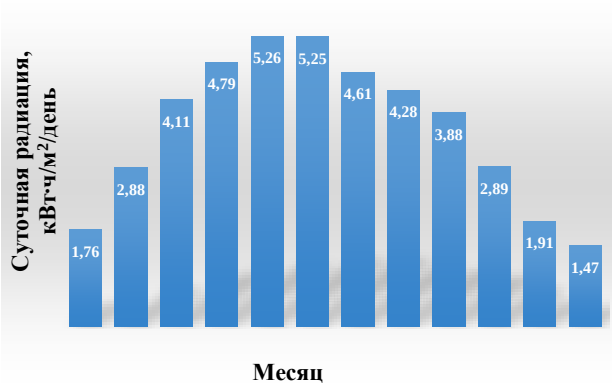


Рис.3. Среднемесячные значения суммарной солнечной радиации

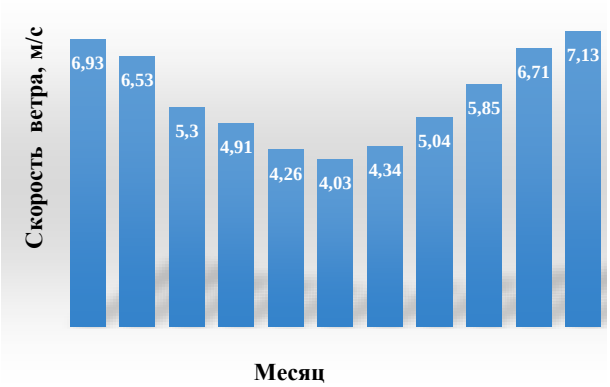


Рис.4. Среднемесячные значения скорости ветра

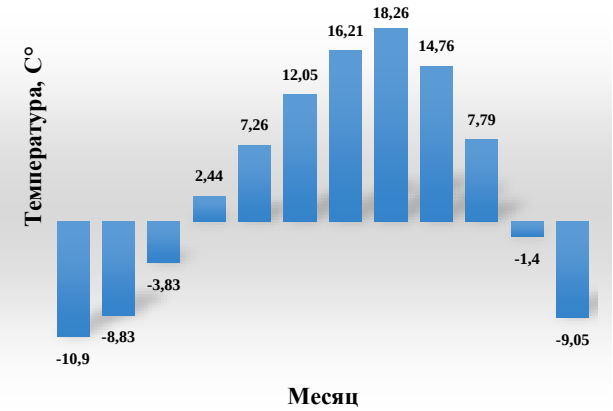


Рис.5. Среднемесячные значения температуры воздуха

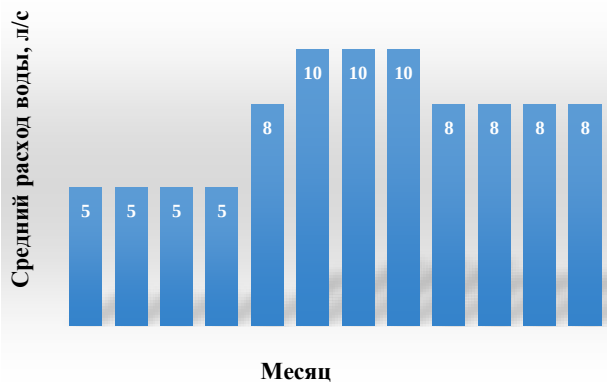


Рис.6. Среднемесячные значения скорости течения реки Соболевка

В ходе исследования с использованием программы HOMER PRO были выделены 20 альтернативных конфигураций ГЭК (рис. 1, этап 4). Каждая конфигурация представлена определенным набором оборудования и имеет оценки по критериям (таблица 2). Экономическую эффективность конфигурации отражают капитальные затраты и LCOE.

Выбросы CO_2 характеризуют воздействие ГЭК на климат и экологию. Количество запусков ДГУ отражает техническую эффективность конфигурации, так как большое количество запусков ведет к большему числу переходных режимов с возможным значительным отклонением параметров. Доля энергии ВИЭ характеризует социальную эффективность варианта ГЭК.

Таблица 1. Технико-экономические характеристики оборудования

1. ДГУ – дизель-генераторная установка, модель «Generic Small Genset»		
1.1	Удельные капитальные затраты, долл./кВт	475
1.2	Затраты на замену, долл./кВт	450
1.3	Стоимость эксплуатации и обслуживания, долл./кВт·ч	0,031
1.4	Срок службы, ч.	15000
1.5	Цена на дизельное топливо, долл./л	0,875
2. ФЭП – солнечные панели, модель «Generic flat plate»		
2.1	Удельные капитальные затраты, долл./кВт	1400
2.2	Затраты на замену, долл./ кВт	1260
2.3	Стоимость эксплуатации и обслуживания, долл./кВт/год	14
3. ВЭУ – ветроэнергетическая установка, модель «Generic 10 kW»		
3.1	Удельные капитальные затраты, долл./кВт	3000
3.2	Затраты на замену, долл./кВт	2700
3.3	Стоимость эксплуатации и обслуживания, долл./кВт/год	60
4. АКБ – аккумуляторные батареи, модель «Generic 1 kWh Li-ion»		
4.1	Удельные капитальные затраты, долл./кВт·ч	250
4.2	Затраты на замену, долл./кВт·ч	225
4.3	Стоимость эксплуатации и обслуживания, долл./кВт·ч/год	5
5. ПБ – преобразователь, модель «Leonics MTP-413F 25kW»		
5.1	Удельные капитальные затраты, долл./кВт	250
5.2	Затраты на замену, долл./кВт	250
5.3	Стоимость эксплуатации и обслуживания, долл./ кВт/год	2
6. МГЭС – микрогидроэлектростанция, модель «10 kW Generic Hyd10».		
6.1	Удельные капитальные затраты, долл./кВт	2500
6.2	Затраты на замену, долл./кВт	2000
6.3	Стоимость эксплуатации и обслуживания, долл./ кВт/год	240

Для исследования качества решений, получаемых с помощью методики многокритериального выбора, были разработаны три сценария. Сценарии отличаются различным вниманием к экономическим, социальным, экологическим и техническим аспектам проблемы. Соответственно в каждом из сценариев назначены различные веса критериев. В первом сценарии максимальный вес назначен критериям «Капитальные вложения» и «LCOE», во втором сценарии наибольший вес дан экологическим и социальным критериям (эмиссия CO2, доля ВИЭ), а в третьем сценарии все критерии имеют одинаковый вес. Веса критериев для сценариев представлены в таблице 3.

Далее с помощью методов TOPSIS и PROMETHEE конфигурации были оценены, проведено их ранжирование (рис. 1, этап 6). Результаты оценки альтернатив методом TOPSIS (значения коэффициента близости (8)), показаны на рис. 7. В таблице 4 показаны данные лучших альтернатив для трех сценариев. На рис. 8 представлены результаты применения метода PROMETHEE II. Результаты расчета методами PROMETHEE II представлены в таблице 5.

Таблица 2. Оценки альтернатив по критериям

Альтернатива	Капитальные вложения, тыс. долл.	LCOE, долл/кВт·ч	Выбросы CO_2 тонны/год	Количество запусков ДГУ в год	Доля энергии от ВИЭ, %
1	503,95	0,218	24,58	350	91,7
2	503,45	0,233	41,91	95	82,4
3	554,01	0,248	32,43	322	88,9
4	576,11	0,266	41,15	255	85,5
5	256,51	0,277	127,43	882	46,2
6	890,81	0,311	0	0	100
7	291,16	0,324	154,31	715	34,8
8	114,97	0,343	212,42	618	9,74
9	1006,35	0,36	0	0	100
10	125,49	0,392	246,98	522	0
11	1079,54	0,395	0	0	100
12	1257,83	0,464	0	0	100
13	83,44	0,51	270,65	4	7,16
14	104,90	0,513	266,35	8	9,20
15	51,21	0,515	284,03	4	0,79
16	81,21	0,516	275,27	8	4,95
17	95,99	0,524	280,35	1	2,55
18	119,24	0,527	274,73	1	5,24
19	77,50	0,533	291,91	1	0
20	47,50	0,534	302,44	1	0

Таблица 3. Веса критериев при различных сценариях

Капитальные вложения	LCOE	Выбросы CO_2	Количество запусков ДГУ	Доля энергии от ВИЭ
Сценарий 1, ориентированный на обеспечение экономической эффективности				
0,4	0,3	0,05	0,2	0,05
Сценарий 2, направленный на обеспечение социальной и экологической эффективности				
0,1	0,2	0,35	0,05	0,3
Сценарий 3 – равный вес всех критериев				
0,2	0,2	0,2	0,2	0,2

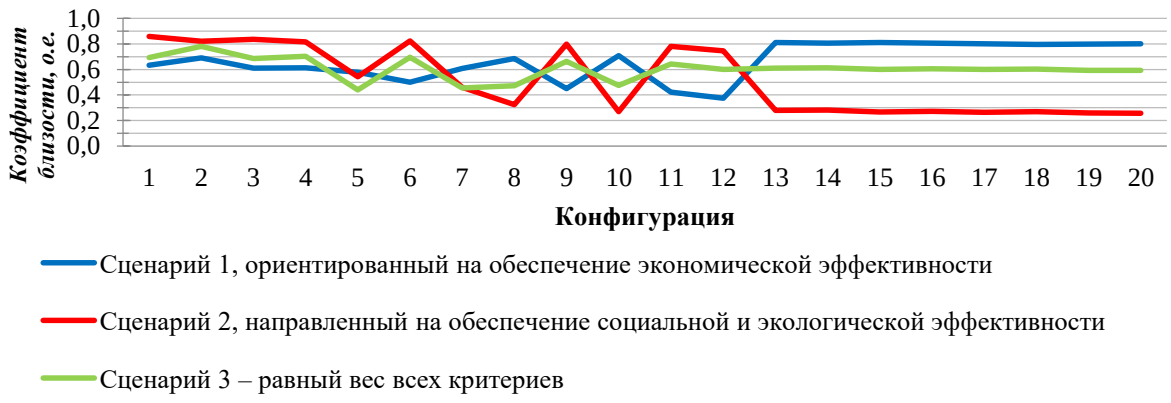


Рис. 7. Многокритериальные оценки конфигураций ГЭК, полученные методом TOPSIS

Таблица 4. Результаты выбора альтернатив методом TOPSIS

ФЭП	АКБ	ВЭУ	ПБ	ДГУ	МГЭС	Капитальные вложения, тыс. долл.	LCOE, долл./ кВт·ч	Выбросы CO2, тонн/год	Количество запусков ДГУ за год	Доля энергии от ВИЭ, %
Сценарий 1 – «экономическая эффективность» – Конфигурация 15										
0	0	0	4,84	100	3,92	51,21	0,515	284,03	4	0,795
Сценарий 2 «социальная и экологическая эффективность» – Конфигурация 1										
164	510	4	85,6	75	3,92	503,95	0,218	24,58	350	91,7
Сценарий 3 «баланс» – Конфигурация 2										
228	711	0	104	50	3,92	503,45	0,233	41,91	95	82,4

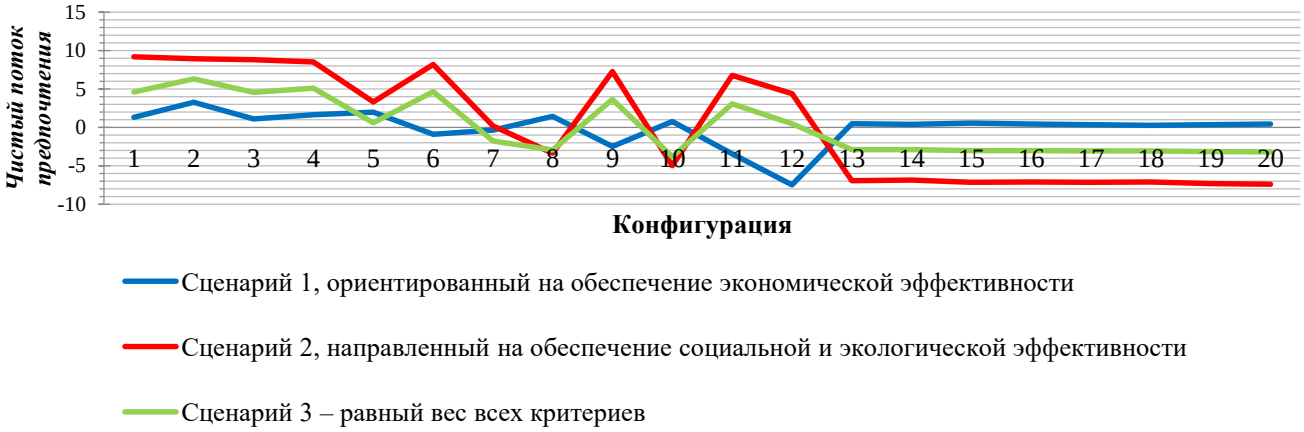


Рис. 8. Многокритериальные оценки рассматриваемых конфигураций ГЭК методом PROMETHEE II

Таблица 5. Результаты расчета методом PROMETHEE I и PROMETHEE II

ФЭП	АКБ	ВЭУ	ПБ	ДГУ	МГЭС	Капитальные вложения, тыс. долл.	LCOE, долл./ кВт·ч	Выбросы CO_2, тонн/год	Количество запусков ДГУ за год	Доля энергии от ВИЭ, %
Сценарий 1 – «экономическая эффективность» – Конфигурация 2										
228	711	0	104	50	3,92	503,45	0,233	41,91	95	82,4
Сценарий 2 «социальная и экологическая эффективность» – Конфигурация 1										
164	510	4	85,6	75	3,92	503,95	0,218	24,58	350	91,7
Сценарий 3 «баланс» – Конфигурация 2										
228	711	0	104	50	3,92	503,45	0,233	41,91	95	82,4

Результаты проведенного многокритериального анализа методами TOPSIS, PROMETHEE I и II представлены в таблице 6.

Результаты проведенного исследования методами многокритериального анализа показали, что лидирующие конфигурации во всех сценариях имеют некоторые расхождения. Это является следствием разных подходов к оценке альтернатив в методах. Например, на многокритериальную оценку каждой альтернативы методом PROMETHEE большое влияние оказывает количество проигравших ей или превосходящих ее альтернатив. На это, в свою очередь, сильно влияют функции предпочтения критериев. Для TOPSIS этот фактор не имеет такого выраженного значения. С другой стороны, TOPSIS оценивает расстояние в пространстве критериев от каждой альтернативы до идеального худшего и лучшего решений. Применение разных методов дает возможность выделить альтернативы, имеющие высокие оценки независимо от подходов к многокритериальному сравнению.

Таблица 6. Результаты многокритериального выбора наиболее эффективных альтернатив

Рейтинг конфигураций		1	2	3
		Конфигурация	Конфигурация	Конфигурация
TOPSIS	Экономическая эффективность	15	13	16
	Социальная и экологическая эффективность	1	3	6
	Равенство весов коэффициентов	2	4	6
PROMETHEE I	Экономическая эффективность	2	1	3
	Социальная и экологическая эффективность	1	3	2
	Равенство весов коэффициентов	2	3	1
PROMETHEE II	Экономическая эффективность	2	5	4
	Социальная и экологическая эффективность	1	3	2
	Равенство весов коэффициентов	2	4	6

Заключение. Применение ВИЭ в дополнение к объектам традиционной генерации, как в зоне централизованного электроснабжения, так и на изолированных и труднодоступных территориях Крайнего Севера, Дальнего Востока и Арктической зоны является все более распространенным. Этот тренд свидетельствует о переходе к новому технико-экономическому укладу, который создаст возможности для социально-экономического роста и повышения качества жизни населения. Для возможности учета разнообразных технических, экономических и социальных факторов в статье предложена методика выбора компонентов ГЭК с применением методов многокритериальной оценки TOPSIS и PROMETHEE.

В предлагаемой методике в качестве основы для формирования альтернатив используется программа HOMER PRO, которая оптимизирует состав ГЭК только на основе экономического критерия. Используя различные ограничения, HOMER PRO позволяет формировать множества разнообразных конфигураций ГЭК, отличающиеся по различным аспектам. Сравнение таких конфигураций и выбор лучшей альтернативы предлагается осуществлять с использованием трех методов. Эти методы имеют различные подходы к оценке альтернатив, что дает возможность повысить обоснованность выбираемых конфигураций ГЭК.

Исследование качества решений, получаемых с помощью предложенного подхода, проводилось для села Усть-Соболевка Приморского края. С использованием HOMER PRO были сформированы 20 альтернатив, существенно отличающихся технико-экономическими показателями. Сравнение альтернатив, выполненное для трех сценариев, показало некоторые отличия в оценках методами TOPSIS и PROMETHEE I, II. Выделенные предлагаемым образом

конфигурации ГЭК представляют наибольший интерес для дальнейшего углубленного сравнения.

Методы многокритериального анализа являются эффективным инструментом для принятия решений при моделировании ГЭК, так как позволяют учитывать различные факторы с учетом их важности для лица, принимающего решение.

Список источников

1. World energy outlook 2023, МЭА, Париж. – URL: https://www.iea.org/reports/world-energy-outlook-2023, Лицензия: CC BY 4.0 (отчет), CC BY NC SA 4.0 (Приложение А)
2. Ежеквартальный информационный обзор рынкаВИЭ в России IV квартал 2023 года (октябрь - декабрь 2023г.), – URL: https://rreda.ru/products/quarterly-reviews/review-1552.
3. Pranav M.S., Karunanithi K., Akhil M. et al. Hybrid renewable energy sources (HRES). International conference on intelligent computing, instrumentation and control technologies (ICICICT), Kerala State, Kannur, India, IEEE, 2017, pp. 162–165, DOI:10.1109/ICICICT1.2017.8342553.
4. Bouaouda A, Sayouti Y. Hybrid meta-heuristic algorithms for optimal sizing of hybrid renewable energy system: a review of the state-of-the-art. Archives of computational methods in engineering, 2022, vol. 29(6), pp. 4049-4083, DOI: 10.1007/s11831-022-09730-x.
5. Diab Ah., Hamdy S., Mohamed S.Ih., Kuznetsov O.N. et al. Application of different optimization algorithms for optimal sizing of pv/wind/diesel/battery storage stand-alone hybrid microgrid. IEEE access, 2019, 99 p., DOI:10.1109/ACCESS.2019.2936656.
6. Agapitidou A.A., Skroufouta S., Baltas E. Methodology for the development of hybrid renewable energy systems (HRES) with pumped storage and hydrogen production on Lemnos Island. Earth 3, 2022, pp. 537-556, DOI:10.3390/earth3020032.
7. Vijayaragavan M., Darly S. Power flow management of grid connected HRES system using a CFANN technique. International journal of electronics, 2021, DOI:10.1080/00207217.2020.1831080.
8. Sharma K.K., Gupta A., Kumar R. et al. Economic evaluation of a hybrid renewable energy system (HRES) using hybrid optimization model for electric renewable (HOMER) software—a case study of rural India. International journal of low-carbon technologies, 2021, vol. 16, iss. 3, pp. 814-821, DOI:10.1093/ijlct/ctab012.
9. Bakhtiari H., Naghizadeh R.A. Multi-criteria optimal sizing of hybrid renewable energy systems including wind, photovoltaic, battery, and hydrogen storage with ε-constraint method. IET Renewable power generation, 2018, 12(8), pp 883–892, DOI:10.1049/iet-rpg.2017.0706
10. Li Z., Wang Y., Xie J. et al. Hybrid multi-criteria decision-making evaluation of multiple renewable energy systems considering the hysteresis band principle. International journal of hydrogen energy, 2024, vol. 49(Part D), pp. 450-462, DOI:10.1016/j.ijhydene.2023.09.059.
11. Polikarpova Ilze & Lauka, Dace & Blumberga, Dagnija & Vigants, Edgars. Multi-criteria analysis to select renewable energy solution for district heating system. Environmental and climate technologies, 2019, pp.101-109, DOI:10.2478/rtuect-2019-0082.
12. Mazzeo D., Matera N., Pierangelo De L., Baglivo C. et al. Giuseppe Worldwide geographical mapping and optimization of stand-alone and grid-connected hybrid renewable system techno-economic performance across Kappen-Geiger climates. Applied energy, 2020, 276, 115507, DOI:10.1016/j.apenergy.2020.115507.
13. Шакиров В.А. Оценка экономической целесообразности подключения удаленных потребителей восточной Арктики к энергосистеме / В.А. Шакиров, И.Ю. Иванова, Р.А. Иванов // Арктика: экология и экономика, 2022. – Т. 12. – № 1. – С. 19-33. – DOI: 10.25283/2223-4594-2022-1-19-33.
14. Assari A., Mahesh T.M., Assari E. Role of public participation in sustainability of historical city: usage of TOPSIS method. Indian journal of science and technology, 2012, vol. 5, iss. 3, pp. 2289-2294.
15. Погодаева И.Н. Многокритериальный выбор состава оборудования гибридного энергокомплекса для удаленных районов / И.Н. Погодаева // ЭНЕРГОСТАРТ. Материалы VI Международной молодежной научно-практической конференции, 17-22 ноября 2023 г., Кузбас. гос. техн. ун-т им. Т. Ф. Горбачева. – Кемерово, 2023.
16. Маленкова И.Н. Многокритериальный анализ вариантов строительства электростанций методом PROMETHEE / И.Н. Маленкова, В.А. Шакиров // Труды Братского государственного университета. Серия: Естественные и инженерные науки, 2016. – № 1. – С. 41-50.

Погодаева Инесса Николаевна. *Аспирант, Братский государственный университет, ORCID: 0000-0003-3176-2422, m.inessa13@yandex.ru, 665709, г. Братск, ул. Макаренко 40, БрГУ.*

Шакиров Владислав Альбертович. *Заведующий кафедрой «Энергетика и электротехника» Иркутского национального исследовательского технического университета, ORCID: 0000-0001-8629-9549, shakirovva@ex.istu.edu.*

UDC 621.311
DOI:10.25729/ESI.2024.34.2.010

Selection of components of a hybrid energy complex for remote areas using multi-criteria assessment methods

Inessa N. Pogodaeva[1], Vladislav A. Shakirov[2]

[1]Bratsk State University, Russia, Bratsk, *m.inessa13@yandex.ru*
[2]Irkutsk National Research Technical University, Russia, Irkutsk

Abstract. The problem of energy supply to areas remote from the power grid remains relevant and is solved mainly in the modernization of diesel power plants and the construction of hybrid energy complexes with renewable energy sources. The life cycle of hybrid energy complexes is several decades, so the design must take into account all the main goals - economic and technical efficiency, minimizing environmental impacts, reliability. The multi-criteria nature of the problem requires the involvement of a decision maker to express preferences regarding the importance of performance indicators of various options for the energy complex. The stochastic nature of renewable generation and the variety of operating restrictions lead to the need to use special software for simulating operating conditions with hourly resolution. The article discusses an approach to multi-criteria selection of a hybrid energy complex from a variety of alternatives generated in the HOMER PRO program, using three methods: TOPSIS and PROMETHEE I, II. The methods have differences in the procedures for evaluating alternatives and thereby increase the validity of the choice. A numerical example is considered for the Ust-Soboleka. Hybrid energy complexes are formed using diesel generation technologies, wind power plants, photovoltaic converters, micro-hydroelectric power plants and energy storage devices. The study examines three scenarios for the development of the region, which entails differences in assessments of the importance of the criteria.

Keywords: hybrid energy complex, multi-criteria analysis, power plant, modeling, renewable energy sources, TOPSIS, PROMETHEE

References

1. World Energy Outlook 2023, IEA. Paris, available at: https://www.iea.org/reports/world-energy-outlook-2023, License: CC BY 4.0 (report), CC BY NC SA 4.0 (Appendix A)
2. Yezhekvartal'nyy informatsionnyy obzor rynkaVIE v Rossii IV kvartal 2023 goda (oktyabr' - dekabr' 2023g.) [Quarterly information review of the renewable energy market in Russia, IV quarter of 2023 (October - December 2023)]. Available at: https://rreda.ru/products/quarterly-reviews/review-1552
3. Pranav M.S., Karunanithi K., Akhil M. et al. Hybrid renewable energy sources (HRES). International conference on intelligent computing, instrumentation and control technologies (ICICICT), Kerala State, Kannur, India, IEEE, 2017, pp. 162–165, DOI:10.1109/ICICICT1.2017.8342553.
4. Bouaouda A, Sayouti Y. Hybrid meta-heuristic algorithms for optimal sizing of hybrid renewable energy system: a review of the state-of-the-art. Archives of computational methods in engineering, 2022, vol. 29(6), pp. 4049-4083, DOI: 10.1007/s11831-022-09730-x.
5. Diab Ah., Hamdy S., Mohamed S.Ih., Kuznetsov O.N. et al. Application of different optimization algorithms for optimal sizing of pv/wind/diesel/battery storage stand-alone hybrid microgrid. IEEE access, 2019, 99 p., DOI:10.1109/ACCESS.2019.2936656.
6. Agapitidou A.A., Skroufouta S., Baltas E. Methodology for the development of hybrid renewable energy systems (HRES) with pumped storage and hydrogen production on Lemnos Island. Earth 3, 2022, pp. 537-556, DOI:10.3390/earth3020032.
7. Vijayaragavan M., Darly S. Power flow management of grid connected HRES system using a CFANN technique. International journal of electronics, 2021, DOI:10.1080/00207217.2020.1831080.
8. Sharma K.K., Gupta A., Kumar R. et al. Economic evaluation of a hybrid renewable energy system (HRES) using hybrid optimization model for electric renewable (HOMER) software—a case study of rural India. International journal of low-carbon technologies, 2021, vol. 16, iss. 3, pp. 814-821, DOI:10.1093/ijlct/ctab012.
9. Bakhtiari H., Naghizadeh R.A. Multi-criteria optimal sizing of hybrid renewable energy systems including wind, photovoltaic, battery, and hydrogen storage with ε-constraint method. IET Renewable power generation, 2018, 12(8), pp 883–892, DOI:10.1049/iet-rpg.2017.0706
10. Li Z., Wang Y., Xie J. et al. Hybrid multi-criteria decision-making evaluation of multiple renewable energy systems considering the hysteresis band principle. International journal of hydrogen energy, 2024, vol. 49(Part D), pp. 450-462, DOI:10.1016/j.ijhydene.2023.09.059.

11. Polikarpova Ilze & Lauka, Dace & Blumberga, Dagnija & Vigants, Edgars. Multi-criteria analysis to select renewable energy solution for district heating system. Environmental and climate technologies, 2019, pp.101-109, DOI:10.2478/rtuect-2019-0082.
12. Mazzeo D., Matera N., Pierangelo De L., Baglivo C. et al. Giuseppe Worldwide geographical mapping and optimization of stand-alone and grid-connected hybrid renewable system techno-economic performance across Kappen-Geiger climates. Applied energy, 2020, 276, 115507, DOI:10.1016/j.apenergy.2020.115507.
13. Shakirov V. A., Ivanova I. Yu., Ivanov R. A. Otsenka ekonomicheskoy tselesoobraznosti podklyucheniya udalennykh potrebiteley vostochnoy Arktiki k energosisteme [Assessment of the economic feasibility of connecting remote consumers of the Eastern Arctic to the energy system]. Arktika, ekologiya i ekonomika [Arctic, ecology and economics], 2022, vol. 12, no. 1, pp. 19-33, DOI: 10.25283/2223-4594-2022-1-19-33.
14. Assari A., Mahesh T.M., Assari E. Role of public participation in sustainability of historical city: usage of TOPSIS method. Indian journal of science and technology, 2012, vol. 5, iss. 3, pp. 2289-2294.
15. Pogodaeva I.N. Mnogokriterial'nyy vybor sostava oborudovaniya gibridnogo energokompleksa dlya udalennykh rayonov [Multi-criteria selection of the equipment composition of a hybrid energy complex for remote areas]. ENERGOSTART. Materialy VI Mezhdunarodnoy molodezhnoy nauchno-prakticheskoy konferentsii, 17-22 noyabrya 2023 g., Kuzbas. gos. tekhn. un-t im. T. F. Gorbacheva, Kemerovo, 2023, [ENERGOSTART. Materials of the VI international youth scientific and practical conference, November 17-22, 2023, Kuzbass. state tech. University named after T. F. Gorbachev, Kemerovo, 2023.]
16. Malenkova I.N., Shakirov V.A. Mnogokriterial'nyy analiz variantov stroitel'stva elektrostantsiy metodom PROMETHEE [Multicriteria analysis of options for constructing power plants using the PROMETHEE method]. Trudy Bratskogo gosudarstvennogo universiteta. Seriya Yestestvennyye i inzhenernyye nauki [Proceedings of Bratsk State University. Series Natural and engineering sciences], 2016, no. 1, pp. 41-50.

Pogodaeva Inessa Nikolaevna. *Postgraduate student, Bratsk state university, ORCID: 0000-0003-3176-2422, m.inessa13@yandex.ru, 665709, Irkutsk region, Bratsk, st. Makarenko 40, BrGU.*

Shakirov Vladislav Albertovich. *Head of the department of Power supply and electrical engineering, Irkutsk national research technical university, ORCID: 0000-0001-8629-9549, shakirovva@ex.istu.edu.*

Статья поступила в редакцию 20.03.2024; одобрена после рецензирования 12.04.2024; принята к публикации 06.06.2024.

The article was submitted 03/20/2024; approved after reviewing 04/12/2024; accepted for publication 06/06/2024.

УДК 621.311.1
DOI:10.25729/ESI.2024.34.2.011

Моделирование процессов плавки гололеда в тяговых сетях постоянного тока

Крюков Андрей Васильевич[1,2], Черепанов Александр Валерьевич[1], Нгуен Куок Хиеу[2]

[1]Иркутский государственный университет путей сообщения,
Россия, Иркутск, *smart_grid@mail.ru*

[2]Иркутский национальный исследовательский технический университет, Россия, Иркутск

Аннотация. Цель представленных в статье исследований состояла в разработке цифровых моделей для определения режимов плавки гололеда на проводах контактных подвесок железных дорог постоянного тока. Они были реализованы в программном комплексе Fazonord, версия 5.3.4.9 –2024. Алгоритм расчета включал следующие этапы: расчет серии режимов, определяемых графиком движения поездов, на основе подхода, использующего фазные координаты; формирование зависимостей токов, протекающих по проводам, от времени; вычисление температур их нагрева; моделирование процесса плавки гололедной канавки; учет испарения пленки воды, оставшейся после отпадения ледяной муфты; моделирование нагрева токоведущих частей без корки льда при наличии или отсутствии мороси и дождя. Описаны исходные данные, необходимые для проведения расчетов. Они включают следующие группы параметров: геометрические, теплотехнические и электрические. Представлены компьютерные модели, учитывающие основные факторы процессов нагрева и удаления гололедных отложений на проводах контактных подвесок тяговых сетей (ТС) постоянного тока. Моделируемая система электроснабжения включала следующие элементы: три питающих ЛЭП 110 кВ и такое же число тяговых подстанций; два участка ТС 3 кВ протяженностью по 20 км. Показано, что гололед удаляется с несущих тросов за 17 минут, а с контактных проводов за 22 минуты. Температуры нагрева токоведущих частей и наиболее нагретых точек трансформаторов во время плавки не превышают допустимых величин. Методика является универсальной и может применяться для ТС любого конструктивного исполнения, включая перспективные ТС повышенного напряжения.

Ключевые слова: удаление гололедных отложений, железные дороги постоянного тока, системы электроснабжения

Цитирование: Крюков А.В. Моделирование процессов плавки гололеда в тяговых сетях постоянного тока / А.В. Крюков, А.В. Черепанов, К.Х. Нгуен // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 112-122. – DOI: 10.25729/ESI.2024.34.2.011.

Введение. В осенне-зимний период в ряде регионов России на проводах линий электропередачи (ЛЭП) и контактных подвесок тяговых сетей (ТС) возникают отложения гололеда, изморози и мокрого снега. Толщина такой ледяной корки может достигать нескольких десятков миллиметров и приводить к следующим негативным последствиям [1]:

- разрегулировке и сближению токоведущих частей;
- возникновению коротких замыканий, повреждениям линейной арматуры и креплений;
- механической перегрузке и обрыву проводов.

Вопросы, связанные с моделированием процессов плавки гололеда на воздушных ЛЭП и ТС, рассматриваются в большом числе работ. Условия образования льда и изморози на ЛЭП и технические решения по их ликвидации описаны в [1, 2]. Современные средства для раннего обнаружения ледяных образований представлены в [3]. Онлайн-метод расчета плавки гололеда на контактной сети описан в [4]. Исследования способа защиты ЛЭП от обледенения, учитывающего снижение мощности плавки, представлены в [5]. Результаты испытаний методов обработки контактных подвесок высокоскоростных линий для борьбы с гололедом рассмотрены в [6]. Анализ методов удаления льда с воздушных ЛЭП низкого напряжения проведен в [7]. Динамика перемещения проводов при сбросе гололедно-изморозевых отложений рассмотрена в [8]. Система оперативного мониторинга обледенения воздушной ЛЭП предложена в [9]. Проект нового противо-гололедного устройства для ЛЭП представлен в [10]. Модель раннего предупреждения о гололеде на ЛЭП, реализованная на

основе нейронной сети, предложена в [11]. Методика расчета значимости ЛЭП по критерию рисков в условиях ледяной погоды описана в [12]. Высокочастотные диэлектрические свойства линий электропередачи при наличии гололедных отложений рассмотрены в [13]. Новый подход к борьбе с обледенением расщепленных фаз ЛЭП сверхвысокого напряжения описан в [14]. Обзор способов борьбы с гололедом на проводах ЛЭП представлен в [15]. Метод определения вероятности обрыва проводов ЛЭП при ледяной катастрофе разработан в [16]. Новый подход к защите от обледенения линий кольцевой сети предложен в [17]. Практика борьбы с гололедом на линиях распределительных сетей описана в [18]. Результаты исследования методов плавления льда на ЛЭП среднего напряжения представлены в [19]. Система мониторинга обледенения ЛЭП предложена в [20].

При эксплуатации ТС постоянного тока применяются тепловые и механические способы освобождения проводов контактных подвесок от накопившегося на них льда. Наиболее широкое распространение получили методы плавки гололеда повышенными токами. В условиях цифровизации для выбора рациональных схем теплового удаления гололедных образований следует использовать компьютерные технологии, позволяющие адекватно моделировать сложные процессы нагрева и плавки льда на проводах ЛЭП и ТС.

Ниже представлены результаты исследований, направленных на создание цифровых моделей систем электроснабжения железных дорог постоянного тока (СЭЖД ПТ) в режимах плавки гололеда [21]. Методы моделирования СЭЖД ПТ описаны в [22].

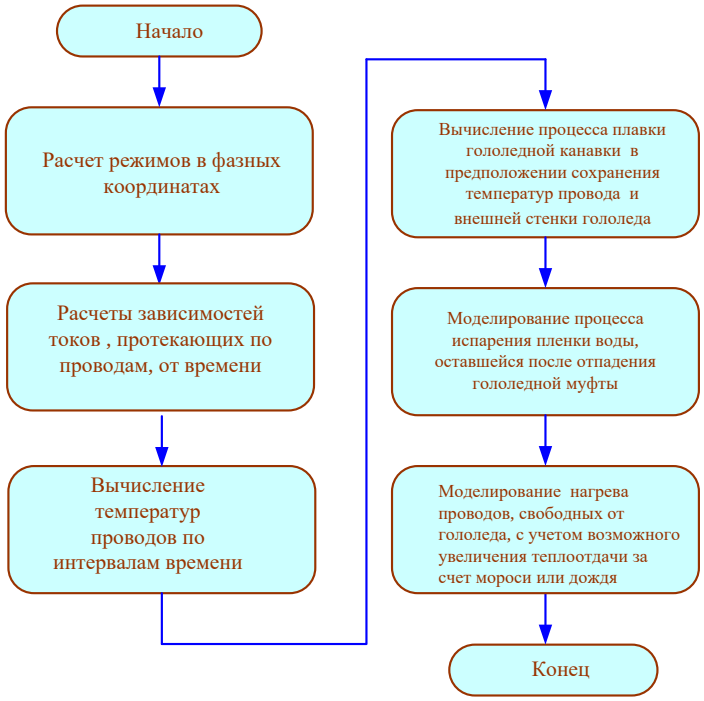


Рис. 1. Основные этапы алгоритма

Методика и результаты моделирования. Алгоритм моделирования процессов плавки гололеда (рис. 1), подробно описанный в [21] и реализованный в программном комплексе Fazonord, версия 5.3.4.9 – 2024 (рис. 2), включает следующие этапы:

1. Определение серии режимов с дискретностью $\delta t_k = t_{k+1} - t_k$.
2. Расчеты зависимостей токов I_{ij}, протекающих по проводам, от времени τ: $I_{ij} = I_{ij}(t)$, i, j -1…n, где n- число узлов сети.
3. Вычисление температур проводов по интервалам времени δt_k .
4. Расчет показателей процесса плавки гололедной канавки.
5. Учет испарения пленки воды, оставшейся после отпадения ледяной муфты.

6. Моделирование нагрева проводов, свободных от гололеда, с учетом возможного увеличения теплоотдачи за счет мороси или дождя при $\tau_{ОКР} < 0$ °С.

– Для моделирования сложных электрических и тепловых процессов плавки гололеда использовались следующие группы исходных данных (рис. 3):
– общие параметры для многопроводной линии (рис. 4);
– данные для конкретного провода с возможностью их выбора из базы с корректировкой отдельных значений (рис. 5);
– информация по стенке гололеда (рис. 6).

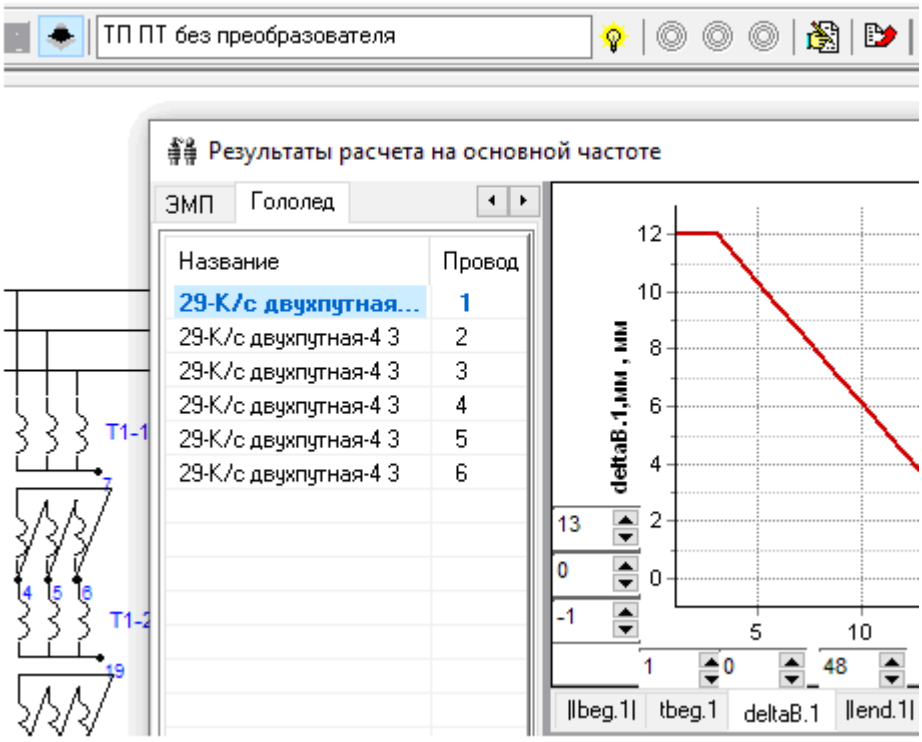


Рис. 2. Окно ПК с результатами моделирования процесса плавки гололеда (б)

Рис. 3. Группы исходных данных

Рис. 4. Параметры многопроводной линии

Рис. 5. Параметры провода

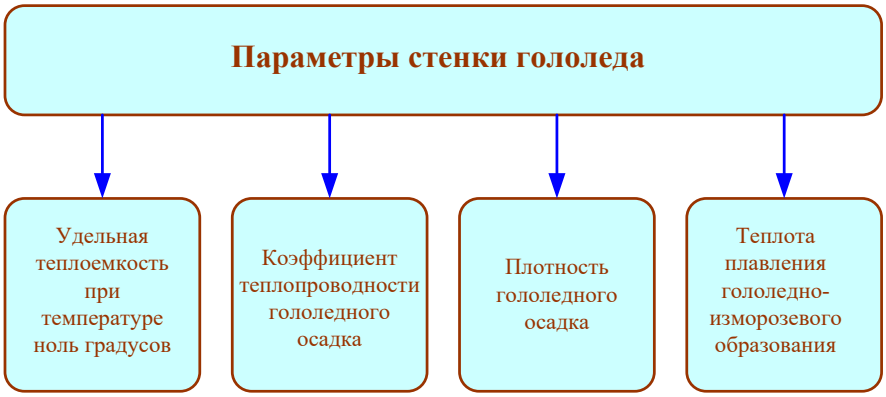
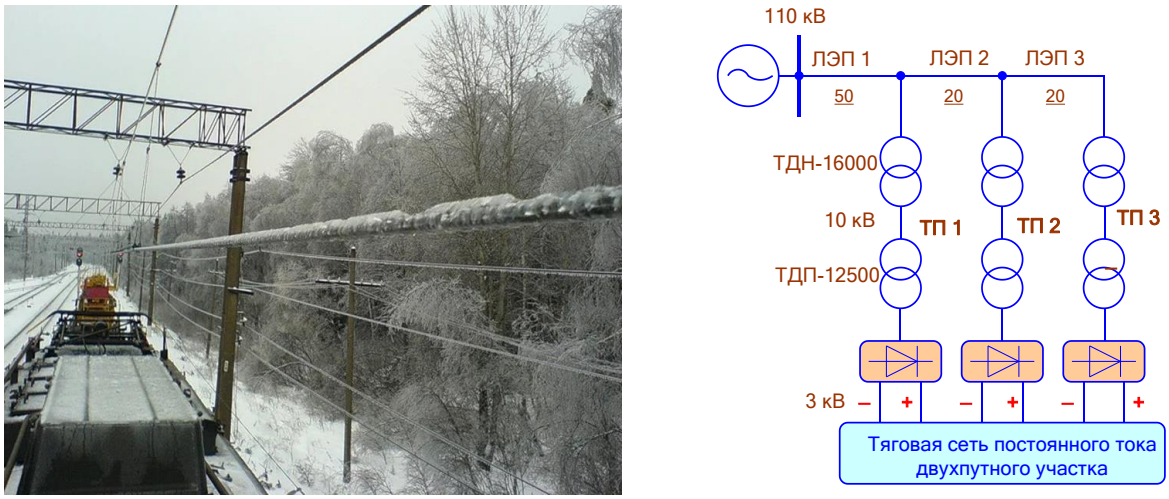


Рис. 6. Параметры стенки гололеда

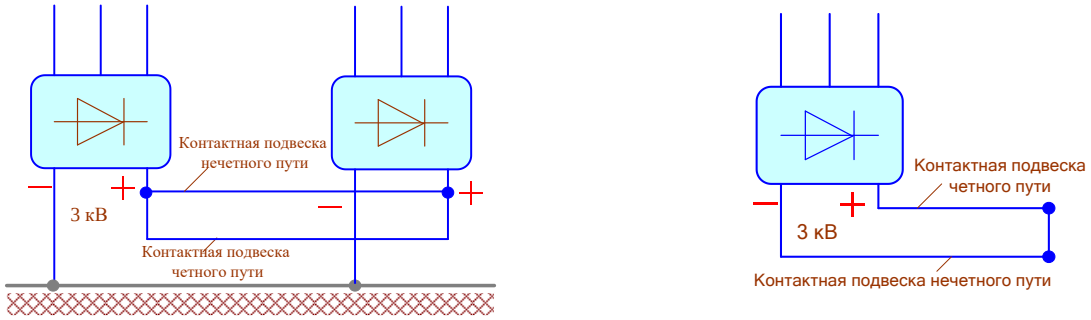
Расчеты проводились для СЭЖД ПТ, схема которой показана на рис. 7 (б). Плавка гололеда (рис. 7 (а)) выполнялась на левой межподстанционной зоне (МПЗ) путем создания петлевой схемы, приведенной на рис. 8. На правой МПЗ (между ТП 2 и ТП 3) предусматривалось движение поездов.

Часть визуального представления разработанной модели показана на рис. 9. Расположение токоведущих частей контактной сети (КС) с четырьмя проводами МФ-100 и двумя тросами ПБСМ-120 представлено на рис. 10. Моделируемая СЭЖД включала три тяговых подстанции и две МПЗ, протяженностью по 20 км. График движения поездов с интервалом в тридцать минут приведен на рис. 11.



а) б)

Рис. 7. Гололед на проводах ТС (а) и схема моделируемой сети (б)



а) б)

Рис. 8. Схемы тяговой сети; а – исходная; б – плавки гололеда

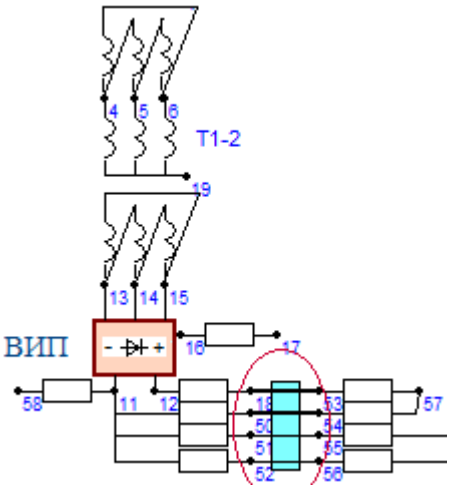


Рис. 9. Фрагмент визуального отображения расчетной модели, отвечающей участку с плавкой гололеда

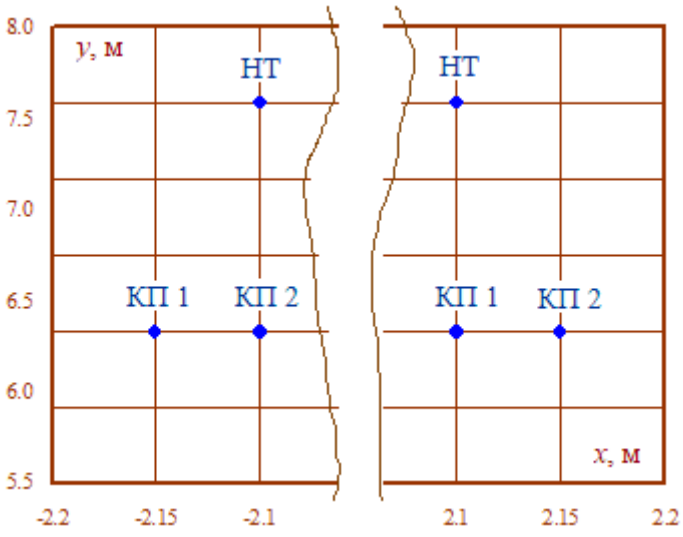
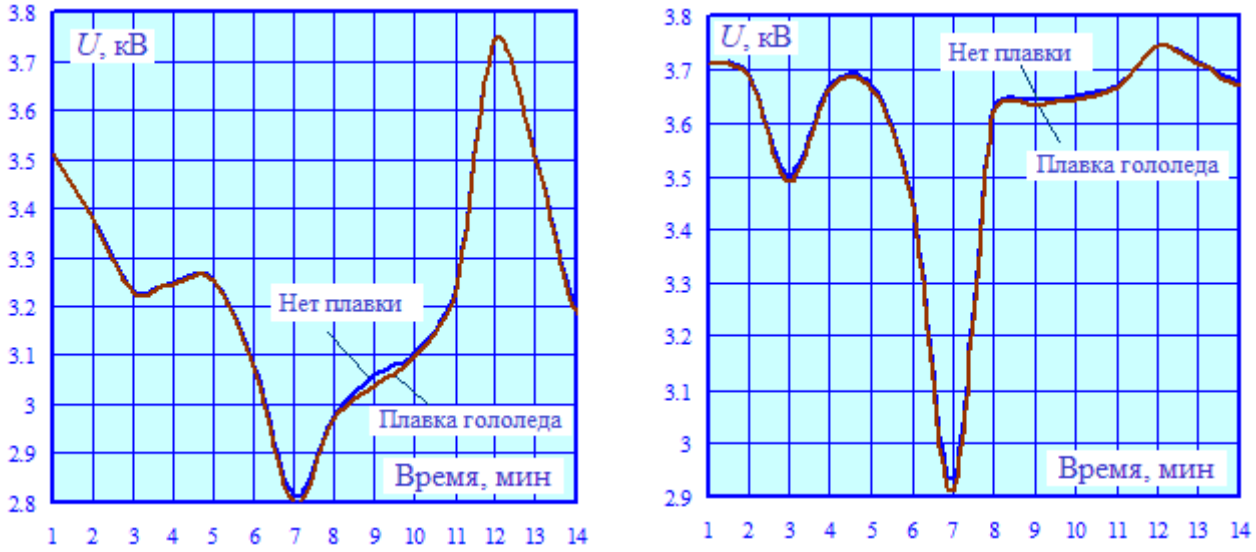


Рис. 10. Расположение токоведущих частей
НТ – несущий трос; КП – контактный провод

Рис. 11. График движения ЭПС

Результаты моделирования представлены на рис. 12–17. На рис. 12 проиллюстрирована динамика изменений напряжений на пантографах электровозов в двух режимах: исходном по рис. 8 (а); при плавке льда на проводах (рис. 8 (б)). Из представленных на этих рисунках зависимостей $U = U(t)$ можно сделать вывод о том, что режим устранения гололеда с толщиной стенки в 12 мм на первой МПЗ практически не оказывает влияние на процессы перевозок во второй зоне. Это подтверждает возможность применения петлевой схемы плавки.



а) б)

Рис. 12. Напряжения на пантографах электровозов:
а – нечетный поезд № 1; б – четный поезд № 1

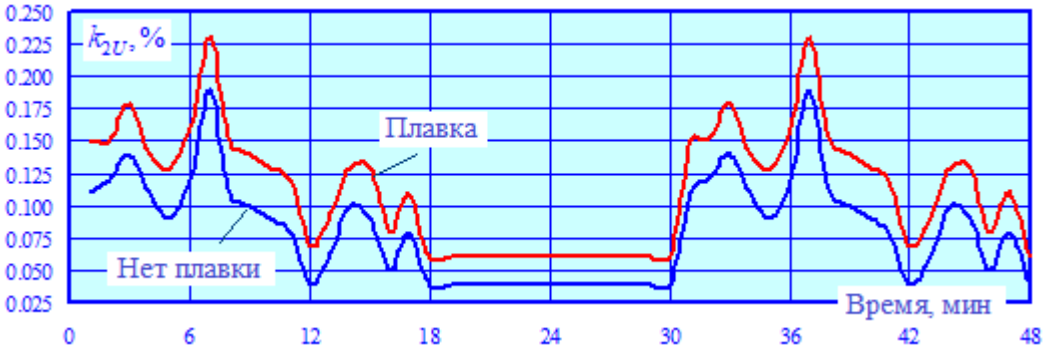


Рис. 13. Вариации коэффициента несимметрии на шинах 110 кВ ТП 3

На рис. 13 показаны графики, характеризующие коэффициенты несимметрии на шинах 110 кВ ТП. Из них видно, что, в отличие от систем тягового электроснабжения переменного тока, в СЭЖД ПТ проблема повышенных уровней k_{2U} отсутствует. В режимах плавки

гололеда наблюдается незначительный рост максимальных и средних значений этого параметра (рис. 14).

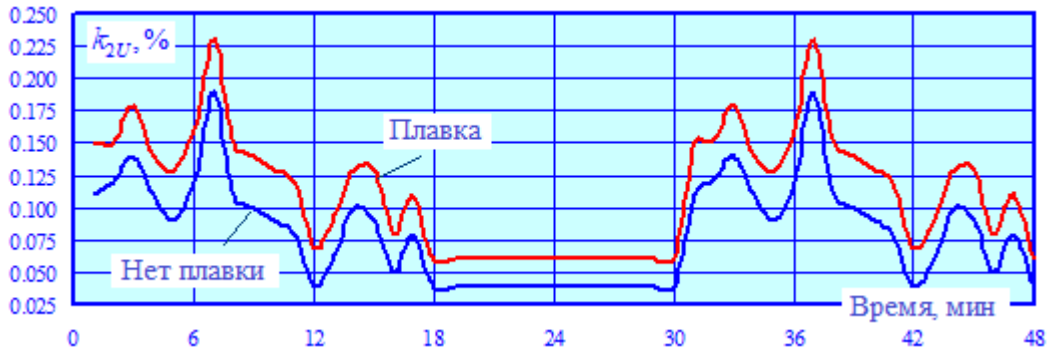
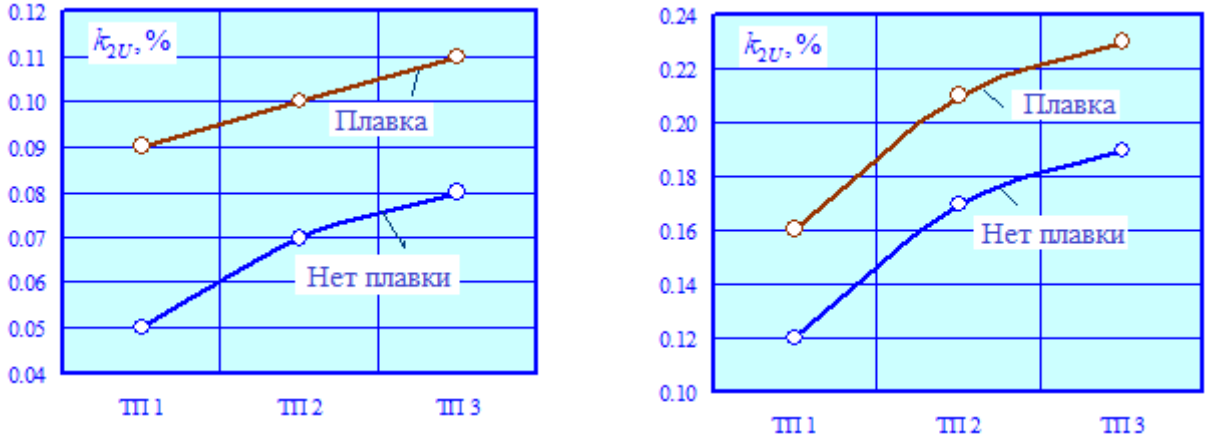


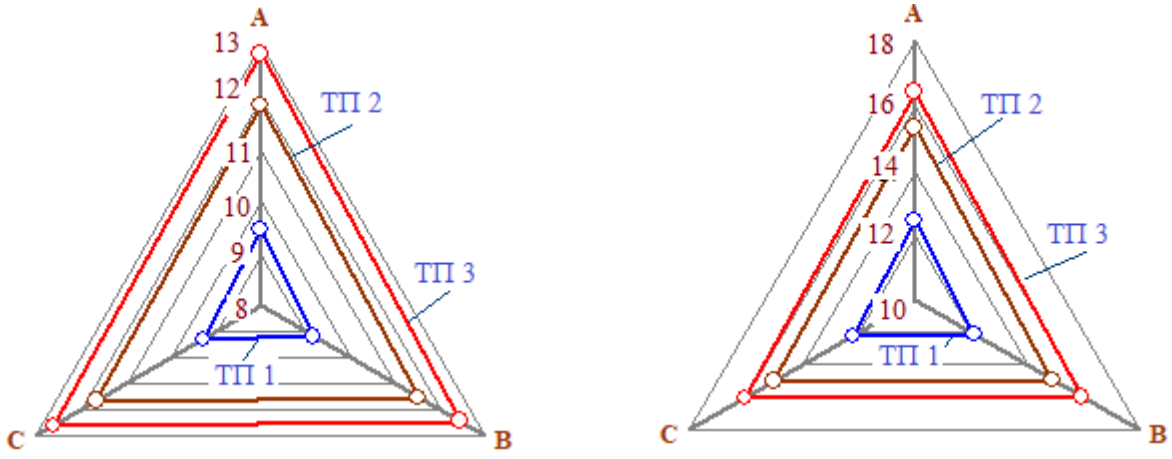
Рис. 14. Вариации коэффициента несимметрии на шинах 110 кВ ТП 3



а) б)

Рис. 15. Коэффициенты несимметрии на вводах 110 кВ тяговых трансформаторов

Преобразовательные агрегаты тяговых подстанций СЭЖД ПТ создают повышенные уровни гармонических искажений на шинах 110 кВ ТП (рис. 16).



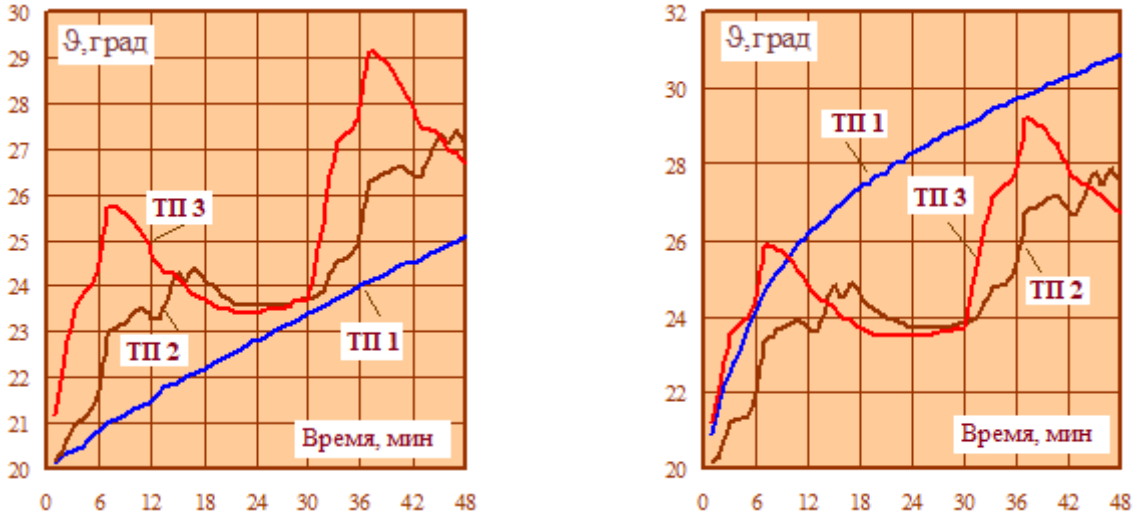
а) б)

Рис. 16. Коэффициенты гармоник на вводах 110 кВ тяговых трансформаторов

В режиме плавки гололеда максимальные значения k_U возрастают на 22…23 %. Для снижения гармонических искажений можно эффективно использовать активные фильтры [23], которые не создают опасности возникновения резонансных процессов.

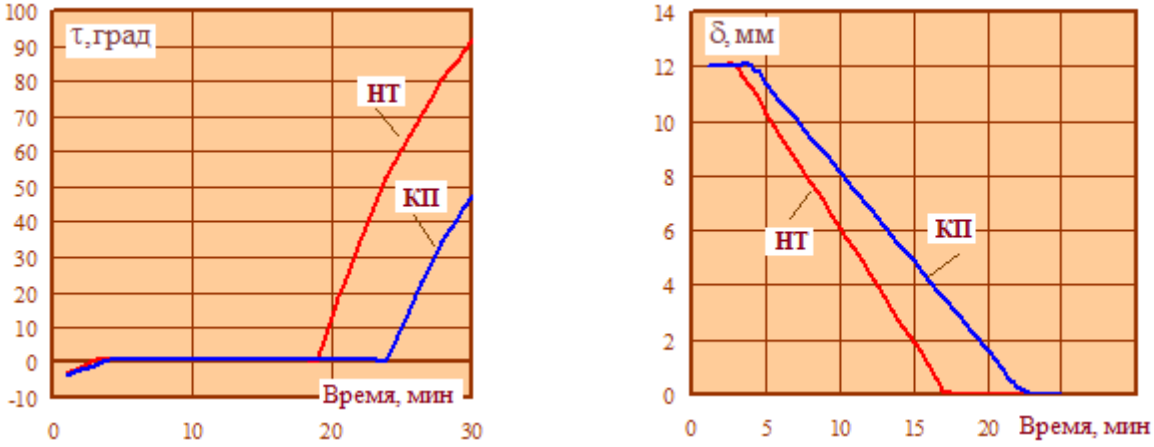
На рис. 17 показаны графики, отвечающие температурам наиболее нагретых точек (ТННТ) трансформаторов ТДН-16000-115/11. Из них видно, что при плавке гололеда максимальное увеличение ТННТ по сравнению с исходным режимом равно 6 ºС. Величины

ТННТ не превышают допустимого значения в 98 ºС. Это свидетельствует о применимости рассматриваемой схемы плавки гололеда.



а) б)

Рис. 17. Температуры наиболее нагретых точек трансформаторов ТДН-16000-115/11: а – исходный режим; б – режим плавки гололеда



а) б)

Рис. 18. Динамика процесса плавки гололеда: а – температуры провода; б – толщина стенки гололеда

На рис. 18 приведены результаты расчёта процессов плавки в виде зависимостей:

$$\tau = \tau(t);\ \delta = \delta(t),$$

где τ – температура, град; δ – толщина стенки гололедных отложений, мм.

Представленные на рис. 18 зависимости позволяют сделать следующие выводы:

- гололедные образования удаляются с несущих тросов за 17 минут, а с контактных проводов за 22 минуты;
- температуры токоведущих частей во время плавки не превышают допустимых величин;
- после освобождения ото льда начинается интенсивный рост τ, поэтому для исключения перегрева проводов плавку надо завершать через 30 минут.

Заключение. Разработаны модели, обеспечивающие определение режимов плавки гололеда в тяговых сетях постоянного тока. Методика их создания является универсальной и может применяться для ТС любого конструктивного исполнения. Алгоритм расчета процессов правки включал следующие этапы: расчет серии режимов, определяемых графиком движения поездов, на основе подхода, использующего фазные координаты; формирование зависимостей токов, протекающих по проводам, от времени; вычисление

температур их нагрева; расчет показателей процесса плавки гололедной канавки; учет испарения пленки воды, оставшейся после отпадения ледяной муфты; моделирование нагрева проводов, свободных от гололеда, при наличии или отсутствии мороси или дождя.

В результате моделирования подтверждена эффективность применения петлевой схемы подогрева токоведущих частей. Установлено, что гололедные образования удаляются с несущих тросов за 17 минут, а с контактных проводов за 22 минуты. Их температуры во время плавки не превышают допустимых величин. Аналогичный вывод можно сделать и для наиболее нагретых точек трансформаторов ТДН-16000-115/11, что свидетельствует об отсутствии перегрузок.

Список источников

1. Дьяков, А.Ф. Предотвращение и ликвидация гололедных аварий в электрических сетях энергосистем / А. Ф. Дьяков, А. С. Засыпкин, И. И. Левченко. – Пятигорск: РП "Южэнерготехнадзор", 2000. – 284 с. – EDN YASQMX.
2. Мясоедов Ю.В. Плавка гололеда на линиях электропередачи: методы, инновации, эксплуатация / Ю.В. Мясоедов, О.Н. Коржова // Вестник Амурского государственного университета. Серия: Естественные и экономические науки. – № 87. –2019. – С. 65-66. – EDN LNYWCF.
3. Кольцов А.В. Новые решения в системах плавки гололеда на воздушных линиях электропередачи высокого напряжения 220 и 330 КВ / А.В. Кольцов, Г.Н. Ковтун, Ю.А. Горюшин и др. // Электротехнические и информационные комплексы и системы, 2016. – Т. 12. – № 4. – С. 21-27. – EDN YTXTLD.
4. Yunda W., Gang Z., Zhongbei T. et al. An online thermal deicing method for urban rail transit catenary. IEEE Transactions on transportation electrification, 2021, vol: 7, Iss. 2.
5. Qunwei Jing, Yusheng Zhou, Liang Liu Research on excitation deicing for transmission line with considering the deicing power attenuation. 2015 5th International conference on electric utility deregulation and restructuring and power technologies (DRPT), 2015, DOI:10.1109/DRPT.2015.7432506.
6. Enrico C., Alvaro F. Deicing of the contact lines of the high-speed electric railways: deicing configurations. experimental test results. IEEE Transactions on power delivery, 2014, vol. 29, iss. 6.
7. Kunpeng Ji, Lichun Zhang, Xueping Zhan, Bin Liu, Jialun Yang Analysis of mechanical deicing method for low voltage overhead transmission lines. 2020 IEEE 5th Information technology and mechatronics engineering conference (ITOEC).
8. Luo Yong, Deng Shaoping, Xu Zhengtao, Huang Xianxu. Analysis of deicing jump characteristics of double circuit overhead transmission lines. IEEE 2nd International conference on electrical engineering, Big Data and algorithms (EEBDA), 2023
9. Shiguang Nie, Guangyu Qu, Haifeng Ye, Xiuchen Jiang On-line monitoring system for icing state of overhead transmission line. IEEE, Power engineering and automation conference, 2012.
10. Sarbajit Paul, Sobia Bashir, Junghwan Chang. Design of a novel electromagnetic energy harvester with dual core for deicing device of transmission lines. IEEE Transactions on magnetics, 2019, vol. 55, iss. 2.
11. Mingguan Zhao, Jianlin Hu, Meng Li et al. Early warning model for transmission line deicing jumps based on bp neural network. IEEE Sustainable power and energy conference (iSPEC), 2023.
12. Yushen Hou, Xiuli Wang, Zechen Wu, Zece Zhang Risk-based transmission lines importance calculation under ice storm weather. 5th International conference on electric utility deregulation and restructuring and power technologies (DRPT), 2015.
13. Xiaohong Ma, Zhou Dai, Yusheng Zhou et al. High frequency dielectric property of transmission lines ice. International conference on computer systems, electronics and control (ICCSEC), 2017.
14. Chang Guanghui, Su Sheng, Li Mingming, Chao Daifeng Novel deicing approach of overhead bundled conductors of EHV transmission systems. IEEE Transactions on power delivery, 2009, vol. 24, iss. 3.
15. Sullivan C.R., Petrenko V.F., McCurdy J.D., Kozliouk V. Breaking the ice [transmission line icing]. IEEE Industry applications magazine, 2003, vol. 9, iss. 5.
16. Dong Wang, Min Qi, Qun Zhang et al. Prediction of breaking probability of transmission line galloping under ice storm disaster. IEEE 3rd Conference on energy internet and energy system integration (EI2), 2019,
17. Xiao-Long Luo, Xiang-Jun Zeng, Min Zeng et al. A novel protection method for icing lines of ring network. IEEE PES Innovative smart grid technologies, 2012.
18. Lyuzerui Yuan, Wei Huang, Xinhao Lin et al. Exploration and practice of guarding against icing on distribution network lines in southern region. 7th International conference on smart grid and smart cities (ICSGSC), 2023.
19. Yuan Zhu, Yanjun Tan, Qingjun Huang et al. Research on melting and de-icing methods of lines in distribution Network. IEEE 3rd Conference on energy Internet and energy system integration (EI2), 2019.
20. Jiankun Zhao, Baofeng Yan, Jianli Zhao, Bo Chen. Icing monitoring system of transmission lines based on image and stress. 2nd IEEE Conference on energy Internet and energy system integration (EI2), 2018.

21. Закарюкин, В.П. Моделирование динамики плавки гололеда на проводах воздушных линий электропередачи и контактных подвесок тяговых сетей / В.П. Закарюкин, А.В. Крюков // System analysis and mathematical modeling, 2022. – Т. 4. – № 1. – С. 11-37. – DOI 10.17150/2713-1734.2022.4(1).11-37. – EDN HLXMBJ.
22. Закарюкин В.П. Моделирование систем тягового электроснабжения постоянного тока на основе фазных координат / В.П. Закарюкин, А.В. Крюков. – Москва: Директ-Медиа, 2023. – 156 с.
23. Крюков А.В. Уменьшение гармонических искажений в электрических сетях, питающих тяговые подстанции железных дорог постоянного тока / А.В. Крюков, К.В. Суслов, А.В. Черепанов, К.Х. Нгуен // Промышленная энергетика, 2024. – № 4. – С. 51-57.

Крюков Андрей Васильевич. *Д.т.н., профессор кафедры Электроэнергетика транспорта, Иркутский государственный университет путей сообщения, AuthorID: 1172600, SPIN: 8463-3200, ORCID: 0000-0001-6543-1790, and_kryukov@mail.ru.*

Черепанов Александр Валерьевич. *К.т.н., доцент кафедры Электроэнергетика транспорта, Иркутский государственный университет путей сообщения, AuthorID: 850588, SPIN: 4280-2137, ORCID: 0000-0002-7712-9537, smart_grid@mail.ru.*

Нгуен Куок Хиеу. *Аспирант, Иркутский национальный исследовательский технический университет, hieu12829@mail.ru.*

UDC 621.311.1
DOI:10.25729/ESI.2024.34.2.011

Modeling modes of ice melting in DC traction networks

Andrey V. Kryukov[1,2], Aleksandr V. Cherepanov[1], Nguyen Quoc Hieu[2]

[1]Irkutsk State Transport University, Russia Irkutsk, *smart_grid@mail.ru*
[2]Irkutsk National Research Technical University, Russia, Irkutsk

Abstract. The purpose of the research presented in this article was to develop digital models for determining the melting modes of ice on the overhead wires of DC railways. The models were implemented in the Fazonord software package, version 5.3.4.9 –2024. The calculation algorithm included the following stages: determination of a series of modes determined by the train schedule, based on an approach using phase coordinates; calculations of the dependences of currents flowing through wires on time; calculation of their heating temperatures over time intervals; calculation of indicators of the process of melting ice-covered grooves; taking into account the evaporation of the water film remaining after the ice sleeve falls off; modeling of heating of wires free from ice, taking into account a possible increase in heat transfer due to drizzle or rain. The initial data necessary for the calculations are described. They include the following groups of parameters: geometric, thermal and electrical. Computer models are presented that take into account the main factors of the processes of heating and removing icing deposits on overhead wires of DC traction networks. The simulated traction power supply system included the following elements: three 110 kV supply power lines; three traction substations; two sections of 3 kV traction network with a length of 20 km. It has been shown that ice is removed from support cables in 17 minutes, and from contact wires in 22 minutes. The heating temperatures of current-carrying parts during melting do not exceed permissible values. The temperatures of the hottest points of traction transformers were determined. It has been shown that melting does not cause unacceptable overheating of transformers. The technique is universal and can be used for traction networks of any design.

Keywords: deicing, DC railways, power supply systems

References

1. Dyakov, A.F., Zasypkin A.S., Levchenko I.I. Predotvrashcheniye i likvidatsiya gololednykh avariy v elektricheskikh setyakh energosistem [Prevention and elimination of ice accidents in electrical networks of power systems]. Pyatigorsk: RP “Yuzhenergotekhnadzor”, 2000, 284 p.
2. Myasoedov, Yu.V. Korzhova O.N. Plavka gololeda na liniyakh elektroperedachi: metody, innovatsii, ekspluatatsiya [Melting ice on power lines: methods, innovations, operation] Vestnik Amurskogo gosudarstvennogo universiteta. Seriya: Yestestvennyye i ekonomicheskiye nauki [Bulletin of the Amur State University. Series: Natural and economic sciences], 2019, no. 87, pp. 65-66.
3. Koltsov A.V., Kovtun G.N., Goryushin Yu.A. et al. Novyye resheniya v sistemakh plavki gololeda na vozdushnykh liniyakh elektroperedachi vysokogo napryazheniya 220 i 330 KV [New solutions in systems for melting ice on overhead power lines of high voltage 220 and 330 kV]. Elektrotekhnicheskiye i informatsionnyye kompleksy i sistemy [Electrical engineering and information complexes and systems], 2016, vol. 12, no. 4, pp. 21-27.

4. Yunda W., Gang Z., Zhongbei T. et al. An online thermal deicing method for urban rail transit catenary. IEEE Transactions on transportation electrification, 2021, vol: 7, Iss. 2.
5. Qunwei Jing, Yusheng Zhou, Liang Liu Research on excitation deicing for transmission line with considering the deicing power attenuation. 2015 5th International conference on electric utility deregulation and restructuring and power technologies (DRPT), 2015, DOI:10.1109/DRPT.2015.7432506.
6. Enrico C., Alvaro F. Deicing of the contact lines of the high-speed electric railways: deicing configurations. experimental test results. IEEE Transactions on power delivery, 2014, vol. 29, iss. 6.
7. Kunpeng Ji, Lichun Zhang, Xueping Zhan, Bin Liu, Jialun Yang Analysis of mechanical deicing method for low voltage overhead transmission lines. 2020 IEEE 5th Information technology and mechatronics engineering conference (ITOEC).
8. Luo Yong, Deng Shaoping, Xu Zhengtao, Huang Xianxu. Analysis of deicing jump characteristics of double circuit overhead transmission lines. IEEE 2nd International conference on electrical engineering, Big Data and algorithms (EEBDA), 2023
9. Shiguang Nie, Guangyu Qu, Haifeng Ye, Xiuchen Jiang On-line monitoring system for icing state of overhead transmission line. IEEE, Power engineering and automation conference, 2012.
10. Sarbajit Paul, Sobia Bashir, Junghwan Chang. Design of a novel electromagnetic energy harvester with dual core for deicing device of transmission lines. IEEE Transactions on magnetics, 2019, vol. 55, iss. 2.
11. Mingguan Zhao, Jianlin Hu, Meng Li et al. Early warning model for transmission line deicing jumps based on bp neural network. IEEE Sustainable power and energy conference (iSPEC), 2023.
12. Yushen Hou, Xiuli Wang, Zechen Wu, Zece Zhang Risk-based transmission lines importance calculation under ice storm weather. 5th International conference on electric utility deregulation and restructuring and power technologies (DRPT), 2015.
13. Xiaohong Ma, Zhou Dai, Yusheng Zhou et al. High frequency dielectric property of transmission lines ice. International conference on computer systems, electronics and control (ICCSEC), 2017.
14. Chang Guanghui, Su Sheng, Li Mingming, Chao Daifeng Novel deicing approach of overhead bundled conductors of EHV transmission systems. IEEE Transactions on power delivery, 2009, vol. 24, iss. 3.
15. Sullivan C.R., Petrenko V.F., McCurdy J.D., Kozliouk V. Breaking the ice [transmission line icing]. IEEE Industry applications magazine, 2003, vol. 9, iss. 5.
16. Dong Wang, Min Qi, Qun Zhang et al. Prediction of breaking probability of transmission line galloping under ice storm disaster. IEEE 3rd Conference on energy internet and energy system integration (EI2), 2019,
17. Xiao-Long Luo, Xiang-Jun Zeng, Min Zeng et al. A novel protection method for icing lines of ring network. IEEE PES Innovative smart grid technologies, 2012.
18. Lyuzerui Yuan, Wei Huang, Xinhao Lin et al. Exploration and practice of guarding against icing on distribution network lines in southern region. 7th International conference on smart grid and smart cities (ICSGSC), 2023.
19. Yuan Zhu, Yanjun Tan, Qingjun Huang et al. Research on melting and de-icing methods of lines in distribution Network. IEEE 3rd Conference on energy Internet and energy system integration (EI2), 2019.
20. Jiankun Zhao, Baofeng Yan, Jianli Zhao, Bo Chen. Icing monitoring system of transmission lines based on image and stress. 2nd IEEE Conference on energy Internet and energy system integration (EI2), 2018.
21. Zakaryukin, V.P., Kryukov A.V. Modelirovaniye dinamiki plavki gololeda na provodakh vozdushnykh liniy elektroperedachi i kontaktnykh podvesok tyagovykh setey [Modeling the dynamics of melting ice on the wires of overhead power lines and catenary suspensions of traction networks]. System analysis and mathematical modeling, 2022, vol. 4, no.. 1, pp. 11-37, DOI 10.17150/2713-1734.2022.4(1).11-37.
22. Zakaryukin V.P., Kryukov A.V. Modelirovaniye sistem tyagovogo elektrosnabzheniya postoyannogo toka na osnove faznykh koordinat [Modeling of DC traction power supply systems based on phase coordinates]. Moscow, Direct-Media, 2023, 156 p.
23. Kryukov A.V., Suslov K.V., Cherepanov A.V., Nguyen Q.H. Umen'sheniye garmonicheskikh iskazheniy v elektricheskikh setyakh, pitayushchikh tyagovyye podstantsii zheleznykh dorog postoyannogo toka [Reducing harmonic distortions in electrical networks feeding traction substations of DC railways]. Promyshlennaya energetika [Industrial energy], 2024, no. 4, pp. 51-57.

Kryukov Andrei Vasilevich. *Doctor of engineering science, professor, Irkutsk state transport university, AuthorID: 1172600, SPIN: 8463-3200, ORCID: 0000-0001-6543-1790, and_kryukov@mail.ru.*

Cherepanov Aleksandr Valerevich. *Ph.D. of engineering science, associate professor, Irkutsk state transport university, AuthorID: 850588, SPIN: 4280-2137, ORCID: 0000-0002-7712-9537, smart_grid@mail.ru.*

Nguyen Quoc Hieu. *Postgraduate student, Irkutsk national research technical university, hieu12829@mail.ru.*

Статья поступила в редакцию 28.05.2024; одобрена после рецензирования 04.06.2024; принята к публикации 06.06.2024.

The article was submitted 05/28/2024; approved after reviewing 06/04/2024; accepted for publication 06/06/2024.

Цифровая экономика и управление

УДК 332.02:004.02
DOI:10.25729/ESI.2024.34.2.012

Поддержка принятия решений в территориальном управлении на основе метода интегрального оценивания качества жизни (на примере Красноярского края)

Алдошкина Анна Михайловна, Пенькова Татьяна Геннадьевна, Ноженкова Людмила Федоровна

Институт вычислительного моделирования СО РАН,
Россия, Красноярск, *metus@ism.krasn.ru*

Аннотация. В работе представлен подход к поддержке принятия решений в территориальном управлении на основе метода интегрального оценивания качества жизни. Предложена модель формирования управляющих рекомендаций на основе анализа текущего уровня качества жизни. Представлена платформа для мониторинга качества жизни в муниципальных образованиях региона и поддержки формирования рейтинговых оценок для измерения качества жизни в разрезе реализации национальных проектов. Представлены результаты оценивания качества жизни в муниципальных образованиях Красноярского края по данным за 2021 год.

Ключевые слова: поддержка принятия решений, качество жизни, интегральное оценивание, управляющие рекомендации

Цитирование: Алдошкина А.М. Поддержка принятия решений в территориальном управлении на основе метода интегрального оценивания качества жизни (на примере Красноярского края) / А.М. Алдошкина, Т.Г. Пенькова, Л.Ф. Ноженкова // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 123-134. – DOI: 10.25729/ESI.2024.34.2.012.

Введение. Повышение качества жизни является основным показателем эффективности государственной политики и важнейшей стратегической задачей территориального управления всех уровней [1, 2]. В современных условиях управления, направленного на «повышение сбалансированности пространственного развития экономики и социальной сферы» и «создание опережающего социально-экономического прогресса», необходимым инструментом принятии решений становится аналитическое обобщение информации о процессах, происходящих в различных сферах жизнедеятельности территорий [3].

Решением проблемы оценки качества жизни современные исследователи занимаются последние 40 лет. Качество жизни оценивается как в разрезе отдельных населенных пунктов, так и стран в целом. Специалисты Института государственной политики AARP США регулярно выпускают отчеты с индексами оценки качества жизни [4]. Применяются различные подходы к измерению индекса, включая опросы и тематические исследования с использованием статистических методов и различных онлайн баз данных. Так, одной группой авторов используются методы стохастического анализа для составления рейтинга стран на основе единого составного индекса качества жизни [5]. Другой группой авторов применяется метод опорных векторов для анализа многомерных данных и составления индекса качества жизни [6]. Для сравнительного анализа уровня устойчивого развития Европейского Союза применяется метод многомерного шкалирования [7]. При этом в разных странах оценка качества жизни производится с привязкой к местной специфике, климатическим и геодезическим условиям. Например, исследователи из Японии составляют индекс качества жизни с учетом риска землетрясения в городах и округах страны [8]. Много исследований посвящено анализу качества жизни в связи с тем или иным заболеванием [9-11]. Отечественными авторами также проводятся исследования и расчеты на основе интегрального индекса качества жизни. С помощью кластерного и факторного анализа значений индекса и его частных показателей делаются выводы об особенностях качества жизни в различных

странах мира. Отдельно рассматривается ситуация в группе стран БРИКС и формулируются рекомендации по повышению качества жизни [12]. В последнее время особенно востребована информация о качестве жизни населения с детализацией на уровне региона и муниципальных образований для сглаживания дифференциации между территориями субъекта. Например, ведутся работы над созданием системы мониторинга для оценки качества жизни населения на уровне муниципальных образований в Якутии [13]. Здесь разработана система показателей с учетом особенностей северных территорий и утвержденных индикаторов реализации национальных проектов в соответствии с Указом Президента России № 204. Предложена методика расчета интегрального показателя и выполнена оценка качества жизни муниципальных образований Республики Саха. При этом основной акцент делается на показатели, отражающие доход и занятость населения [14-15].

С целью обеспечения контроля текущей ситуации и возможности выявлять проблемные направления экономики и социальной сферы территорий, авторами работы был предложен метод расчета рейтинговых оценок для измерения качества жизни муниципальных образований в разрезе реализации национальных проектов. Метод обеспечивает формирование комплексного показателя на основе иерархии оценок показателей социально-экономического развития территорий с применением территориально-ориентированной нормативной модели. Данный метод позволяет получать обобщенные количественные оценки текущего уровня жизни с учетом индивидуальных особенностей территорий, выполнять сравнительный анализ территорий и, в случае необходимости, детализировать оценки до конкретных показателей, что дает возможность определять первопричины сложившейся ситуации и формировать управляющие рекомендации.

Целью данного исследования является разработка инструмента, обеспечивающего поддержку принятия решений в территориальном управлении по результатам рейтингового оценивания качества жизни, а именно разработка модели формирования управляющих рекомендаций на основе детализации интегральных оценок.

1. Интегральное оценивание качества жизни муниципальных образований. Оценивание качества жизни территорий основано на формировании территориально-ориентированной нормативной модели, которая описывает иерархическую структуру комплексного показателя качества жизни и задает параметры, позволяющие учитывать индивидуальные особенности территории для корректного измерения фактического уровня качества жизни [16]. Нормативная модель разрабатывается на основе спецификаций территорий (паспортов территорий, общероссийских и региональных классификаторов) и особенностей их социально-экономического развития с привлечением экспертов.

Формирование нормативной модели состоит из следующих этапов: 1) формирование иерархии показателей; 2) определение коэффициентов значимости показателей; 3) определение нормативных значений базовых показателей; 4) определение коэффициентов чувствительности оценок; 5) формирование оценочной шкалы. На первом этапе строится иерархическая система показателей. Нижний уровень иерархии представляет набор базовых статистических социально-экономических показателей. Остальные уровни представляют комплексные показатели. В качестве комплексных показателей на промежуточных уровнях иерархии рассматриваются показатели, характеризующие направления национальных проектов – «Человеческий капитал», «Комфортная среда», «Экономический рост», каждый из этих комплексных показателей обобщает показатели, характеризующие соответствующие национальные проекты. Структура иерархии показателей может меняться в зависимости от целей и задач управления. На следующем этапе для каждой территории с учетом её физико-географических и социально-экономических особенностей оценивается относительный вклад показателя в показатель верхнего уровня и определяется коэффициент значимости. Далее, по

результатам статистического анализа для каждого базового показателя рассчитывается интервал нормативных значений, определяются коэффициенты чувствительности оценок. Коэффициент чувствительности определяет, насколько резко отреагирует оценка на отклонение фактического значения показателя от установленного норматива. Далее для интерпретации оценок формируется оценочная шкала, для которой задается лингвистическая переменная «Уровень качества жизни» с терм-множеством значений: «Улучшенный», «Хороший», «Приемлемый», «Удовлетворительный», «Пониженный», «Низкий», «Критический». Формирование нормативной модели завершается экспертизой, в ходе которой выполняется проверка «норматива» на соответствие актуальным данным с необходимой корректировкой.

Оценивание качества жизни территорий представляет собой пошаговый расчет оценки комплексного показателя – «интегральной оценки». Интегральная оценка формируется путем обобщения оценок вверх по дереву иерархии показателей в соответствии с нормативной моделью и определенными алгоритмами [17]. Интегральное оценивание состоит в последовательном выполнении процедур: формирование оценок базовых показателей, формирование оценок комплексных показателей и интерпретация оценок. Расчет оценок базовых показателей выполняется путем сопоставления фактических значений показателей с интервалами их нормативных значений. Выход текущего значения показателя за границу интервала нормативных значений влияет положительно или отрицательно на оценку состояния территории в зависимости от желаемого тренда показателя. Расчет оценок комплексных показателей выполняется с учетом коэффициентов значимости обобщаемых показателей. Интерпретация оценок выполняется путем преобразования их количественных значений в эквивалентные качественные значения с помощью оценочной шкалы.

2. Детализация интегральных оценок для выявления причин текущего качества жизни и формирования управляющих рекомендаций. Для выявления причин текущего уровня качества жизни и формирования управляющих рекомендаций предлагается модель, позволяющая анализировать сложившуюся ситуацию по результатам интегрального оценивания. Модель представляет собой когнитивную карту и строится на основе иерархии оценок, позволяя выделять «проблемы» и «причины» текущего состояния.

Согласно теории когнитивного анализа и моделирования [18, 19], под «ситуацией» понимается сочетание условий и обстоятельств, создающих определенную обстановку, в которой возникла «проблема». Под «проблемой» понимается уровень качества жизни ниже ожидаемого, возникший в результате определенных «причин». Для конкретного сочетания «проблем» и «причин» формируются управляющие рекомендации, направленные на достижение желаемого уровня качества жизни. Особенность предложенной модели состоит в применении иерархических зависимостей между «причинами» возникновения «проблемных» ситуаций и формированием комплекса рекомендаций с учетом детализации факторов, влияющих на качество жизни.

На рисунке 1 представлена обобщенная модель выявления «причин» текущего состояния и формирования управляющих рекомендаций. В ходе моделирования ситуации определяется показатель «проблема» P_i, характеризующий текущее состояние, где $i = \overline{1,I}$ – количество возникших «проблем». Для каждого показателя «проблема» определяется совокупность иерархически подчиненных показателей «причина» $S_{lj}(S_{l-1j}(\dots S_{1j}(P_i)))$, которые характеризуют возникшую «проблемную» ситуацию, где $l = \overline{1,L}$ – количество уровней иерархии «причин», $j = \overline{1,J}$ – количество показателей «причина» на уровне иерархии. В модели используются знаковые и беззнаковые типы связей: знаковые соответствуют причинно-следственной связи показателей, которые находятся в состоянии «не норма»,

беззнаковые обеспечивают перебор «причин» одного уровня или переход от «причин» к рекомендации. Состояние «не норма» определяется, исходя из полученных значений оценок показателей. Например, в качестве состояния «не норма» могут рассматриваться значения оценок «Пониженный», «Низкий» или «Критический», либо ухудшение оценки показателя по сравнению с предыдущим отчетным периодом. Если «проблема» P_i в состоянии «не норма», то на каждом уровне иерархии существует показатель «причина» в состоянии «не норма», что обеспечивает движение вниз по иерархии. Частные рекомендации $R_j(S_{Lj}(\dots S_{1j}(P_i)))$ формируются для «причин» $S_{Lj}(S_{L-1j}(\dots S_{1j}(P_i)))$, соответствующих базовым показателям на нижнем уровня иерархии, которые находятся в состоянии «не норма». Для каждого уровня «причин» формируются общие рекомендации $R_l^*(S_{lj}(\dots S_{1j}(P_i)))$. Для уровня «проблем» формируется общая рекомендация R^*. Управляющие рекомендации могут меняться в зависимости от значения оценки.

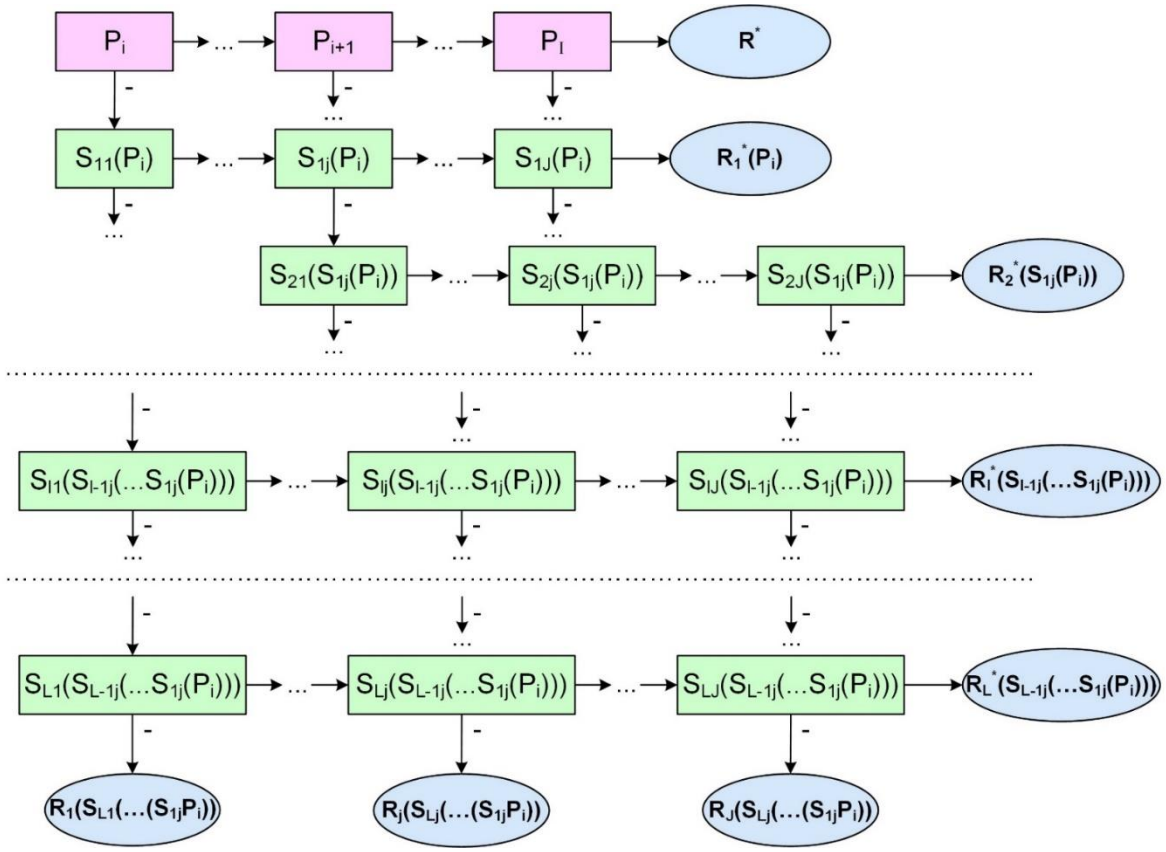


Рис. 1. Обобщенная модель выявления причин текущего состояния и формирования управляющих рекомендаций

Например, в соответствии с нормативной моделью качества жизни Красноярского края [16], модель будет иметь три уровня причин: уровень базовых показателей социально-экономического развития, уровень национальных проектов и уровень сфер качества жизни. Пусть определена «проблема» P_1: оценка комплексного показателя «Качество жизни» имеет значение «Пониженный». Согласно иерархии комплексных показателей, данная «проблема» связана со следующими «причинами»: «Человеческий капитал» $S_{11}(P_1)$, «Комфортная среда» $S_{12}(P_1)$ и «Экономический рост» $S_{13}(P_1)$. В свою очередь, каждый из указанных показателей «причина», содержит дочерние показатели «причина». Так, для показателя $S_{11}(P_1)$ определяются «причины»: $S_{21}(S_{11}(P_1))$ «Здравоохранение», $S_{22}(S_{11}(P_1))$ «Демография», $S_{23}(S_{11}(P_1))$ «Образование» и $S_{24}(S_{11}(P_1))$ «Культура». Для показателя $S_{12}(P_1)$ определяются «причины»: $S_{21}(S_{12}(P_1))$ «Безопасные и качественные дороги», $S_{22}(S_{12}(P_1))$ «Жилье и городская среда» и $S_{23}(S_{12}(P_1))$ «Экология». Для показателя $S_{13}(P_1)$ определяются

«причины»: $S_{21}(S_{13}(P_1))$ «Малое и среднее предпринимательство», $S_{22}(S_{13}(P_1))$ «Цифровая экономика», $S_{23}(S_{13}(P_1))$ «Производительность труда и поддержка занятости», $S_{24}(S_{13}(P_1))$ «Международная кооперация и экспорт» и $S_{25}(S_{13}(P_1))$ «Магистральная инфраструктура». И так далее, спускаясь вниз по иерархии до базовых показателей.

Частные управляющие рекомендации на уровне базовых показателей формируются, исходя из интервалов нормативных значений показателей, определенных в нормативной модели. Общие рекомендации на уровне базовых показателей формируются, исходя из задач и результатов, определенных в паспортах региональных проектов. Региональные проекты Красноярского края описывают конкретные меры, направленные на достижения национальных целей в Красноярском крае. Общие рекомендации на уровне национальных проектов формируются, исходя из паспортов федеральных проектов, входящих в состав соответствующего национального проекта, которые описывают системные меры, направленные на достижения национальных целей. Общие рекомендации на уровне сфер качества жизни формируются в соответствии с целями и целевыми показателями, определенными в паспортах национальных проектов. Таким образом, модель формирования рекомендаций представляет собой инструмент исследования интегральных оценок состояния территорий и обеспечивает поддержку принятия обоснованных решений.

3. Поддержка принятия решений в территориальном управлении на примере Красноярского края. Для поддержки принятия решений в территориальном управлении при реализации национальных проектов в Красноярском крае разработана аналитическая платформа [20]. Платформа предназначена для мониторинга уровня качества жизни и формирования рейтинговых оценок для измерения качества жизни в разрезе реализации национальных проектов в муниципальных образованиях региона.

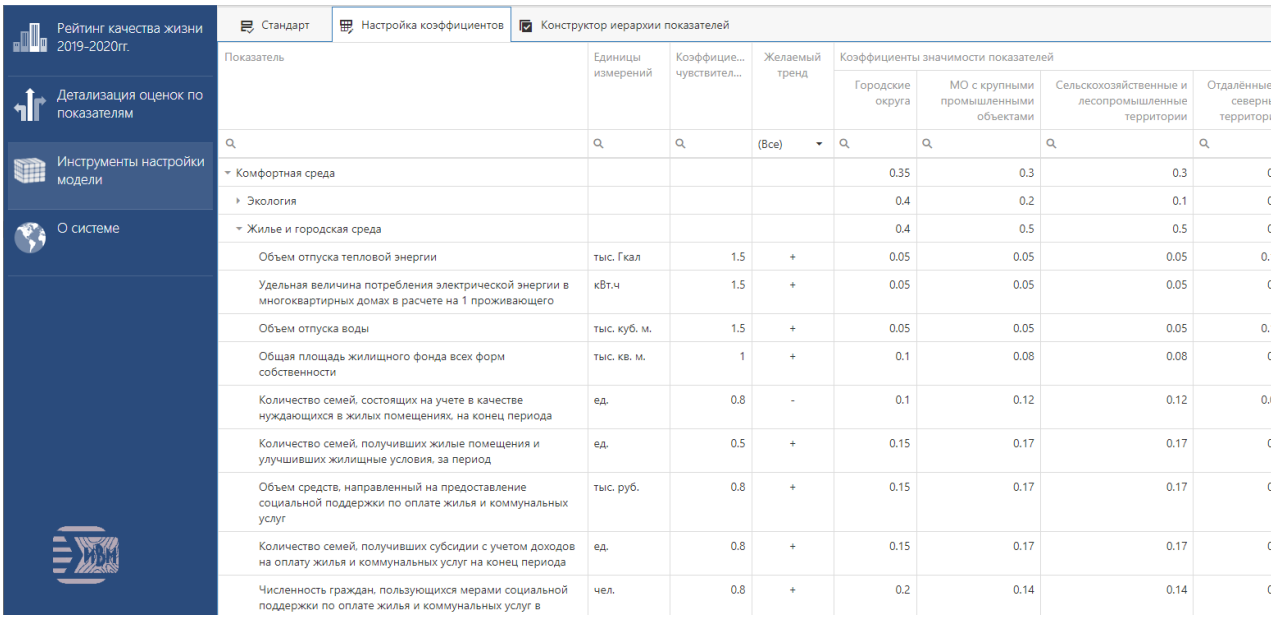
К основным задачам, решение которых поддерживает платформа для информационно-аналитической поддержки принятия решений, относятся:

- формирование нормативной модели на основе накопленных статистических данных о состоянии экономики и социальной сферы территорий, а также экспертных знаний о физико-географических характеристиках и особенностях социально-экономического развития территорий;
- расчет и качественная интерпретация оценок показателей в соответствии с алгоритмами, предусмотренными методом интегрального оценивания;
- визуальное представление полученных рейтинговых оценок с помощью диаграмм и картограмм;
- формирование управляющих рекомендаций по результатам расчета оценок показателей.

Аналитическая платформа представляет собой клиент-серверное приложение с web-интерфейсом. Создание и редактирование нормативной модели реализовано в табличной форме с поддержкой фильтрации, поиска, сортировки, постраничной разбивки больших данных, редактирования и валидации вводимых пользователем значений (рисунок 2).

Перечень базовых показателей качества жизни основан на данных многолетнего мониторинга муниципальных образований, который ведется службами государственной статистики в сфере экономики и социальной политики Красноярского края. Для каждого базового показателя определены коэффициенты чувствительности и желаемый тренд значений, положительный (+) или отрицательный (-). Иерархия показателей едина для всех территорий края, а весовые коэффициенты определены, исходя из типа территории. Территории Красноярского края разделены на четыре территориальных кластера: городские округа, муниципальные образования с крупными промышленными объектами, муниципальные образования с развитой сельскохозяйственной и лесопромышленной

отраслью, отдаленные (северные) муниципальные образования. Нормативные значения базовых показателей и оценочная шкала вычисляются расчетно-статистическим модулем на серверной стороне.



Показатель	Единицы измерений	Коэффицие... чувствител...	Желаемый тренд	Коэффициенты значимости показателей			
				Городские округа	МО с крупными промышленными объектами	Сельскохозяйственные и лесопромышленные территории	Отдалённые северны территори
			(Все)				
Комфортная среда				0.35	0.3	0.3	
Экология				0.4	0.2	0.1	
Жилье и городская среда				0.4	0.5	0.5	
Объем отпуска тепловой энергии	тыс. Гкал	1.5	+	0.05	0.05	0.05	0.
Удельная величина потребления электрической энергии в многоквартирных домах в расчете на 1 проживающего	кВт.ч	1.5	+	0.05	0.05	0.05	
Объем отпуска воды	тыс. куб. м.	1.5	+	0.05	0.05	0.05	0.
Общая площадь жилищного фонда всех форм собственности	тыс. кв. м.	1	+	0.1	0.08	0.08	
Количество семей, состоящих на учете в качестве нуждающихся в жилых помещениях, на конец периода	ед.	0.8	-	0.1	0.12	0.12	0.
Количество семей, получивших жилые помещения и улучшивших жилищные условия, за период	ед.	0.5	+	0.15	0.17	0.17	
Объем средств, направленный на предоставление социальной поддержки по оплате жилья и коммунальных услуг	тыс. руб.	0.8	+	0.15	0.17	0.17	
Количество семей, получивших субсидии с учетом доходов на оплату жилья и коммунальных услуг на конец периода	ед.	0.8	+	0.15	0.17	0.17	
Численность граждан, пользующихся мерами социальной поддержки по оплате жилья и коммунальных услуг в	чел.	0.8	+	0.2	0.14	0.14	

Рис. 2. Определение параметров нормативной модели

Представление оценок показателей реализовано в виде кросс-таблицы, которая имеет сложную вертикальную и горизонтальную шапку и обеспечивает возможность исследования, сортировки, детализации и агрегации оценок в разрезе территорий и показателей (рисунок 3). Количественные значения оценки отражают степень соответствия текущего состояния территорий нормативу. Значение оценки равно единице в случае нахождения всех дочерних базовых показателей в интервале нормативных значений, значения выше единицы указывают на улучшение качества жизни территорий, значения ниже единицы указывают на ухудшение состояния.

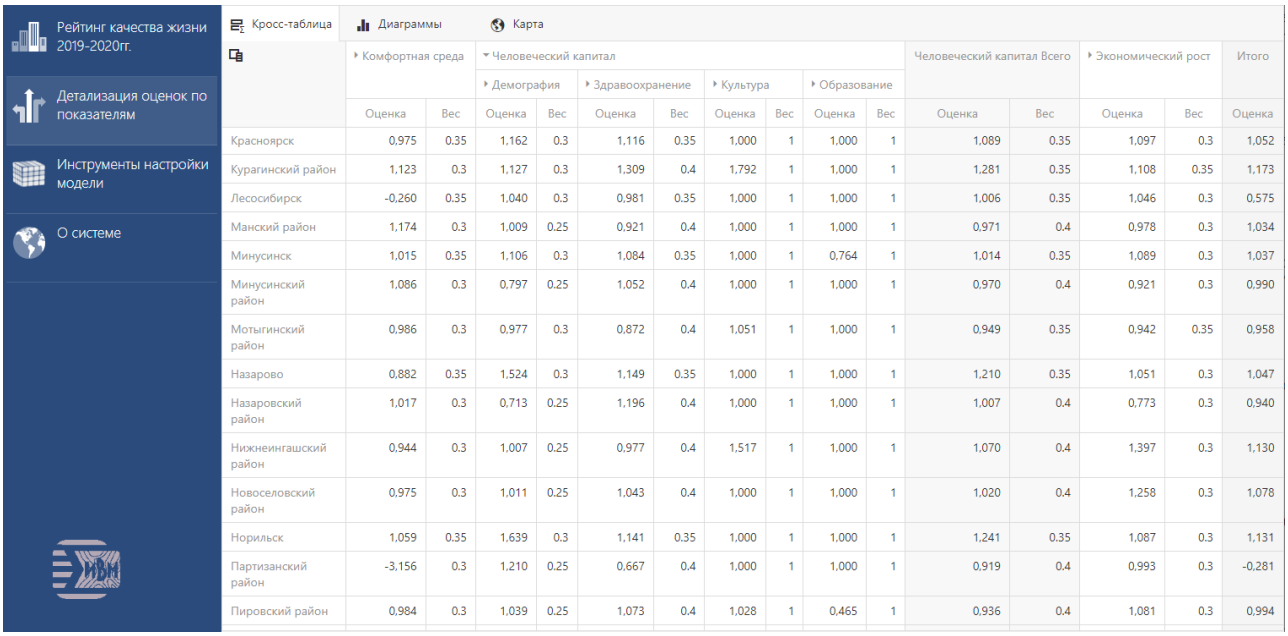
Сравнительный анализ отдельных территорий и исследование динамики показателей по годам выполняется при помощи гистограмм и линейных графиков (рисунок 4). Анализ территории выполняется в разрезе фактических значений показателей и рейтинговых оценок. Динамика показателя описывает изменение как его фактического значения в течение ряда лет, так и интервала его нормативных значений.

Пространственное распределение интегральных оценок с учетом их качественной интерпретации отображается с помощью картограммы (рисунок 5). Картограмма позволяет исследовать распределение оценок с географической привязкой на любом уровне иерархии показателей: в разрезе интегральной оценки качества жизни, в разрезе сфер качества жизни, в разрезе национальных проектов и в разрезе базовых показателей.

Результаты оценивания качества жизни муниципальных образований Красноярского края по данным за 2021 год показали, что большинство районов Красноярского края демонстрируют удовлетворительный и пониженный уровень качества жизни, 18 и 27 районов края соответственно.

Наихудший уровень качества жизни зафиксирован в Северо-Енисейском районе, интегральная оценка качества жизни составила 0,36, что соответствует критическому уровню. Такая ситуация определяется как «проблема». Согласно иерархии показателей, данная «проблема» связана со следующими «причинами»: «Человеческий капитал», «Комфортная среда» и «Экономический рост». В свою очередь, для показателя «Человеческий капитал» определены «причины»: «Здравоохранение» и «Демография». Спускаясь ниже на уровень

базовых показателей, для показателя «Здравоохранение» определены «причины»: «Обеспеченность врачами, на 10 тыс. человек населения», «Доля населения, охваченного профилактическими осмотрами» и «Число заболеваний, зарегистрированных у больных с впервые в жизни установленным диагнозом, на 1 тыс. человек населения». В зависимости от выявленных «причин» формируются управляющие рекомендации.



	Комфортная среда		Человеческий капитал								Человеческий капитал Всего		Экономический рост		Итого
			Демография		Здравоохранение		Культура		Образование						
	Оценка	Вес	Оценка	Вес	Оценка	Вес	Оценка	Вес	Оценка	Вес	Оценка	Вес	Оценка	Вес	Оценка
Красноярск	0,975	0.35	1,162	0.3	1,116	0.35	1,000	1	1,000	1	1,089	0.35	1,097	0.3	1,052
Курагинский район	1,123	0.3	1,127	0.3	1,309	0.4	1,792	1	1,000	1	1,281	0.35	1,108	0.35	1,173
Лесосибирск	-0,260	0.35	1,040	0.3	0,981	0.35	1,000	1	1,000	1	1,006	0.35	1,046	0.3	0,575
Манский район	1,174	0.3	1,009	0.25	0,921	0.4	1,000	1	1,000	1	0,971	0.4	0,978	0.3	1,034
Минусинск	1,015	0.35	1,106	0.3	1,084	0.35	1,000	1	0,764	1	1,014	0.35	1,089	0.3	1,037
Минусинский район	1,086	0.3	0,797	0.25	1,052	0.4	1,000	1	1,000	1	0,970	0.4	0,921	0.3	0,990
Мотыгинский район	0,986	0.3	0,977	0.3	0,872	0.4	1,051	1	1,000	1	0,949	0.35	0,942	0.35	0,958
Назарово	0,882	0.35	1,524	0.3	1,149	0.35	1,000	1	1,000	1	1,210	0.35	1,051	0.3	1,047
Назаровский район	1,017	0.3	0,713	0.25	1,196	0.4	1,000	1	1,000	1	1,007	0.4	0,773	0.3	0,940
Нижнеингашский район	0,944	0.3	1,007	0.25	0,977	0.4	1,517	1	1,000	1	1,070	0.4	1,397	0.3	1,130
Новоселовский район	0,975	0.3	1,011	0.25	1,043	0.4	1,000	1	1,000	1	1,020	0.4	1,258	0.3	1,078
Норильск	1,059	0.35	1,639	0.3	1,141	0.35	1,000	1	1,000	1	1,241	0.35	1,087	0.3	1,131
Партизанский район	-3,156	0.3	1,210	0.25	0,667	0.4	1,000	1	1,000	1	0,919	0.4	0,993	0.3	-0,281
Пировский район	0,984	0.3	1,039	0.25	1,073	0.4	1,028	1	0,465	1	0,936	0.4	1,081	0.3	0,994

Рис. 3. Представление оценок показателей

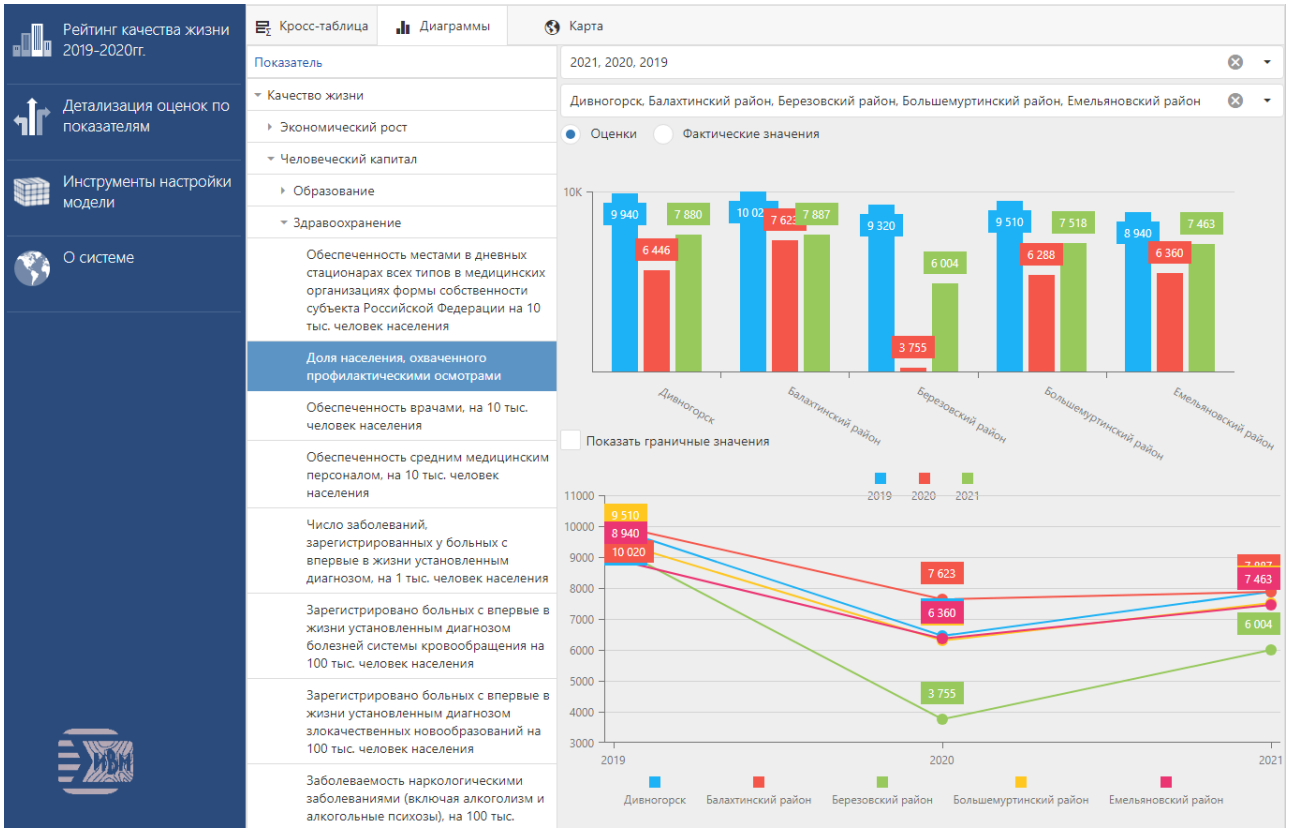


Рис. 4. Сравнительный анализ отдельных территорий

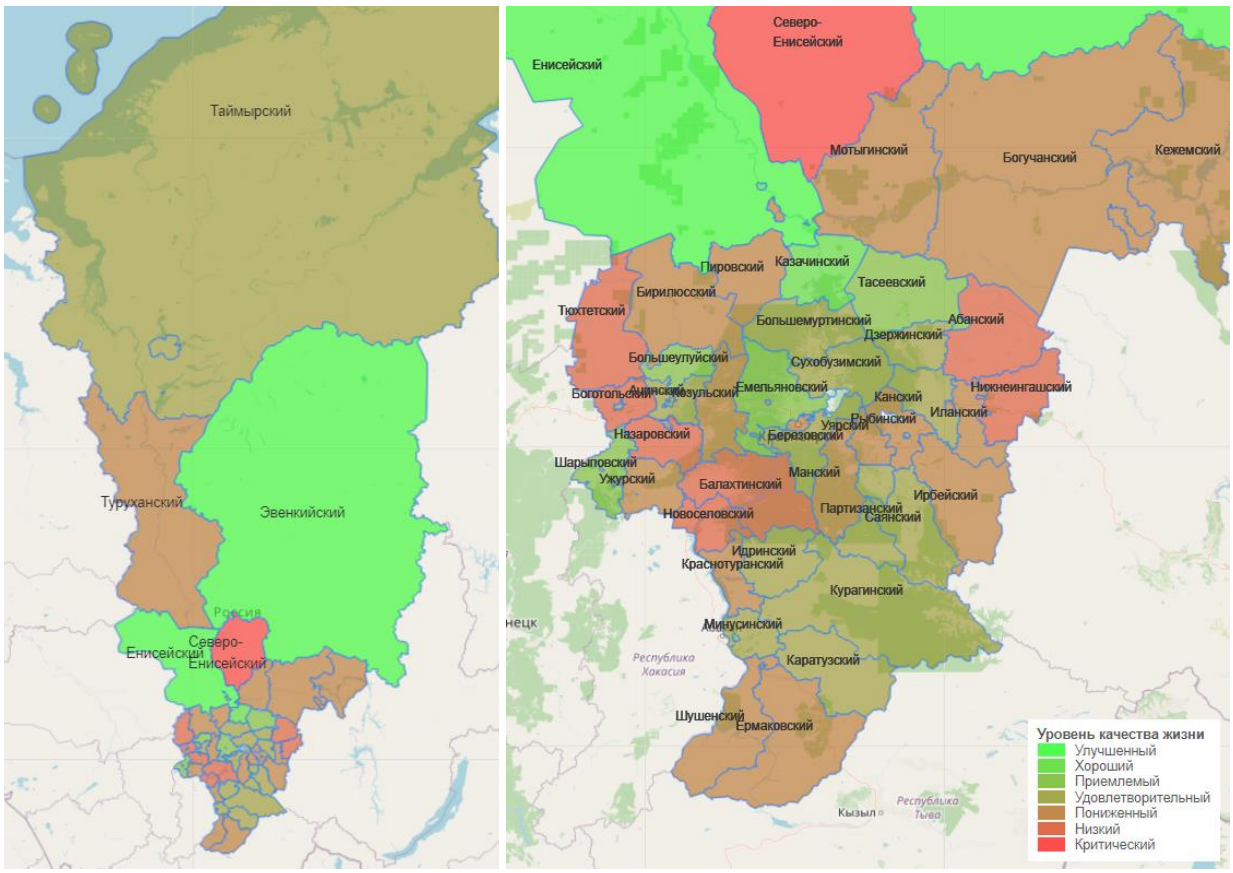


Рис. 5. Картограмма распределения интегральной оценки

Примерами формулировок частных рекомендаций с учетом разницы фактических и нормативных значений показателей в этом случае могут быть следующие: «Следует увеличить обеспеченность врачами, на 10 тыс. человек населения до не менее чем 6,85 чел.», «Следует обеспечить увеличение доли населения, охваченного профилактическими осмотрами, до не менее чем 1,58 %» и «Следует обеспечить снижение числа заболеваний, зарегистрированных у больных с впервые в жизни установленным диагнозом, на 1 тыс. человек населения до не менее чем 35,19 ед.» соответственно. Например, общая рекомендация для уровня базовых показателей «Здравоохранения», сформированная на основе паспортов региональных проектов для Красноярского края, может быть следующей: «Следует обеспечить увеличение укомплектованности фельдшерских пунктов, повысить обеспеченность населения врачами, оказывающими первичную медико-санитарную помощь, повысить обеспеченность медицинскими работниками, оказывающими скорую медицинскую помощь, повысить обеспеченность населения врачами, оказывающими специализированную медицинскую помощь, увеличить число специалистов, участвующих в системе непрерывного образования медицинских работников, в том числе с использованием дистанционных образовательных технологий, обеспечить цифровизацию здравоохранения и доступность гражданам цифровых сервисов, в том числе телемедицинских технологий, электронной записи к врачу, электронных рецептов». Аналогичным образом устанавливаются «причины» и формируются управляющие рекомендации для показателей «Комфортная среда» и «Экономический рост» для территории Северо-Енисейского района. Данные рекомендации выдают необходимую информацию о «причинах» текущего состояния и возможных мер для выработки экспертами конкретных мероприятий по улучшению сложившейся ситуации.

Заключение. В работе представлен инструмент, обеспечивающий информационно-аналитическую поддержку принятия решений в территориальном управлении на основе метода интегрального оценивания качества жизни. Метод обеспечивает формирование

рейтинговых оценок территорий с применением нормативной модели. Предложена модель, позволяющая выявлять «причины» текущего уровня качества жизни территорий и формировать комплекс управляющих рекомендаций в соответствии с иерархией получаемых оценок. Представлена программная реализация метода интегрального оценивания качества жизни для муниципальных образований Красноярского края.

Возможность интегрального оценивания качества жизни в разрезе реализации национальных проектов на уровне муниципальных образований и формирования управляющих рекомендаций на основе иерархически зависимых «причин» текущей ситуации определяют новизну предложенных решений. Развитие метода интегрального оценивания за счет включения модели формирования управляющих рекомендаций дает возможность определять первопричины состояния, выделять проблемные и перспективные направления развития для территорий и планировать мероприятия при реализации национальных проектов на уровне муниципальных образований.

Список источников

1. Айвазян С.А. Анализ синтетических категорий качества жизни населения субъектов Российской Федерации: их измерение, динамика, основные тенденции / С.А.Айвазян // Уровень жизни населения регионов России, 2002. – № 11. – С. 13-24.
2. Спиридонов С.П. Индикаторы качества жизни и методологии их формирования / С.П. Спиридонов // Вопросы современной науки и практики. Университет им. В.И. Вернадского, 2010. – № 10-12. – С. 208-223.
3. Указ Президента Российской Федерации от 21.07.2020 № 474 О национальных целях развития Российской Федерации на период до 2030 года. – URL: http://www.kremlin.ru/acts/bank/45726 (дата обращения: 30.08.2023).
4. Harrell Rodney, Jana Lynott, Shannon Guzman. "Is this a good place to live? Measuring community quality of life for all ages." AARP Public policy institute, Washington, 2014, 28 p.
5. Mehdi T. Stochastic dominance approach to OECD's Better Life Index. Social indicators research, 2019, v. 143, no. 3, pp. 917-954.
6. Piwowarski M., Maison D., Wątróbski J. Application of VMCM method (Vector Measure Construction Methods) to estimate consumer's quality of life in EU countries–dynamic perspective. Procedia Computer Science, 2019, v. 159, pp. 2404-2413.
7. Seko M. Earthquake risk and a quality of life index. Housing markets and household behavior in Japan. – Springer, Singapore, 2019, v. 19, pp. 157-175.
8. Bartlomowicz T., Cheba K. Multidimensional comparative analysis of the European Union countries in the field of sustainable development. Prace naukowe uniwersytetu ekonomicznego we Wrocławiu, 2017, no. 476, pp. 118-126.
9. Beamer L.C., Grant M. Using the dermatology life quality index to assess how breast radiodermatitis affects patients' quality of life. Breast cancer: basic and clinical research, 2019, v. 13, pp. 1-7.
10. Daye M. et al. Evaluation of the effect of COVID-19 pandemic on dermatological diseases with dermatological quality life index. Dermatologic therapy, 2020, v. 33, no. 6, pp. 1-4.
11. Baek H.J. et al. Health-related quality of life of moderate and severe haemophilia patients: Results of the haemophilia-specific quality of life index in Korea. Plos one, 2020, v. 15, no. 9, pp. 1-12.
12. Малков С. Ю. Межстрановый анализ качества жизни на основе индекса "Life quality index" / С.Ю. Малков, Д.М. Мусиева, С.Э. Билюга // Информационные войны, 2020. – №. 1. – С. 61-69.
13. Сукнёва С.А. Система мониторинга качества жизни населения муниципальных образований северных регионов / С.А. Сукнёва, Е.Н. Семенова // Устойчивый Север: общество, экономика, экология, политика, 2019. – Т. 1. – С. 216-221.
14. Сукнёва С.А., Семёнова Е.Н. Мониторинг качества жизни в муниципальных образованиях: доходы населения, занятость / С.А. Сукнёва, Е.Н. Семенова // Проблемы современной экономики, 2019. – №. 1. – С. 149-152.
15. Борисов М. А. Индекс качества городской среды муниципальных образований российской федерации как индикатор местного и регионального развития / М.А.Борисов //Мировая наука, 2019. – №. 12. – С. 81-87.
16. Пенькова Т.Г. Метод построения территориально-ориентированной нормативной модели для оценивания качества жизни с учетом спецификации территорий (на примере Красноярского края). / Т.Г. Пенькова, А.М. Метус, В.В. Ничепорчук, А.И. Ноженков // Информатизация и связь,2021. – № 5. – С. 39-46.

17. Пенькова Т.Г. Метод расчета интегральной оценки качества жизни муниципальных образований региона в разрезе реализации национальных проектов / Т.Г. Пенькова, А.М. Метус, Л.Ф. Ноженкова, Р.В. Морозов // Современные наукоемкие технологии, 2022. – № 7 – С. 79-87.
18. Куприенко П.С. Теория систем: системный подход к моделированию ситуации в техносфере / П.С. Куприенко, Д.Д. Лапшин. – Воронеж: Воронежский государственный технический университет, 2012. – 105 с.
19. Busemeyer J.R., Diederich A. Cognitive modelling. London, SAGE Publication Ltd., 2010, 211 p.
20. Метус А.М. Программные средства Интернет-платформы рейтингового оценивания качества жизни территорий / А.М. Метус, Р.В. Морозов, А.И. Ноженков // Информационные и математические технологии в науке и управлении, 2022. – № 3(27). – С. 5-14.

Алдошкина Анна Михайловна. *Программист I категории отдела прикладной информатики Института вычислительного моделирования СО РАН. Основные направления исследования: методы и технологии интеллектуального анализа данных, базы данных, информационные системы. AuthorID: 847927, SPIN: 6760-0584, ORCID: 0000-0003-0547-5999, metus@icm.krasn.ru, 660036, Россия, г. Красноярск, Академгородок, д. 50, стр.44.*

Пенькова Татьяна Геннадьевна. *Кандидат технических наук, старший научный сотрудник отдела прикладной информатики Института вычислительного моделирования СО РАН, доцент кафедры информатики Сибирского федерального университета. Основные направления исследований включают методы и технологии интеллектуального анализа данных, экспертные системы, системы поддержки принятия решений. AuthorID: 604225, SPIN: 2281-3852, ORCID: 0000-0002-0057-0535, penkova_t@icm.krasn.ru, 660036, Россия, г. Красноярск, Академгородок, д. 50, стр. 44.*

Ноженкова Людмила Федоровна. *Доктор технических наук, профессор, главный научный сотрудник отдела прикладной информатики Института вычислительного моделирования СО РАН. Научная деятельность посвящена исследованиям в области интеллектуальных информационных технологий и анализа данных. AuthorID: 70114, SPIN: 8354-3536, ORCID: 0000-0002-9277-8981, expert@icm.krasn.ru, 660036, Россия, г. Красноярск, Академгородок, д. 50, стр. 44.*

UDC 332.02:004.02
DOI:10.25729/ESI.2024.34.2.012

Decision making support in territory management based on the method of integral estimation of the life quality (in case of Krasnoyarsk region)

Anna M. Aldoshkina, Tatiana G. Penkova, Ludmila F. Nozhenkova

Institute of computational modeling SB RAS,
Russia, Krasnoyarsk, *metus@ism.krasn.ru*

Abstract. This paper presents the technique of decision making support in territory management based on the method of integral estimation of the life quality. A model of management recommendations production based on the analysis of the current state of life quality is proposed. The Internet platform for municipalities monitoring and rating assessment of the life quality of territories in the context of the national projects is presented. The results of a study of rating assessment of the life quality of Krasnoyarsk region territories are presented.

Keywords: decision making support, life quality, integral estimation, management recommendations

References

1. Ajvazyan S.A. Analiz sinteticheskih kategorij kachestva zhizni naseleniya sub"ektov Rossijskoj Federacii: ih izmerenie, dinamika, osnovnye tendencii [Analysis of synthetic categories of the life quality of the population of the Russian Federation regions: measurement, dynamics, trends]. Uroven' zhizni naseleniya regionov Rossii [Life quality of the Russian Federation regions], 2002, no. 11, pp. 13-24.

2. Spiridonov S.P. Indikatory kachestva zhizni i metodologii ih formirovaniya [Life quality indicators and methodology of its formation]. Voprosy sovremennoj nauki i praktiki. Universitet im. V.I. Vernadskogo [Issues of modern science and practice. V.I. Vernadsky National University], 2010, no. 10-12, pp. 208-223.
3. Ukaz Prezidenta Rossiyskoy Federatsii ot 21.07.2020 № 474 “O natsional′nykh tselyakh razvitiya Rossiyskoy Federatsii na period do 2030 goda” [Decree of the President of the Russian Federation of July 21, 2020, no. 474 “On the national development goals of the Russian Federation for the period up to 2030”]. Prezident Rossii: ofitsial′nyy sayt [President of Russia, official website], available at: http://www.kremlin.ru/acts/bank/45726 (accessed 08/30/2023).
4. Harrell Rodney, Jana Lynott, Shannon Guzman. "Is this a good place to live? Measuring community quality of life for all ages." AARP Public policy institute, Washington, 2014, 28 p.
5. Mehdi T. Stochastic dominance approach to OECD’s Better Life Index. Social indicators research, 2019, v. 143, no. 3, pp. 917-954.
6. Piwowarski M., Maison D., Wątróbski J. Application of VMCM method (Vector Measure Construction Methods) to estimate consumer’s quality of life in EU countries–dynamic perspective. Procedia Computer Science, 2019, v. 159, pp. 2404-2413.
7. Seko M. Earthquake risk and a quality of life index. Housing markets and household behavior in Japan. – Springer, Singapore, 2019, v. 19, pp. 157-175.
8. Bartlomowicz T., Cheba K. Multidimensional comparative analysis of the European Union countries in the field of sustainable development. Prace naukowe uniwersytetu ekonomicznego we Wrocławiu, 2017, no. 476, pp. 118-126.
9. Beamer L.C., Grant M. Using the dermatology life quality index to assess how breast radiodermatitis affects patients’ quality of life. Breast cancer: basic and clinical research, 2019, v. 13, pp. 1-7.
10. Daye M. et al. Evaluation of the effect of COVID-19 pandemic on dermatological diseases with dermatological quality life index. Dermatologic therapy, 2020, v. 33, no. 6, pp. 1-4.
11. Baek H.J. et al. Health-related quality of life of moderate and severe haemophilia patients: Results of the haemophilia-specific quality of life index in Korea. Plos one, 2020, v. 15, no. 9, pp. 1-12.
12. Malkov S.YU., Musieva D.M., Bilyuga S.E. Mezhstranovyj analiz kachestva zhizni na osnove indeksa “Life quality index” [Cross-country analysis of the quality of life based on the index “Life quality index”]. Informacionnye vojny [Information wars], 2020, no. 1, pp. 61-69.
13. Suknyova S.A., Semenova E.N. Sistema monitoringa kachestva zhizni naseleniya municipal'nyh obrazovanij severnyh regionov [The system of monitoring the quality of life of the population of municipalities of the northern regions]. Ustojchivyj Sever, obshchestvo, ekonomika, ekologiya, politika [Sustainable North, society, economy, ecology, politics], 2019, v. 1, pp. 216-221.
14. Suknyova S.A., Semyonova E.N. Monitoring kachestva zhizni v municipal'nyh obrazovaniyah: dohody naseleniya, zanyatosti [Monitoring the quality of life in municipalities: incomes of the population, employment]. Problemy sovremennoj ekonomiki [Problems of modern economy], 2019, no. 1, pp. 149-152.
15. Borisov M.A. Indeks kachestva gorodskoj sredy municipal'nyh obrazovanij Rossijskoj Federacii kak indikator mestnogo i regional'nogo razvitiya [Urban environment quality index of municipalities of the Russian Federation as an indicator of local and regional development]. Mirovaya nauka [World Science], 2019, no. 12, pp. 81-87.
16. Penkova T.G., Metus, A.M., Nicheporchuk V.V., Nozhenkov A.I. Metod postroyeniya territorial′nooriyentirovannoy normativnoy modeli dlya otsenivaniya kachestva zhizni s uchetom spetsifikatsii territoriy (na primere Krasnoyarskogo kraya) [Method for constructing a territorial-oriented normative model for estimation of the life quality taking into account the territories specification (in the case of Krasnoyarsk region)]. Informatizatsiya i svyaz′ [Informatization and communication], 2021, no. 5, pp. 39-46.
17. Penkova T.G., Metus, A.M., Nozhenkova L.F. , Morozov R.V., Metod rascheta integral'noj ocenki kachestva zhizni municipal'nyh obrazovanij regiona v razreze realizacii nacional'nyh proektov [Method of integral estimation of the life quality for municipal territories in the context of national projects implementation]. Sovremennye naukoemkie tekhnologii [Modern high-tech technologies], 2022, no. 7, pp. 79-87.
18. Kuprienko P.S., Lapshin D.D. Teoriya sistem: sistemnyj podhod k modelirovaniyu situacii v tekhnosfere [Systems theory: a systematic approach to modeling a situation in the technosphere]. Voronezh, Voronezhskij gosudarstvennyj tekhnicheskij universitet, 2012, 105 p.
19. Jerome R. Busemeyer, Adele Diederich Cognitive modelling, London, SAGE Publication Ltd., 2010, 211 p.
20. Metus, A.M., Morozov R.V., Nozhenkov A.I. Programmnye sredstva Internet-platformy rejtingovogo ocenivaniya kachestva zhizni territorij [Software tools of the Internet platform for rating assessment of life the quality of territories]. Informacionnye i matematicheskie tekhnologii v nauke i upravlenii [Information and mathematical technologies in science and management], 2022, no. 3, pp. 5-14.

Aldoshkina Anna Mikhailovna. *Programmer of the 1st category, department of applied informatics, Institute of computational modeling, SB RAS. Main areas of research: methods and technologies of data mining, databases, information systems. AuthorID: 847927, SPIN: 6760-0584, ORCID: 0000-0003-0547-5999, metus@icm.krasn.ru, 660036, Russia, Krasnoyarsk, Akademgorodok, 50/44.*

Penkova Tatiana Gennad'evna. *Ph.D., senior research associate, department of applied informatics, Institute of computational modeling, SB RAS. An associate professor at the department of informatics of the Siberian Federal university. Main areas of research: methods and technologies of data mining, expert systems, decision support systems. AuthorID: 604225, SPIN: 2281-3852, ORCID: 0000-0002-0057-0535, penkova_t@icm.krasn.ru, 660036, Russia, Krasnoyarsk, Akademgorodok, 50/44.*

Nozhenkova Ludmila Federovna. *Doctor of technical sciences, professor, chief researcher, department of applied informatics, institute of computational modeling, SB RAS. Scientific activity is devoted to research in the field of intelligent information technologies and data analysis. AuthorID: 70114, SPIN: 8354-3536, ORCID: 0000-0002-9277-8981, expert@icm.krasn.ru, 660036, Russia, Krasnoyarsk, Akademgorodok, 50/44.*

Статья поступила в редакцию 01.09.2023; одобрена после рецензирования 31.05.2024; принята к публикации 06.06.2024.

The article was submitted 09/01/2023; approved after reviewing 05/31/2024; accepted for publication 06/06/2024.

УДК 004.725
DOI:10.25729/ESI.2024.34.2.013

Построение оптимизационной модели беспроводной внутренней сети для использования технологии интернета вещей

Сиротинин Анатолий Андреевич, Володько Ольга Станиславовна

Институт вычислительного моделирования СО РАН,
Россия, Красноярск, *slitch@icm.krasn.ru*

Аннотация. В работе предложен подход для построения оптимальной беспроводной сенсорной сети внутри помещения для технологии интернета вещей. Для повышения энергоэффективности сети используется гибридная структура сети на основе стандартов Wi-Fi и ZigBee. Преимуществом предложенного подхода является то, что сеть проектируется на трехмерной модели здания, учитывая затухания сигнала не только в стенах, но и в перекрытиях этажей. Для оптимального расположения связующих и сенсорных узлов предлагается использовать комбинацию радиоволнового (модели Мотли-Кинана) и оптимизационного (генетического алгоритма) методов.

Ключевые слова: беспроводные сенсорные сети, модель Мотли-Кинана, уровень мощности сигнала, интернет вещей, ячеистые сети, задача оптимизации, генетические алгоритмы

Цитирование: Сиротинин А.А. Построение оптимизационной модели беспроводной внутренней сети для использования технологии интернета вещей / А.А. Сиротинин, О.С. Володько // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 135-143. – DOI: 10.25729/ESI.2024.34.2.013.

Введение. В настоящее время существует проблема организации беспроводной сенсорной сети для технологии интернета вещей [1, 2]. Беспроводная сенсорная сеть (БСС) является перспективным инструментом для создания безопасного и комфортного автоматизированного рабочего пространства. Важными преимуществами БСС являются отсутствие кабельной инфраструктуры, малые размеры устройств, низкое энергопотребление, встроенный радиоинтерфейс, достаточно высокая вычислительная мощность и относительно низкая стоимость. БСС – это распределенная система миниатюрных электронных устройств (сетевых узлов), таких, как датчики дыма, открытия дверей, влажности, протечки, температуры и исполнительных устройств. Такая сеть собирает данные о параметрах среды и передает их на сервер по узлам связи с помощью беспроводного подключения. Несколько узлов такой сети выступают в качестве шлюзов (хабов), которые осуществляют связь с глобальной сетью. БСС имеет тип ячеистой структуры, с базовыми точками доступа, выступающими в роли ведущих, прерывание связи с которыми ведет к потере данных и не позволяет оперативно предпринимать действия по устранению нештатного изменения параметров. Поскольку большинство параметров от датчиков являются критически важными для пользователя, необходимо обеспечить стабильную и надежную работу сети.

В устройствах интернета вещей для обмена данными внутри помещений самыми распространёнными являются стандарты Wi-Fi, Bluetooth, Z-Wave, Thread и ZigBee [3–5]. У каждого из них есть свои преимущества и недостатки. Например, Wi-Fi имеет преимущество в универсальности, при этом недостатками являются высокая стоимость качественного современного оборудования и высокое энергопотребление. Bluetooth имеет небольшой радиус действия, но низкое энергопотребление, особенно у версии LE (Low Energy), и низкую скорость передачи данных. ZigBee [6] является промежуточным вариантом, так как имеет низкое энергопотребление, радиус действия сопоставим с Wi-Fi, но скорость передачи существенно ниже. Плюсом стандартов ZigBee и Wi-Fi (нового поколения) является то, что стандарты имеют ячеистую, самоорганизующуюся структуру и большинство точек доступа оснащены одновременно модулями ZigBee и Wi-Fi. Стандарт Z-Wave работает на более низкой частоте, что является преимуществом и недостатком одновременно. Низкая частота

передатчика позволяет распространять сигнал на бóльшие расстояния при той же мощности, но при этом существенно снижаются скорость передачи и количество передаваемой информации. Thread является прямым конкурентом ZigBee, но с обновленными алгоритмами шифрования AES-CCM. Его существенным недостатком является малая распространенность.

В связи с тем, что для функционирования беспроводных сетей, включая сенсорные, требуются точки доступа, то при выборе их мест расположения должны учитываться конструктивные особенности здания, материалы конструкций и отделки. Для решения проблемы оптимального расположения точек доступа предлагаются как различные готовые программные пакеты [7], которые позволяют оценивать уровень мощности сигнала в зависимости от физических свойств окружающей среды, так и методы поиска размещения точек доступа для обеспечения оптимальной зоны покрытия сигнала. В то же время, программные пакеты имеют недостатки, которые не позволяют использовать их большинству людей. Крупные производители оборудования, например, Cisco [8], D-Link [9], предоставляют собственные программные решения, рассчитанные только под собственные продукты. Существуют программные пакеты, рассматривающие точки доступа от разных производителей, но они предлагают пользователю минимальный набор опций за высокую стоимость [10–12]. Кроме того, имеющееся программные пакеты в основном рассматривают двумерные модели помещений и учитывают фиксированные типы препятствий из встроенных информационных таблиц, не позволяя делать поправку, например, на слои декоративной отделки.

В литературе существуют различные способы для постановки задач оптимального размещения беспроводных точек доступа в сенсорных сетях и методов их решения. Эти методы разделяются на детерминированные и стохастические. В детерминированных методах гарантируется достижение глобального или, по крайней мере, локального минимума, в то время, как в стохастических алгоритмах минимум достигается только с некоторой вероятностью. Однако, стохастические методы быстрее находят глобальный оптимум по сравнению с детерминированными, особенно, когда найти глобальный оптимум – сложная вычислительная задача. К детерминированным относятся, например, такие методы, как методы перебора, алгоритмы теории графов [13, 14]. К стохастическим относятся алгоритмы, вдохновленные природой, например, оптимизации роя частиц, муравьиные и генетические алгоритмы [15].

Для решения задачи оптимального расположения точек доступа необходимо комбинировать радиоволновое моделирование с подходящими оптимизационными методами, например, такими, как генетические алгоритмы. Что касается радиоволновых моделей, в большинстве исследований используется модель простых логарифмических потерь на расстоянии (LDPL) [16, 17], которая не учитывает затуханий в стенах и перекрытиях в отличие от моделей распространения лучей [18] или более практической модели Мотли-Кинана [19]. Преимущества эмпирических моделей (например, модели Мотли-Кинана) перед детерминистическими моделями распространения (например, моделью трассировки лучей) включают вычислительную эффективность, небольшое количество параметров и простоту применения [20]. Дополненная модель Мотли-Кинана, предложенная в [21], учитывает затухания сигнала в стенах и перекрытиях этажей и показывает более точные результаты, чем мультистенная модель COST 231 [22].

Целью настоящей работы является построение оптимальной БСС ячеистой структуры внутри помещений для технологии интернета вещей. Предложенная БСС рассматривается на примере здания научного института.

1. Беспроводная сенсорная сеть. Для проектирования сети предлагается структура БСС, показанная на рис. 1. Для обеспечения надежности и отказоустойчивости сеть должна быть построена следующим образом: каждый сенсор должен видеть хотя бы два связующих узла (хаба), каждый хаб должен видеть хотя бы ещё один хаб, в случае отказа ведущего узла (лидера) любой из хабов может взять на себя право быть ведущим. Вывод данных в интернет осуществляется через хаб, работающий в режиме шлюза. Благодаря этим условиям ячеистый тип сетей легко масштабируется и являются очень надёжным.

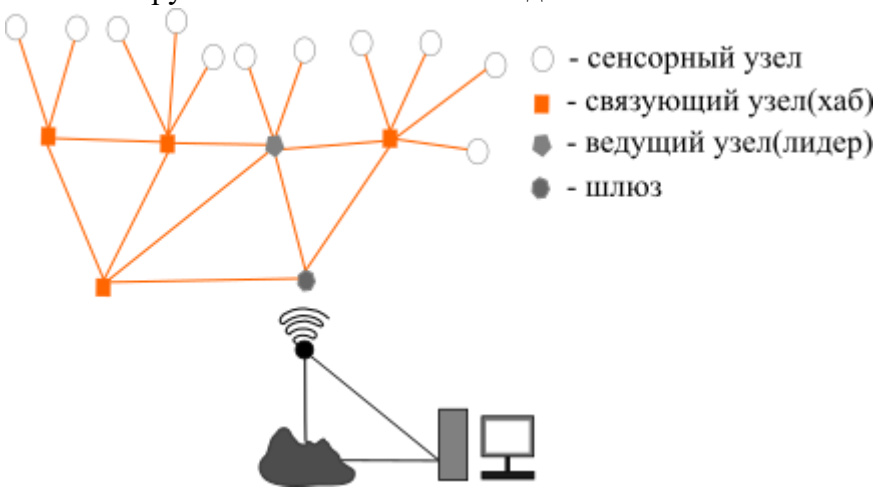


Рис. 1. Структура беспроводной сенсорной сети

В настоящей работе стандарт Wi-Fi используется для соединения ведущего узла с хабами и хабов между собой. При этом хабы могут располагаться как в пределах одного этажа, так и на разных этажах. Таким образом, рассматривается трехмерная модель здания. Сенсоры будут связываться с хабами по стандарту ZigBee для увеличения энергоэффективности сети.

2. Оценка уровня мощности сигнала. Для получения начальной картины покрытия уровнем мощности сигнала Wi-Fi контрольных точек, в качестве которых выбраны несколько хабов, в обследуемом здании методом измерений была построена цифровая модель. В качестве радиоволновой модели для расчета уровней мощности сигнала была выбрана дополненная модель Мотли-Кинана. Для обследуемого здания проведено сравнение измеренных уровней мощности сигнала с рассчитанными по дополненной модели Мотли-Кинана, чтобы убедиться в точности расчетов.

Уровень мощности сигнала стандарта Wi-Fi по модели Мотли-Кинана [23] в зоне прямой видимости рассчитывается следующим образом:

$$RSS(d) = RSS(d_0) - 10a\log\left(\frac{d}{d_0}\right), \quad (1)$$

где $RSS(d)$ – мощность сигнала (RSS – Received Signal Strength) (дБм), определяемая устройством на расстоянии d от точки доступа, $RSS(d_0)$ – мощность сигнала (дБм), измеренная на расстоянии $d_0 = 1$ м от точки доступа, a – коэффициент, учитывающий потери сигнала в окружающей среде.

В работе [21] представлена дополненная модель Мотли-Кинана, которая учитывает также затухание сигнала в стенах, перегородках и межэтажных перекрытиях:

$$RSS(d) = RSS(d_0) - 10a\log\left(\frac{d}{d_0}\right) + \sum_{r=1}^{L} l_r\omega_r + FAF, \quad (2)$$

где L – количество стен, l_r – количество стен с затуханием ω_r, FAF – затухание в межэтажном перекрытии (таблица 1). Усредненные значения затуханий приведены в таблице 2 [24].

Для получения начальной картины уровня мощности сигнала от беспроводных точек доступа в здании научного института методом измерений была проведена оценка уровней мощности сигнала от точек доступа 802.11n. В качестве примера рассмотрены уровни

мощности сигнала на 4-ом этаже. На рис. 2 (а) показано покрытие уровнем мощности сигнала 4-го этажа от всех точек доступа, на рис. 2 (б) – уровень мощности сигнала с отключенными точками доступа 4-ого этажа, на рис. 2 (в) – уровень мощности только от точек доступа 4-го этажа, на рис. 2 (г) – покрытие уровнем мощности сигнала на 3-ем этаже от точек доступа 4-го этажа. В дальнейшем измеренные уровни мощности сигналов были использованы для валидации с рассчитанными по дополненной модели Мотли-Кинана (2).

Таблица 1. Зависимость FAF от количества этажей

Этаж	FAF, дБ
1	-12,9
2	-18,7
3	-24,4
4	-27

Таблица 2. Затухания сигнала Wi-Fi при прохождении через препятствия

Класс препятствия	Затухание, дБ
Несущая стена (30,5 см)	-20-25
Межкомнатная стена (15 см)	-15-20
Бетонная плита перекрытия	-15-25
Монолитное железобетонное перекрытие	-20-25
Деревянная стена	-10

Для валидации уровней мощности сигналов был проведен расчёт по модели Мотли-Кинана для шести контрольных точек, находящихся в различных условиях:

1. Контрольная точка находится в прямой видимости от точки доступа, поляризация вертикальная.
2. Контрольная точка находится за межкомнатной стеной, поляризация вертикальная.
3. Контрольная точка находится под точкой доступа этажом ниже, поляризация вертикальная.
4. Контрольная точка находится в прямой видимости, горизонтальная поляризация.
5. Контрольная точка находится за межкомнатной стеной, горизонтальная поляризация.
6. Контрольная точка находится под точкой доступа этажом ниже, горизонтальная поляризация.

Для 1-го случая уровень мощности сигнала рассчитывается по формуле (1), где $a = 20$ дБ – стандартное значение затухания в воздухе для частоты 2,4 ГГц, $RSS(d_0) = -39$ дБм, $d = 4$ м и тогда $RSS(d) = -51$ дБм.

Для случав 2–6 уровень мощности сигнала рассчитывается по формуле (2):

Для 2-го случая уровень мощности сигнала $RSS(d) = -68$ дБм, где $RSS(d_0) = -39$ дБм, $d = 5$ м, $l_r = 1$, $\omega_r = -15$ дБ.

Для 3-го случая уровень мощности сигнала $RSS(d) = -68$ дБм, где $RSS(d_0) = -39$ дБм, $d = 6{,}4$ м, $FAF = -12{,}9$ дБ.

Для 4-го случая уровень мощности сигнала $RSS(d) = -70$ дБм, где $RSS(d_0) = -37{,}5$ дБм, $d = 7{,}8$ м, $l_r = 1$.

Для 5-го случая уровень мощности сигнала $RSS(d) = -72$ дБм, где $RSS(d_0) = -37{,}5$ дБм, $d = 9$ м, $l_r = 1$.

Для 6-го случая уровень мощности сигнала $RSS(d) = -62$ дБм, где $RSS(d_0) = -37$ дБм, $d = 4$ м, $FAF = -12{,}9$ дБ.

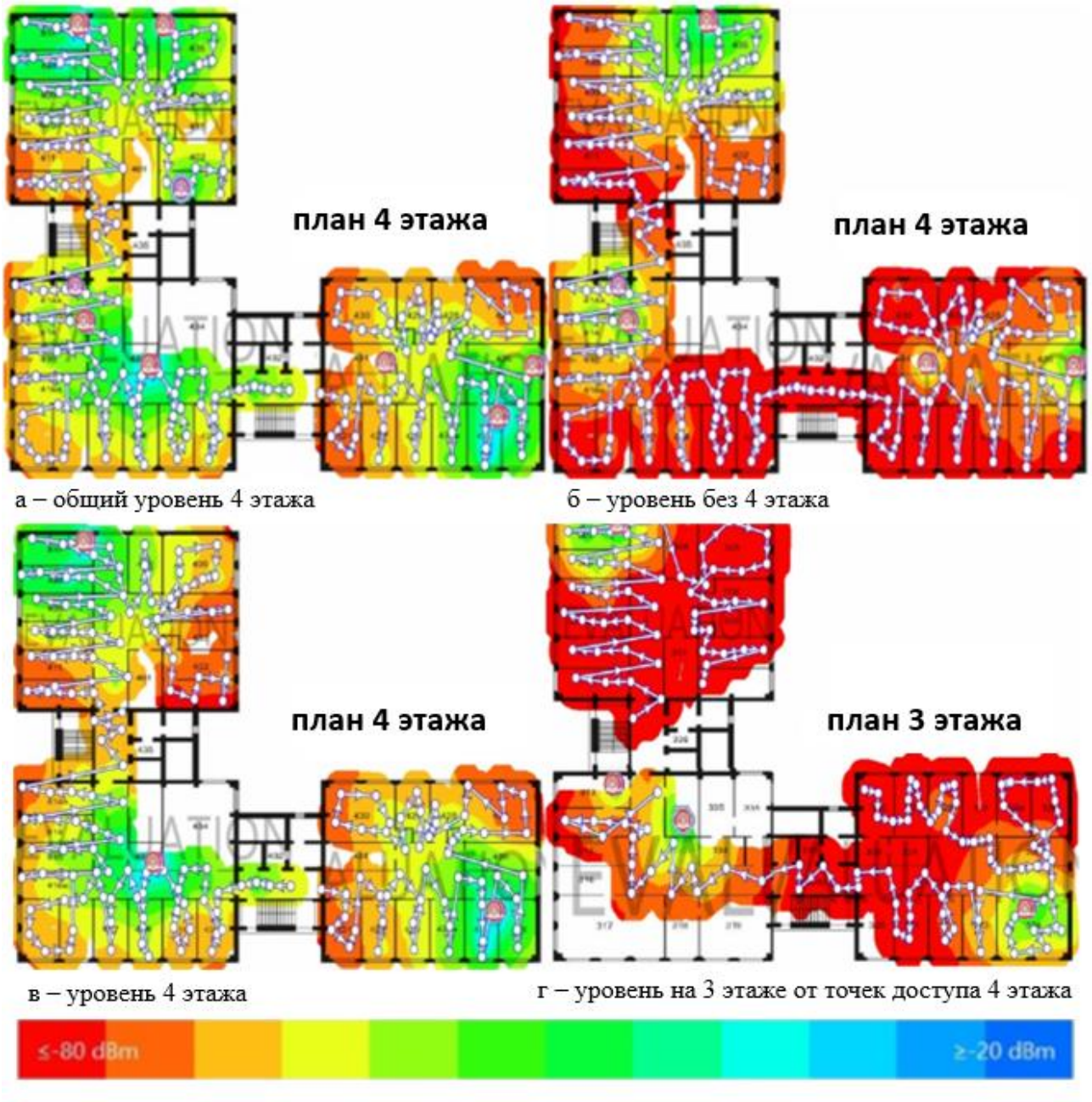


Рис. 2. Уровень мощности сигнала на 4-ом этаже здания

Валидация, измеренных уровней мощности сигнала с рассчитанными по дополненной модели Мотли–Кинана, показала соответствие значений с минимальной погрешностью 2 дБ.

После получения полной картины состояния беспроводных сетей в здании, выделены параметры, которые необходимо оптимизировать и будет поставлена соответствующая задача оптимизации. В качестве начальных данных в задаче оптимизации предполагается использовать скорости, определенные в зависимости от уровней мощности сигнала, которые рассчитаны по дополненной модели Мотли-Кинана. (таблица 3).

Таблица 3. Зависимость скорости от уровня мощности сигнала

Скорость, Мбит/с	Уровень мощности, дБм
150	-59
100	-78
50	-87
0	-97

3. Задача оптимизации. Поставим задачу оптимизации для оптимального расположения беспроводных точек доступа (хабов) [25]. Задача оптимизации заключается в оптимальной расстановке хабов с учётом максимально возможной скорости передачи данных и минимально возможных затрат на оборудование и коммуникации.

Определим начальные условия. Так как качественное оборудование для беспроводной передачи данных стоит довольно дорого, то первым условием зададим снизить количество хабов, оптимально расположив их в здании. Вторым важным условием будет скорость передачи данных с контрольной точки, она должна стремиться к максимуму.

Предположим, что есть несколько зон с хорошим уровнем приема. Определяем приоритет для каждой из них. Также следует установить минимальную скорость для любой зоны. Таким образом, имеем N_c ячеек, для которых необходимо задать приоритет F_j и минимальную скорость V_j. Для начала нам необходимо разбить обследуемый объект на клетки, количество клеток будет N_p, в каждой клетке можно разместить хаб N_{tr}, для каждого хаба нужно учесть стоимость коммуникаций, которая известна и равна $C_i: 1 \le i \le N_p$, также известна стоимость хаба $C_k: 1 \le k \le N_c$. Определяем матрицу X булевых переменных x_{ik}. Значение переменной x_{ik} , равное 1, означает, что хаб k-типа помещен в i-й клетке, а значение, равное 0, означает его отсутствие. Таким образом, необходимо получить максимальную скорость в максимальном количестве клеток в соответствии с приоритетами, при минимальных затратах. Запишем условия:

$$V_j(X)F_j \to max, \qquad V_j \ge V_{min\ j}, \qquad \sum_{i=1}^{N_p}\sum_{k=1}^{N_{tr}}(C_i + C_k)x_{ik} \to min,$$

$$\sum_{k=1}^{N_{tr}} x_{ik} \le 1, \forall i \in [1; N_p], \qquad x_{ik} \in \{0,1\}, \forall i \in [1; N_p], \forall k \in [1, N_{tr}].$$

Здесь $V_j(X)$ – скорость в контрольной точке i (таблица 3), F_j – приоритет i-й контрольной точки, N_p – возможные места для установки хабов, N_{tr} – количество хабов, $V_{min\ j}$ – минимально необходимая скорость.

В дальнейшем задача оптимизации может быть расширена с учетом расстановки сенсоров, подключенных по стандарту ZigBee. Для решения задачи оптимизации планируется применить различные виды генетических алгоритмов [1, 26]. В последующих работах будут представлены решения соответствующих задач оптимизации.

Заключение. В работе предложен подход для построения оптимальной БСС для технологии интернета вещей. Предложенный подход рассматривается на примере здания научного института. Одним из преимуществ данного подхода является то, что БСС проектируется на трехмерной карте здания с учетом затухания сигнала не только в стенах, но и в перекрытиях этажей. Кроме того, для повышения энергоэффективности сети за основу взята гибридная структура сети, в которой технология Wi-Fi используется для связи хабов со шлюзом и хабов между собой, а технология ZigBee будет использоваться для соединения сенсоров с хабами. Начальные данные для решения задачи оптимизации получены с помощью расчетов по дополненной модели Мотли-Кинана, валидация которой была проведена на основе измерений, проведенных в обследуемом здании. Для оптимального расположения хабов поставлена задача оптимизации, которая в дальнейшем будет расширена с учётом расстановки сенсоров. Решение задачи оптимизации будет выполнено с использованием генетических алгоритмов.

Таким образом, предложен подход, который может позволить построить оптимальную БСС для технологии интернета вещей внутри помещений.

Список источников

1. Srinidhi N.N., Kumar S.M.D., Venugopal K.R. Network optimizations in the Internet of Thing: A review. Engineering science and technology, an international journal, 2019, vol. 22, no. 1, pp. 1-21.
2. Арефьев А.В. Проблемы построения больших локальных сетей интернета вещей / А.В. Арефьев, О.В. Афанасьева, Н.А. Вешев и др. // Известия Тульского государственного университета, 2022. – Т. 10. – С. 261-267.
3. Jia M., Komeily A., Wang Y., Srinivasan R.S. Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications. Automation in construction, 2019, vol. 101, pp. 111-126.
4. Wang W., Capitaneanu S.L., Marinca D. et al. Comparative analysis of channel models for industrial IoT wireless communication. IEEE Access, 2019, vol. 7, pp. 91627-91640.
5. Соловьева В.О. Подходы к построению ячеистых сетей / В.О. Соловьева // Информационные технологии, системный анализ и управление, 2021. – С. 354-358.
6. Adi P.D.P., Sihombing V., Siregar V.M.M. et al. A performance evaluation of ZigBee mesh communication on the Internet of Things (IoT), 3rd East Indonesia conference on computer and information technology (eiconcit), ieee, 2021, pp. 7-13.
7. Pendão C., Moreira A. Automatic RF interference maps and their relationship with Wi-Fi positioning errors, International conference on indoor positioning and indoor navigation (IPIN). IEEE, 2019, pp. 1-8.
8. CISCO. Available at: https://www.cisco.com/en/US/docs/solutions/Enterprise/Mobility/emob30dg/Manage.html#wp999519 (accessed: 09/20/2023).
9. D-Link. Available at: https://www.dlink.ru/tools/wi-fi/ (accessed: 09/19/2023).
10. Tamograph. Available at: https://www.tamos.ru/products/wifi-site-survey/ (accessed: 09/20/2023).
11. Ekahau. Available at: https://www.ekahau.com (accessed: 09/17/2023).
12. iBwave. Available at: https://www.ibwave.com/ibwave-design (accessed: 09/20/2023).
13. Тайк А.М. Применение алгоритма перебора для оптимизации топологии беспроводных сетей. / А.М. Тайк, С.А. Лупин, Ю.Ф. Вагапов // International journal of open information technologies, 2016. – Т. 4(9). – С.80-86.
14. Priyadarshi R., Gupta B. and Anurag A. Wireless sensor networks deployment: a result-oriented analysis. Wireless personal communications, 2020, vol. 113, pp.843-866.
15. Singh A., Sharma S., Singh J. Nature-inspired algorithms for wireless sensor networks: A comprehensive survey. Computer science review, 2021, vol 39, p.100342.
16. Valadares D.C.G., de Araújo J.M.F.R. et al. E.U.K., 802.11g signal strength evaluation in an industrial environment. Internet of Things, 2020, vol 9, p.100163.
17. Zhao B., Zheng C., Ren X., et al. An optimization method combining RSSI and PDR data to estimate distance between smart devices. TechRxiv, 2020, DOI:10.36227/techrxiv.13488738.v1.
18. El-Maghrabi H.M., A standard ray tracing technique for predicting signal strength of wireless sensor network in smart building. Progress in electromagnetics research letters, 2022, vol 105, pp.79-84.
19. Tian Y., Huang B., Jia B., et al. Optimizing AP and Beacon Placement in WiFi and BLE hybrid localization. Journal of Network and computer applications, 2020, vol 164, p. 102673.
20. Kar K., Datta S., Pal M., et al. Motley Keenan model of in-building coverage analysis of IEEE 802.11 n WLAN signal in electronics and communication engineering department of National Institute of Technology Durgapur. In 2016 international conference on microelectronics, computing and communications (MicroCom), IEEE, 2016, pp. 1-6.
21. Zhang Y., Wang F., Shen Y. et al. A study of indoor distributed calculation model of mobile communication. Information computing and applications: second international conference, Qinhuangdao, China. Proceedings, Part I 2, 2011, pp. 458-465.
22. Andrade C.B., Hoefel R.P. F. IEEE 802.11 WLANs: A comparison on indoor coverage models. CCECE, Calgary, Canada, 2010, pp. 1-6, DOI:10.1109/CCECE.2010.5575205.
23. Motley A.J., Keenan J.M.P. Personal communication radio coverage in buildings at 900 MHz and 1700 MHz. Electronics Letters, 1988, vol. 24, no. 12, pp. 763-764.
24. Ахметшин Д.А. Постановка математической задачи организации бесшовной беспроводной сети по технологии WiFi с фильтрацией контента / Д.А. Ахметшин, Д. Ку, Н.К. Нуриев и др // Современные наукоемкие технологии, 2019. – №. 11. – С. 15-23.
25. Kazakovtsev L.A. Wireless coverage optimization based on data provided by built-in measurement tools. World applied sciences journal, 2013, 22, pp. 08-15.
26. Hu J., Goodman E. Wireless access point configuration by genetic programming. Proceedings of the 2004 congress on evolutionary computation, IEEE, 2004, vol. 1, pp. 1178-1184.

Сиротинин Анатолий Андреевич. *Аспирант, институт вычислительного моделирования Сибирского отделения Российской академии наук - обособленное подразделение ФИЦ КНЦ СО РАН, AuthorID: 1215100, SPIN: 9617-6083, slitch@icm.krasn.ru, 660036, Россия, Красноярск, Академгородок, дом. 50, стр. 44.*

Володько Ольга Станиславовна. *Кандидат физико-математических наук, институт вычислительного моделирования Сибирского отделения Российской академии наук, м.н.с., AuthorID: 907835, SPIN 8080-9331, ORCID 0000-0002-0580-9103, osv@icm.krasn.ru, 660036, Россия, Красноярск, Академгородок, дом. 50, стр. 44.*

UDC 004.725
DOI:10.25729/ESI.2024.34.2.013

Building an optimization model of a wireless internal network for using internet of things technology

Anatoliy A. Sirotinin, Olga S. Volodko

Institute of computational modeling SB RAS,
Russia, Krasnoyarsk, *slitch@icm.krasn.ru*

Abstract. In the paper, an approach for project an optimal wireless sensor network indoors for the technology of the Internet of Things is proposed. To improve the energy efficiency of the network is used a hybrid network structure based on Wi-Fi and ZigBee standards. The advantage of the approach is that the network is designed on a three-dimensional model of the building, given the attenuation of the signal not only in the walls, but also in the floors. The combination of radio-wave (Motley-Keenan model) and optimization (genetic algorithm) methods for optimal arrangement of connecting and sensory nodes is used.

Keywords: wireless sensor network, Motley-Keenan model, signal strength, Internet of Things, mesh networks, optimization problem, genetic algorithms

References

1. Srinidhi N.N., Kumar S.M.D., Venugopal K.R. Network optimizations in the Internet of Thing: A review. Engineering science and technology, an international journal, 2019, vol. 22, no. 1, pp. 1-21.
2. Arefyev A.V., Afanasyeva O. V., Veshev N. A. et al. Problemy postroyeniya bol'shikh lokal'nykh setey interneta veshchey [Problems of building large local networks of the Internet of Things]. Izvestiya Tul'skogo gosudarstvennogo universiteta [Proceedings of Tula state university], 2022, vol. 10, pp. 261-267.
3. Jia M., Komeily A., Wang Y., Srinivasan R.S. Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications. Automation in construction, 2019, vol. 101, pp. 111-126.
4. Wang W., Capitaneanu S.L., Marinca D. et al. Comparative analysis of channel models for industrial IoT wireless communication. IEEE Access, 2019, vol. 7, pp. 91627-91640.
5. Solovyova V.O. Podkhody k postroyeniyu yacheistykh setey [Approaches to the construction of mesh networks]. nformatsionnyye tekhnologii, sistemnyy analiz i upravleniye [Information technologies, system analysis and management], 2021, pp. 354-358.
6. Adi P.D.P., Sihombing V., Siregar V.M.M. et al. A performance evaluation of ZigBee mesh communication on the Internet of Things (IoT), 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), IEEE, 2021, pp. 7-13.
7. Pendão C., Moreira A. Automatic RF interference maps and their relationship with Wi-Fi positioning errors, International conference on indoor positioning and indoor navigation (IPIN). IEEE, 2019, pp. 1-8.
8. CISCO. Available at: https://www.cisco.com/en/US/docs/solutions/Enterprise/Mobility/emob30dg/Manage.html#wp999519 (accessed: 09/20/2023).
9. D-Link. Available at: https://www.dlink.ru/tools/wi-fi/ (accessed: 09/19/2023).
10. Tamograph. Available at: https://www.tamos.ru/products/wifi-site-survey/ (accessed: 09/20/2023).
11. Ekahau. Available at: https://www.ekahau.com (accessed: 09/17/2023).
12. iBwave. Available at: https://www.ibwave.com/ibwave-design (accessed: 09/20/2023).

13. Tajk A.M., Lupin S.A., Vagapov YU.F. Primeneniye algoritma perebora dlya optimizatsii topologii besprovodnykh setey [Primenenie algoritma perebora dlya optimizacii topologii besprovodnyh setej]. International journal of open information technologies, 2016, 4(9), pp.80-86.
14. Priyadarshi R., Gupta B. and Anurag A. Wireless sensor networks deployment: a result-oriented analysis. Wireless personal communications, 2020, vol. 113, pp.843-866.
15. Singh A., Sharma S., Singh J. Nature-inspired algorithms for wireless sensor networks: A comprehensive survey. Computer science review, 2021, vol 39, p.100342.
16. Valadares D.C.G., de Araújo J.M.F.R. et al. E.U.K., 802.11g signal strength evaluation in an industrial environment. Internet of Things, 2020, vol 9, p.100163.
17. Zhao B., Zheng C., Ren X., et al. An optimization method combining RSSI and PDR data to estimate distance between smart devices. TechRxiv, 2020, DOI:10.36227/techrxiv.13488738.v1.
18. El-Maghrabi H.M., A standard ray tracing technique for predicting signal strength of wireless sensor network in smart building. Progress in electromagnetics research letters, 2022, vol 105, pp.79-84.
19. Tian Y., Huang B., Jia B., et al. Optimizing AP and Beacon Placement in WiFi and BLE hybrid localization. Journal of Network and computer applications, 2020, vol 164, p. 102673.
20. Kar K., Datta S., Pal M., et al. Motley Keenan model of in-building coverage analysis of IEEE 802.11 n WLAN signal in electronics and communication engineering department of National Institute of Technology Durgapur. In 2016 international conference on microelectronics, computing and communications (MicroCom), IEEE, 2016, pp. 1-6.
21. Zhang Y., Wang F., Shen Y. et al. A study of indoor distributed calculation model of mobile communication. Information computing and applications: second international conference, Qinhuangdao, China. Proceedings, Part I 2, 2011, pp. 458-465.
22. Andrade C.B., Hoefel R.P. F. IEEE 802.11 WLANs: A comparison on indoor coverage models. CCECE, Calgary, Canada, 2010, pp. 1-6, DOI:10.1109/CCECE.2010.5575205.
23. Motley A.J., Keenan J.M.P. Personal communication radio coverage in buildings at 900 MHz and 1700 MHz. Electronics Letters, 1988, vol. 24, no. 12, pp. 763-764.
24. Ahmetshin D.A., Ku D., Nuriev N.K. et al. Postanovka matematicheskoy zadachi organizatsii besshovnoy besprovodnoy seti po tekhnologii WiFi s fil'tratsiyey kontenta [Statement of the mathematical problem of organizing a seamless wireless network using WiFi technology with content filtering]. Sovremennye naukoemkie tekhnologii [Modern high technology], 2019, no. 11, pp. 15-23.
25. Kazakovtsev L.A. Wireless coverage optimization based on data provided by built-in measurement tools. World applied sciences journal, 2013, 22, pp. 08-15.
26. Hu J., Goodman E. Wireless access point configuration by genetic programming. Proceedings of the 2004 congress on evolutionary computation, IEEE, 2004, vol. 1, pp. 1178-1184.

Sirotinin Anatoly Andreevich. *Postgraduate student, the institute of computational modeling of the Siberian Branch of the Russian academy of sciences is a separate division of the FITC KNC SB RAS, SPIN: 9617-6083, AuthorID: 1215100, slitch@icm.krasn.ru, 660036, Krasnoyarsk, Akademgorodok, ICM SB RAS.*

Volodko Olga Stanislavovna. *Ph. D., the institute of computational modeling of the Siberian Branch of the Russian academy of sciences, junior researcher, ORCID: 0000-0002-0580-9103, SPIN: 8080-9331, AuthorID: 907835, osv@icm.krasn.ru, 660036, Krasnoyarsk, Akademgorodok, ICM SB RAS.*

Статья поступила в редакцию 10.07.2023; одобрена после рецензирования 02.04.2024; принята к публикации 30.05.2024.

The article was submitted 07/10/2023; approved after reviewing 04/02/2024; accepted for publication 05/30/2024.

Информационные системы и технологии

УДК 004.415.25
DOI:10.25729/ESI.2024.34.2.014

Система учета заявок для ИТ-подразделений КНЦ СО РАН

Малимонов Максим Игоревич

Федеральный исследовательский центр «Красноярский научный центр СО РАН»,
Россия, Красноярск, *malimonov.mi@ksc.krasn.ru*

Аннотация. Статья описывает процесс создания системы заявок для КНЦ СО РАН с применением следующих технологий и инструментов: vue, express, node, sequelize, redis и primevue. В статье подробно рассматривается архитектура приложения «Система учета заявок для ИТ-подразделений КНЦ СО РАН», включая клиентскую и серверную части, систему кэширования данных и пользовательский интерфейс. Создание такой системы обеспечивает высокую производительность и удобство применения для пользователей.

Ключевые слова: заявки, система заявок, servicedesk, helpdesk, vue

Цитирование: Малимонов М.И. Система учета заявок для ИТ-подразделений КНЦ СО РАН / М.И. Малимонов // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 144-154. – DOI:10.25729/ESI.2024.34.2.014.

Введение. Существует множество различных систем для взаимодействия поставщика ИТ-сервисов и клиентов [1-7]. Существуют как небольшие системы для учета и обработки заявок, называемые «HelpDesk», так и огромные системы, которые позволяют не только обрабатывать заявки ИТ-сервисов, но обрабатывать заявки других отделов, вести учет оборудования, сотрудников и бизнес-процессов компании. Такие большие системы называются «ServiceDesk». Системы для учета заявок используются не только в ИТ-сегменте, но и в различных других областях, например, в телекоммуникации [8-9]. В статье все системы подобного рода будут именоваться как «система заявок». Для организации понятной работы ИТ-отдела используются документы, называемые SLA и OLA.

SLA (Service Level Agreement) – это документ, описывающий условия и требования качества обслуживания, которые должны быть выполнены поставщиком услуг для клиента. SLA определяет уровень сервиса, который должен быть предоставлен клиенту, а также сроки и способы реагирования на запросы клиента [10]. OLA (Operational Level Agreement) – это документ, который определяет условия и требования между различными подразделениями или командами внутри организации для обеспечения достижения целей SLA. OLA определяет взаимодействие между службами поддержки, чтобы обеспечить выполнение SLA [11].

Использование этих документов необходимо для прозрачности работы ИТ-структур, чтобы пользователи понимали, кто отвечает за решение их проблем, куда и к кому необходимо обращаться, в какие сроки их проблема будет решена и каковы способы получения обратной связи. Системы заявок в ИТ-области строятся на принципах SLA и OLA, но учесть всю структуру компании – весьма сложная работа, которая требует огромных временных затрат. Существующие системы не позволяют точно передать структуру организации и их внутреннюю работу, поэтому большие компании занимаются разработками своих собственных систем заявок, в зависимости от ресурсов и потребностей, в которые можно заложить необходимую функциональность и структуру организации.

Здесь очень показательным примером является функционирование подобной системы в Федеральном исследовательском центре Красноярского научного центра Сибирского отделения Российской академии наук (далее ФИЦ КНЦ СО РАН). ФИЦ КНЦ СО РАН включает 18 обособленных подразделений, часть из которых имеют свои собственные ИТ-отделы. Многие

ИТ-отделы не имеют доступа в системы другого ИТ-отдела и вынуждены, как и обычные пользователи, для решения проблем, связанных, например, с работами в 1С, использовать систему заявок. Данная система заявок должна быть построена таким образом, чтобы при создании заявки, например, «подключить рабочее место сотрудника», эта заявка попадала в нужный отдел и нужному исполнителю, в зависимости от места работы сотрудника, подающего заявку. При этом сам пользователь не должен задумываться над тем, как и куда заявка попадет, система должна сама назначить заявку, пользователю нужно лишь выбрать категорию и заполнить необходимые поля. При этом необходимо учесть то, что некоторые пользователи не должны иметь доступ к некоторым категориям. Например, только начальники отдела или руководители могут подать заявку на подключение рабочего места. С учетом всего этого поставлена цель: создать систему заявок для ФИЦ КНЦ СО РАН.

1. Существующие системы. Существует множество систем подобного рода, рассмотрим некоторые из них.

Amelia 2.0. Отечественный продукт. Эффективное управление коммерческой недвижимостью. Предоставляет модули по управлению заявок, паспортизации объектов, систематизации данных по потреблению ресурсов, отслеживанию услуг клининга, учету оборудования и т.д. Это огромная система, которая покрывает большую часть потребностей. Стоимость такой системы от пятнадцати тысяч рублей в месяц, до двух миллионов, зависит от требований заказчика. Имеет маршрутизацию заявок только по категориям работ и отделам [12].

OkDesk. Отечественный продукт. Называют свою систему «HelpDesk», хотя по размерам системы – это не так. Поддерживает множество модулей. Существует гибкий SLA, который зависит не только от категории, но и от приоритета, категории клиента, исполнителя и типа заявки, что является огромным плюсом. Есть много правил для маршрутизации. В системе есть модули для учета техники и программного обеспечения, бухгалтерии и т.д. Стоимость аренды варьируется от восьми тысяч до шестидесяти тысяч рублей в месяц [13].

ITSM365. «ServiceDesk» система. Система позволяет управлять обращениями, имеет отдельную функцию для разграничения обращений на инциденты, запросы на изменения и проблемы, через создание типов услуг, а не через построение множественного дерева категорий. Является отечественным продуктом. Стоимость от пятнадцати до шестидесяти тысяч рублей за 10 лицензий [14].

OTRS. Имеет две версии продукта: платную и бесплатную. Бесплатная версия является открытой, что позволяет доработать данную систему под себя, но не имеет множества модулей (не умеет работать с почтой), на данный момент бесплатная версия не обновляется и поддерживается только сообществом. Платная версия работает только через облако самой компании, имеет множество отдельных модулей и постоянно обновляется [15].

OsTicket. Зарубежный продукт. Существует несколько версий продукта, в том числе и бесплатная версия. Функциональность направлена на работу с заявками и базами знаний. Система проста в обращении. Поддерживает SLA, но заявка попадает к исполнителю в зависимости от категории. На данный момент используется в ФИЦ КНЦ СО РАН. Поддерживает установку на собственные серверы. Системы, описанные выше, работают в основном через облако [16].

HubEx. Данная система представляет собой «ServiceDesk» и FSM-систему. FSM – система управления мобильными сотрудниками. Включает такие функции, как: учет и управление заявками и объектами, планирование расписаний, учет выездов и работ, автоматизация диспетчерской, электронный паспорт оборудования, GPS-контроль, модуль аналитики и отчетности. Отечественный продукт. Стоимость от тринадцати тысяч рублей [17].

2. Материалы и методы. Для создания системы заявок для организации ФИЦ КНЦ СО РАН были использованы следующие инструменты:

1. Vue – фреймворк для разработки пользовательских интерфейсов [18-21].
2. Vue-router – официальная библиотека для маршрутизации в SPA-приложении.
3. Vuex – библиотека управления состоянием приложения.
4. Express.js – веб-фреймворк для Node.js, который обеспечивает функциональность сервера.
5. Node.js – среда выполнения JavaScript, используемая для создания серверной части приложения [22].
6. Sequelize – ORM (Object-Relational Mapping) для работы с базами данных, позволяет создавать модели и обращаться к базе данных без знания языка SQL.
7. Redis – высокопроизводительная система управления базами данных, используемая для кэширования данных, чтобы ускорить их выдачу, без обращения к основной базе данных.
8. Primevue – библиотека UI элементов для создания пользовательских интерфейсов. Поддерживается фреймворком Vue.
9. PostgreSQL – свободная объектно-реляционная база данных [23].

В системе существуют следующие сущности:

1. Область работ – сущность для группировки категорий. Представляет собой дерево, где на узлах могут быть установлены категории. Данная сущность хранит структуру дерева в формате JSON.
2. Категории – это вид работы или проблемы, для которых требуется ответ или выполнение задачи.
3. Папки – фильтры заявок, созданные пользователями. Данная сущность хранит правила для отбора заявок в формате JSON и редактируется через меню «Папки».
4. Чат – массив сообщений от пользователей и исполнителей, которые прикреплены к конкретной заявке. Также здесь размещены основные элементы управления заявкой: передать другому сотруднику, сменить статус заявки, заявка закрыта или отклонить заявку.
5. Организации – обособленные подразделения и головная организация. Необходимы для составления маршрутизации заявок в зависимости от подразделения.
6. Приоритеты – первенство по времени выполнения заявки. В зависимости от выставленного значения, заявка будет помечена определенным цветом в системе. Также доступна сортировка заявок по приоритетам.
7. Роли – набор полномочий, который необходим пользователю для выполнения рабочих задач и ограничивает доступ к частям системы.
8. Статусы – состояние заявки.
9. Шаги выполнения – составная часть «Пути выполнения». Содержит перечисление сотрудников с описанием выполняемых работ.
10. Системные настройки – набор настроек для старта системы. Содержит поля для настройки статусов при открытии и закрытии заявки. Также содержит допустимые почтовые сервисы, с которых можно регистрироваться в системе.
11. Заявки – обращения пользователей.
12. Пользователи – учетные записи в системе, которые имеют свои настройки уровня доступа для выполнения какой-либо работы.
13. Пути выполнения – маршрут выполнения заявки.

На рисунке 1 представлена структура клиентской части приложения.

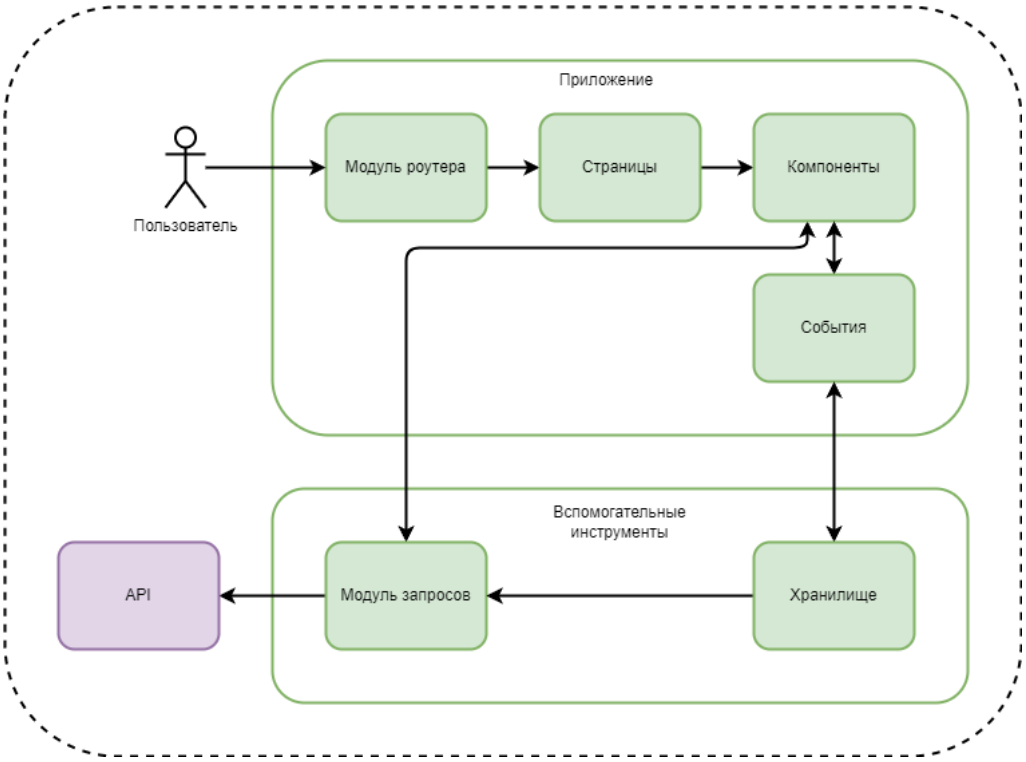


Рис. 1. Структура клиентской части приложения

Данные о пользователя хранятся в JWT. JWT – это JSON объект, используется для передачи данных для авторизации в приложении. Состоит из трех частей – заголовок, тело или полезная нагрузка, и подпись. В теле JWT как раз и хранится вся информация о пользователе. Данная информация передается между клиентом и сервером, когда пользователь заходит в систему, при этом пользователь получает ключ от сервера и ключ хранится в клиентском приложении. При выполнении запросов этот ключ вставляется в заголовок запроса, сервер проверяет заголовок, если заголовок пустой, то значит, пользователь не авторизован, если заголовок не пустой, то сервер выявляет роль пользователя и проверяет, имеет ли пользователь доступ к маршруту для выполнения запроса. Подпись устанавливается в конфигурации сервера, чтобы ключ не могли подделать. Ключ выдается на 8 часов, время можно изменить; при истечении срока система “выбрасывает” пользователя и необходимо заново выполнить вход в систему.

Пользователь при запуске приложения взаимодействует с модулем маршрутизации. Если пользователь не авторизован, то приложение автоматически перенаправляет пользователя на страницу входа. Если авторизован, то загружается та страница, на которую пытается зайти пользователь. Также на некоторых страницах существует уровень доступа, и если пользователь не имеет необходимых прав, то приложение перенаправит пользователя на главную страницу и выдаст ошибку об этом. Каждая страница имеет метаданные, где указывается шаблон страницы, уровень доступа, тип страницы и наименование. Шаблон отвечает за то, какие части страницы будут статистическими, а какие – динамическими. Например, шапка сайта – статическая часть страницы, она должна отображаться везде, а форма для заполнения пользователя или таблица всех пользователей – динамические части страницы и зависят от того, на какой мы странице находимся. Уровень доступа отвечает за то, кто имеет доступ к странице. Тип страницы имеет три значения: создание, изменение и отображение. От этих настроек зависят тип HTTP-запроса и некоторые отображаемые элементы. Например, кнопка «Создать» появится только на странице с типом «Создание», а кнопки удаления и сохранения – при типе

страницы «Изменение». Тип «Отображение» стоит на тех страницах, на которых информация служит только для отображения и ее нельзя изменить.

На рисунке 2 изображена структура серверной части приложения.

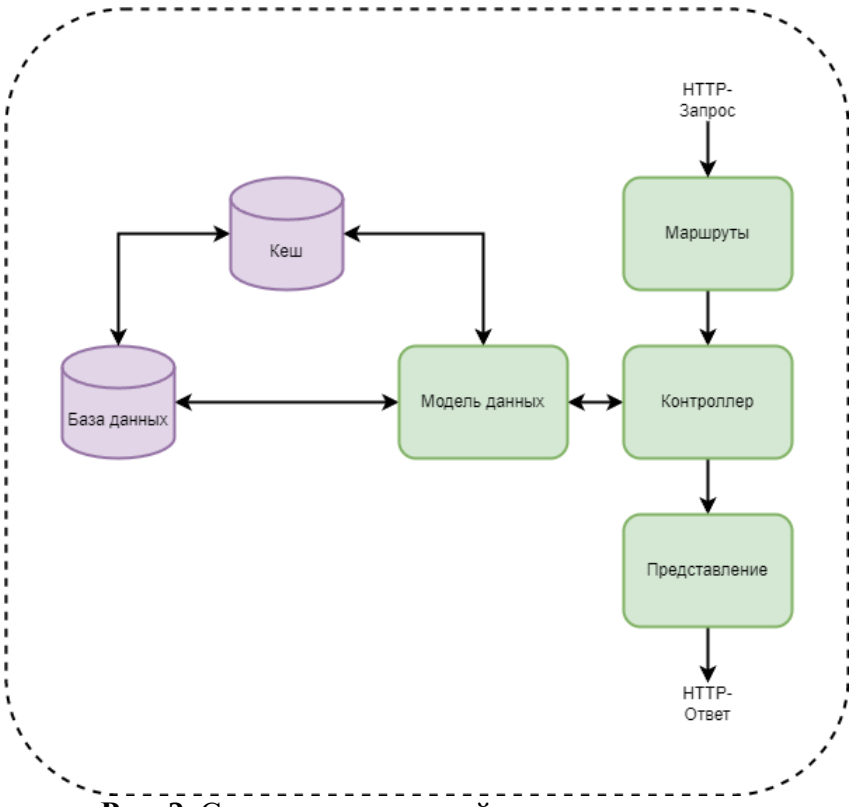


Рис. 2. Структура серверной части приложения

Далее на вход поступает HTTP-запрос. В зависимости от типа запроса и маршрута контроллер выполнит определенное действие. Некоторые данные кэшируются, что ускоряет ответ сервера. Кеширование данных необходимо, потому что на данный момент в организации ФИЦ КНЦ СО РАН работают две с половиной тысячи сотрудников. Основным действием пользователя является просмотр данных, что вызывает обращения к базе данных. При редактировании и создании различных сущностей в системе необходима дополнительная информация, что также требует обращения к базе данных. Кеширование ускоряет процесс выдачи данных для просмотра и редактирования элементов системы. У сервера предусмотрен механизм обновления кеша при обновлении сущностей системы.

Модели данных составлены с помощью библиотеки sequelize, которая позволяет создавать модели данных, указывать, какой тип у поля будет в базе данных и задавать валидацию данных. Также у маршрутов присутствует связующее программное обеспечение. Перед передачей запроса в контроллер вызывается модуль проверки пользователя, где проверяется его роль и авторизован ли пользователь. Если пользователь не авторизован, то выдается сообщение об ошибке. Если авторизован, то сервер проверяет роль пользователя. Некоторые маршруты имеют уровень доступа. Если уровень доступа не соответствует, то система выводит сообщение об ошибке.

3. Структура базы данных. Для отображения структуры базы данных она была разбита на три части: заявки, пользователи и пути. На рисунках 3-5 показана вся структура базы данных. Напротив поля указан тип данных: t – текст; b – булево значение; d – дата и время; # – целочисленный тип; [] – массив.

Если тип не указан, то это тип данных json или jsonb. На рисунке 3 отображен фрагмент базы данных (БД) «Заявки».

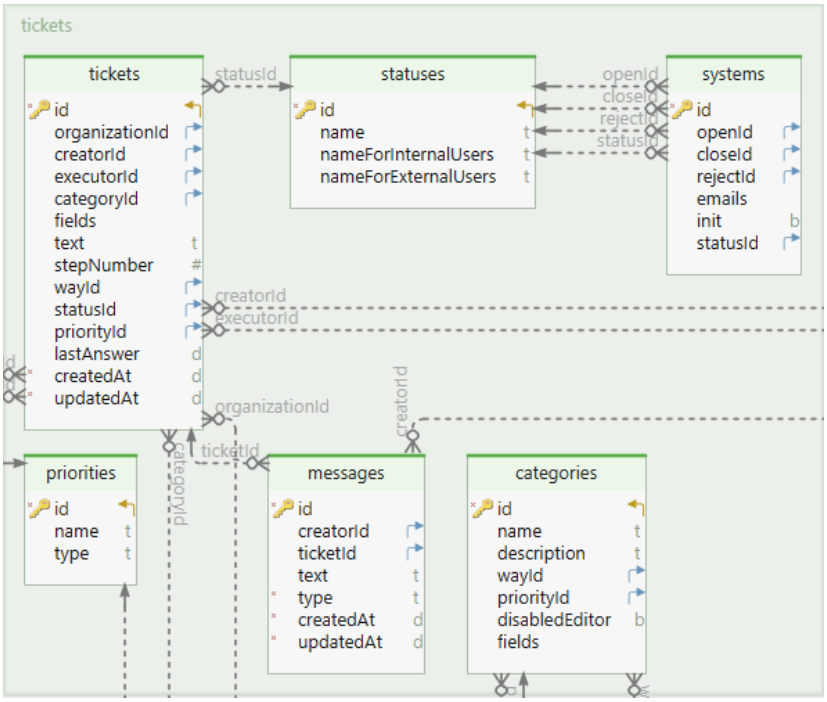


Рис. 3. Фрагмент базы данных «Заявки»

Этот фрагмент включает такие таблицы, как: заявки, статусы, система, приоритеты, сообщения и категории, т.е. содержит основные элементы, связанные с сущностью «Заявка»

На рисунке 4 представлен фрагмент БД «Пользователи».

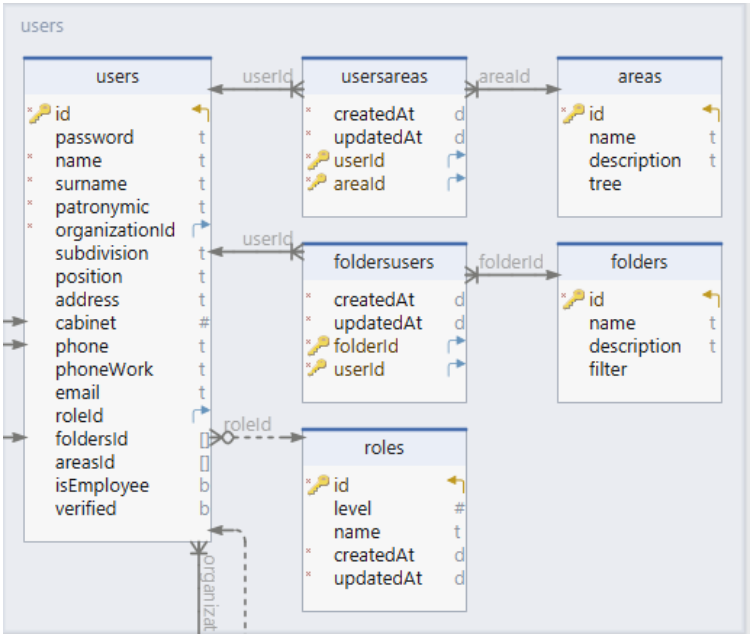


Рис. 4. Фрагмент базы данных «Пользователи»

Этот фрагмент включает таблицы: пользователи, области работ, папки, и роли, т.е. содержит основные элементы, относящиеся к пользователям.

На рисунке 5 представлен фрагмент БД «Пути».

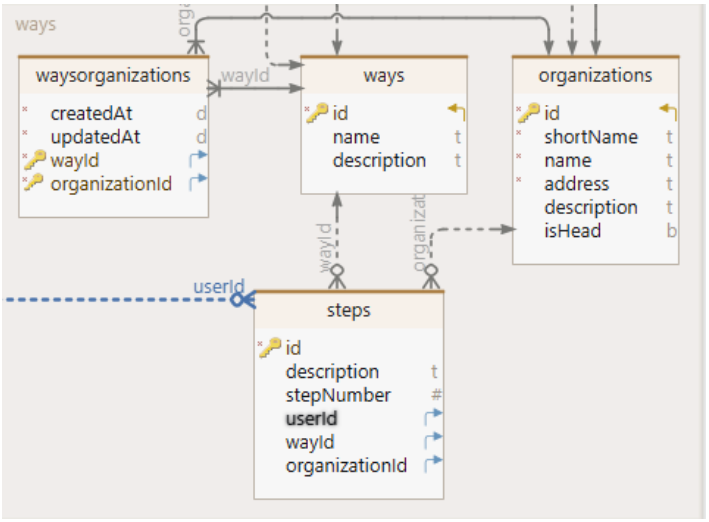


Рис. 5. Фрагмент базы данных «Пути»

Этот фрагмент включает таблицы: пути, организации и шаги, т.е. содержит основные элементы, связанные с сущностью «Путь».

4. Работа системы. Основной задачей системы является маршрутизация заявок в зависимости от организации и категории. Все маршруты заполняются в объекте «Путь», указывается организация, для этой организации перечисляются сотрудники в порядке выполнения работ, затем созданный путь прикрепляется к категории.

На рисунке 6 представлена модель маршрутизации заявки.

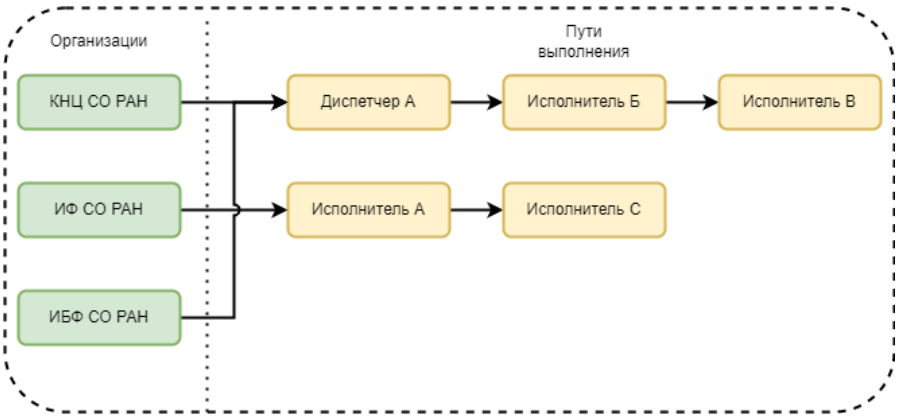


Рис. 6. Маршрутизация заявки с учетом структуры организации

В левой части рисунка перечислены некоторые обособленные подразделения ФИЦ КНЦ СО РАН. Например, необходимо подключить рабочее место работника. Выбирается категория «Подключить рабочее место». В зависимости от того, к какой организации принадлежит заявитель, заявка попадет разным исполнителям:

- если заявитель находится в КНЦ СО РАН, то заявка попадет диспетчеру «А», затем исполнителю «Б», затем исполнителю «В»;
- если заявитель находится в ИФ СО РАН, то заявка попадет исполнителю «А», затем исполнителю «С»;
- если заявитель из ИБФ СО РАН и для этой организации не настроен маршрут выполнения заявки, то заявка пойдет по маршруту головной организации – КНЦ СО РАН. Естественно, головную организацию можно указать в системе.

На рисунке 7 представлен пример, когда исполнитель «Б» не может выполнить свою часть работы по каким-либо причинам.

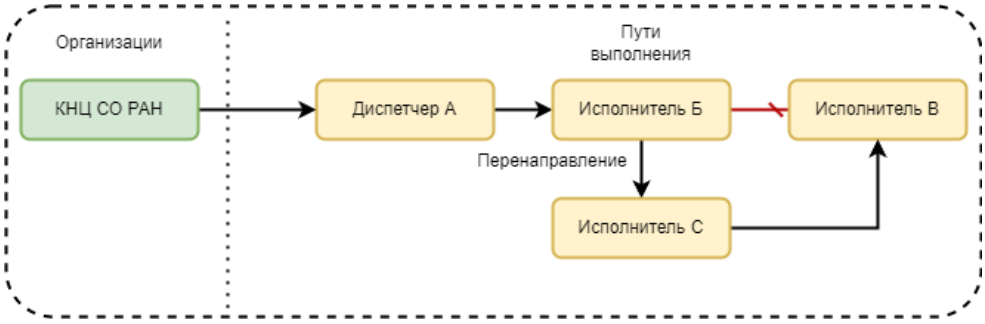


Рис. 7. Маршрутизация заявки с перенаправлением на другого исполнителя

В таком случае исполнитель «Б» перенаправляет свою заявку на исполнителя «С», который выполнит работу и направит следующему исполнителю «В» по маршруту заявки; это перенаправление будет выполнено автоматически, в зависимости от настроенного маршрута заявки. В системе заявка всегда двигается по составленному маршруту. При составлении маршрута указываются необходимые шаги для выполнения этой заявки. К шагу прикрепляется описание работ и пользователь (userId), который является исполнителем (isEmployee). Данный исполнитель меняется автоматически, следуя составленному маршруту, либо вручную. Если требуется переназначить заявку другому сотруднику, то для этого есть инструменты внутри самой заявки, можно выбрать абсолютно любого исполнителя через меню чата. Можно указать исполнителя из любой организации или подразделения, независимо от категории, составленного маршрута заявки и организации заявителя.

На рисунке 8 представлен пример созданной заявки. Слева расположена боковая панель с навигацией и папки. В нижней части меню управления заявкой: открыть чат, передать предыдущему исполнителю, передать следующему, отклонить заявку и выполнить заявку.

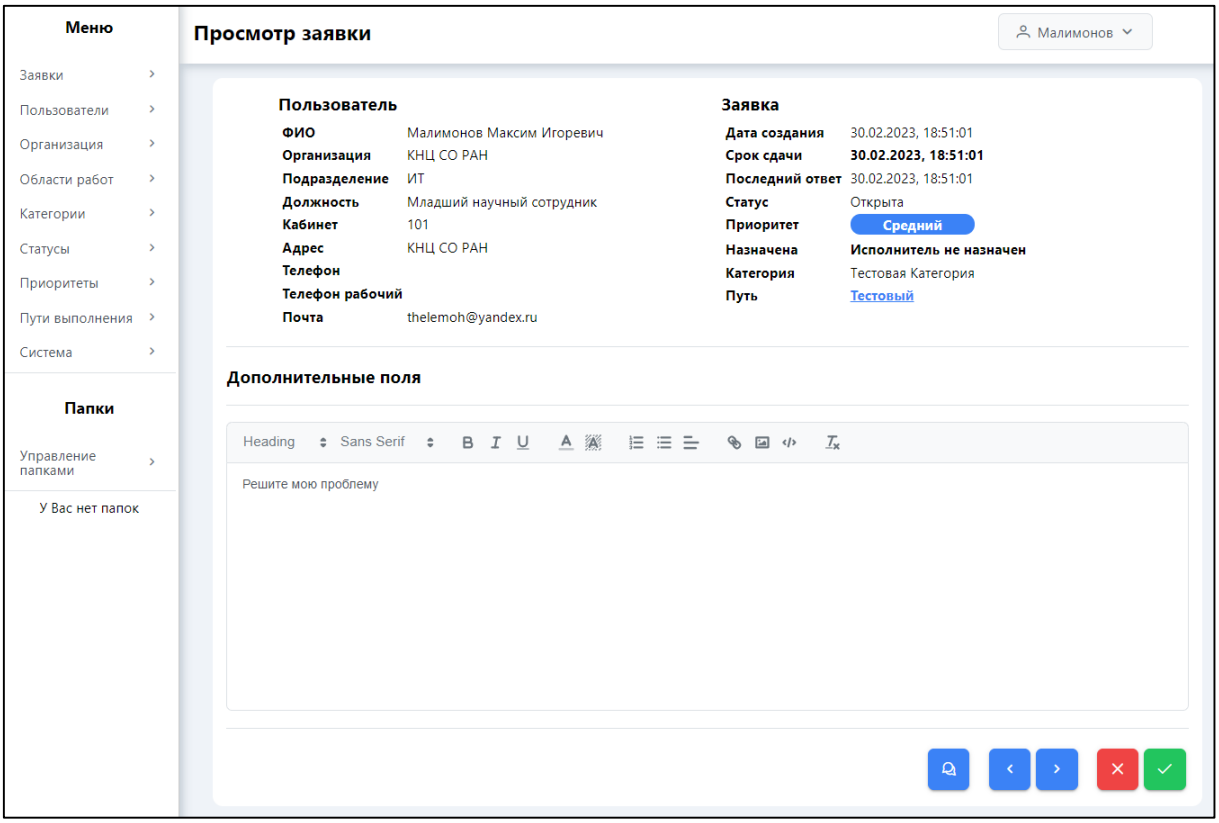


Рис. 8. Визуализация формы заявки

Заявитель может отслеживать выполнение заявки и видеть, к кому попадет заявка на всем пути следования. Также для каждого шага на маршруте описывается работа исполнителя и заявитель знает, что должен сделать исполнитель.

Таким образом, можно обеспечить прозрачную и понятную работу ИТ-отделов, а не представлять ее в виде черного ящика.

Заключение. В статье описана разработка системы заявок для организации ФИЦ КНЦ СО РАН. Следует отметить, что внедрение системы учета заявок с маршрутизацией по организациям для ФИЦ КНЦ СО РАН является значимым событием. Благодаря новой системе процесс обработки заявок станет более прозрачным, быстрым и удобным для всех участников. Кроме того, система позволит более точно контролировать выполнение задач и улучшить взаимодействие между различными подразделениями ФИЦ КНЦ СО РАН. В целом, внедрение системы учета заявок с маршрутизацией по организациям является важным шагом в совершенствовании работы организации и повышении ее эффективности.

Список источников

1. Al-Hawari F., Barham H. A machine learning based help desk system for IT service management. Journal of King Saud university – Computer and information sciences, 2021, no. 33, pp. 702-718.
2. Serdar Serbest, Yilmaz Goksen, Onur Doğan, Anıl Tokdemir Design and implementation of Help Desk system on the effective focus of information system. Procedia economics and finance, 2015, no. 33, pp. 461-467.
3. Akinnuwesi B.A., Enikuomehin O.A., Uzoka, F.E., et al. Electronic helpdesk support system in tertiary institutions in developing countries. Int. J. Comput. Inf. Technol, 2014, 3 (6), pp. 1-12.
4. Wongsakthawom R., Limpiyakorn Y. Development of IT helpdesk with microservices. In: IEEE 8th International conference on electronics information and emergency communication (ICEIEC), 2018, pp. 31-34.
5. Zhou Y., Tong Y., Gu R., Gall H. Combining text mining and data mining for bug report classification. J. Software, Evol. Process, 2016, no. 28 (3), pp. 150-176.
6. Zhu X., Zhou W., Li T. A visual tool for ticket monitoring and management. In: 2017 IEEE 12th International conference on intelligent systems and knowledge engineering (ISKE), 2017, 2016, 107 p.
7. Serbest S., Goksen Y., Dogan O., Tokdemir A. Design and implementation of help desk system on the effective focus of information system. Procedia econics and finance, 2015, 33, pp. 461-467.
8. Tanovic A., Mastorakis N.E. Advantage of using service desk management systems in real organizations. Int. J. Econ. Manage. Syst, 2016, 1, pp. 81-86.
9. Tanovic A., Orucevic F., Butkovic A. Advantages of the implementation of service desk based on ITIL framework in telecommunication industry. Math. Comput. Sci. Ind, 2016. 1, pp. 244–258.
10. Adrian Paschke, Martin Bichler Knowledge representation concepts for automated SLA management. Decision support systems, 2008. no. 46, pp. 187-205
11. Maria-Cruz Valiente, Elena Garcia-Barriocanal, Miguel-Angel Sicilia. Applying an ontology approach to IT service management for business-IT integration. Knowledge-Based systems, 2012, no. 28, pp. 76-87.
12. Amelia 2.0. Цифровая эксплуатация недвижимости. – URL: https://info.amelia-st.ru/ (дата обращения: 10.06.2023).
13. Okdesk. Help desk система учета заявок. – URL: https://okdesk.ru/ (дата обращения: 10.06.2023).
14. ITSM365. Help desk система учета заявок. – URL: https://itsm365.com/ (дата обращения: 12.02.2024).
15. OTRS. Software solutions for customer service, ITSM, ISMS and cyber defense, available at: https://otrs.com/. (accessed:12.02.2024).
16. osTicket. Support ticketing system. Available at: https://osticket.com/ (accessed:10.06.2023).
17. Hubex. Управление мобильными сотрудниками – FSM HubEx. – URL: https://hubex.ru/ (дата обращения: 12.02.2024).
18. Malimonov M.I. Yakubailik O.E. Cloud-based software tools for the rapid development of web gis, CEUR workshop proceedings, 2021, no. 3006, pp. 172-179.
19. Shiyong Xiong, Xue Wang, Zhengpeng Lan. Model research of visual report components. Procedia computer science, 2022, no. 208, pp. 478-485.
20. Nian L, Bo Z. The research on single page application front-end development based on vue. Journal of physics: conference series, 2021, no. 1883(1).
21. Qiu Bin Liu, Sheng Gao, Bi Feng Guo, Hui Liu, Yong Zhao Feng, Hao Xia. Research and development of movie social system. Procedia computer science, 2020, no. 166, pp. 154-159.

22. Fabian Kaimer, Philipp Brune Return of the JS: Towards a Node.js-Based software architecture for combined CMS/CRM applications. Procedia computer science, 2018, no. 141, pp. 454-459.
23. Massimiliano Pirani, Alessandro Cucchiarelli, Luca Spalazzi Paradigms for database-centric application interfaces. Procedia computer science. 2023, no. 217, pp. 835-845.

Малимонов Максим Игоревич. *Младший научный сотрудник Красноярского научного центра СО РАН. Основное направление – программирование. SPIN: 7645-6804, malimonov.mi@ksc.krasn.ru, 660036, г. Красноярск, ул. Академгородок, 50.*

UDC 004.415.25
DOI:10.25729/ESI.2024.34.2.014

Ticketing system for IT-department of KSC SB RAS

Maksim I. Malimonov

Federal researcher center “Krasnoyarsk scientific center SB RAS”,
Russia, Krasnoyarsk, *malimonov.mi@ksc.krasn.ru*

Abstract. This article describes the process of creating an application system for the KSC SB RAS using the following technologies and tools: vue, express, node, sequelize, redis and primevue. The article discusses in detail the architecture of the application “Application accounting system for IT departments of the KSC SB RAS”, including the client and server parts, data caching system and user interface. The creation of such a system ensures high performance and ease of use for users.

Keywords: ticket, ticket system, servicedesk, helpdesk, vue

References

1. Al-Hawari F., Barham H. A machine learning based help desk system for IT service management. Journal of King Saud university – Computer and information sciences, 2021, no. 33, pp. 702-718.
2. Serdar Serbest, Yilmaz Goksen, Onur Doğan, Anıl Tokdemir Design and implementation of Help Desk system on the effective focus of information system. Procedia economics and finance, 2015, no. 33, pp. 461-467.
3. Akinnuwesi B.A., Enikuomehin O.A., Uzoka, F.E., et al. Electronic helpdesk support system in tertiary institutions in developing countries. Int. J. Comput. Inf. Technol, 2014, 3 (6), pp. 1-12.
4. Wongsakthawom R., Limpiyakorn Y. Development of IT helpdesk with microservices. In: IEEE 8th International conference on electronics information and emergency communication (ICEIEC), 2018, pp. 31-34.
5. Zhou Y., Tong Y., Gu R., Gall H. Combining text mining and data mining for bug report classification. J. Software, Evol. Process, 2016, no. 28 (3), pp. 150-176.
6. Zhu X., Zhou W., Li T. A visual tool for ticket monitoring and management. In: 2017 IEEE 12th International conference on intelligent systems and knowledge engineering (ISKE), 2017, 2016, 107 p.
7. Serbest S., Goksen Y., Dogan O., Tokdemir A. Design and implementation of help desk system on the effective focus of information system. Procedia econics and finance, 2015, 33, pp. 461-467.
8. Tanovic A., Mastorakis N.E. Advantage of using service desk management systems in real organizations. Int. J. Econ. Manage. Syst, 2016, 1, pp. 81-86.
9. Tanovic A., Orucevic F., Butkovic A. Advantages of the implementation of service desk based on ITIL framework in telecommunication industry. Math. Comput. Sci. Ind, 2016. 1, pp. 244–258.
10. Adrian Paschke, Martin Bichler Knowledge representation concepts for automated SLA management. Decision support systems, 2008. no. 46, pp. 187-205
11. Maria-Cruz Valiente, Elena Garcia-Barriocanal, Miguel-Angel Sicilia. Applying an ontology approach to IT service management for business-IT integration. Knowledge-Based systems, 2012, no. 28, pp. 76-87.
12. Amelia 2.0. Tsifrovaya ekspluatatsiya nedvizhimosti [Amelia 2.0. Digital operation of buildings]. Available at: https://info.amelia-st.ru/(accessed: 06/10/2023).
13. Okdesk. Help desk sistema ucheta zayavok [Okdesk. Help desk ticket system]. Available at: https://okdesk.ru/ (accessed: 06/10/2023).
14. ITSM365. Help desk sistema ucheta zayavo [ITSM365. Help desk ticket system]. Available at: https://itsm365.com/ (accessed: 02/12/2024).

15. OTRS. Software solutions for customer service, ITSM, ISMS and cyber defense, available at: https://otrs.com/. (accessed:12.02.2024).
16. osTicket. Support ticketing system. Available at: https://osticket.com/ (accessed:10.06.2023).
17. Hubex. Upravleniye mobil'nymi sotrudnikami – FSM HubEx [Hubex. Mobile employee management – FSM HubEx]. Available at: https://hubex.ru/ (accessed: 02/12/2024)
18. Malimonov M.I. Yakubailik O.E. Cloud-based software tools for the rapid development of web gis, CEUR workshop proceedings, 2021, no. 3006, pp. 172-179.
19. Shiyong Xiong, Xue Wang, Zhengpeng Lan. Model research of visual report components. Procedia computer science, 2022, no. 208, pp. 478-485.
20. Nian L, Bo Z. The research on single page application front-end development based on vue. Journal of physics: conference series, 2021, no. 1883(1).
21. Qiu Bin Liu, Sheng Gao, Bi Feng Guo, Hui Liu, Yong Zhao Feng, Hao Xia. Research and development of movie social system. Procedia computer science, 2020, no. 166, pp. 154-159.
22. Fabian Kaimer, Philipp Brune Return of the JS: Towards a Node.js-Based software architecture for combined CMS/CRM applications. Procedia computer science, 2018, no. 141, pp. 454-459.
23. Massimiliano Pirani, Alessandro Cucchiarelli, Luca Spalazzi Paradigms for database-centric application interfaces. Procedia computer science. 2023, no. 217, pp. 835-845.

***Malimonov Maxim Igorevich**. Junior researcher, Krasnoyarsk scientific center SB RAS. The main direction is programming. SPIN: 7645-6804, malimonov.mi@ksc.krasn.ru, 660036, Krasnoyarsk, st. Akademgorodok, 50.*

Статья поступила в редакцию 21.06.2023; одобрена после рецензирования 02.04.2024; принята к публикации 30.05.2024.

The article was submitted 06/21/2023; approved after reviewing 04/02/2024; accepted for publication 05/30/2024.

Программные системы и комплексы

УДК 004.421+004.4’2+004.771
DOI:10.25729/ESI.2024.34.2.015

Инструментальные средства автоматизации разработки и применения пакета прикладных микросервисов

**Опарин Геннадий Анатольевич, Богданова Вера Геннадьевна,
Пашинин Антон Алексеевич**

Институт динамики систем и теории управления им. В.М. Матросова СО РАН,
Россия, Иркутск, *apcrol@gmail.com*

Аннотация. Представлены результаты, связанные с применением разработанного ранее метода булевых ограничений для исследования динамики и параметрического синтеза систем управления, в частности, булевых сетей, в микросервисной инфраструктуре с использованием методов и средств построения баз знаний. В качестве базы знаний выступает вычислительная модель предметной области, под которой понимается совокупность сведений об объектах предметной области и отношениях между ними. Каждое функциональное отношение реализуется вычислительным микросервисом, созданным на основе программного модуля, вычисляющего значения выходных параметров по заданным значениям входных параметров. Набор вычислительных микросервисов составляет функциональное наполнение пакета прикладных микросервисов. Системное наполнение составляют сервисы постановки задачи, планирования и мониторинга вычислений, управления данными, обработки результатов. Управление вычислениями (децентрализованное, централизованное или иерархическое) осуществляется агентами пакета прикладных микросервисов в зависимости от постановки задачи и способа организации базы знаний. В связи с переходом к исследованию новых классов булевых сетей (управляемых, сингулярных, неявных, асинхронных, двудольных и других) выполняется расширение созданного ранее на основе вышеизложенного подхода пакета прикладных микросервисов, предназначавшегося для задач качественного исследования автономных булевых сетей. Целью исследования является разработка нового инструментария микросервисной интеллектуальной вычислительной платформы HPCSOMAS-MSC для создания, настройки и доступа к вычислительным микросервисам, позволяющего взаимодействовать с микросервисами функционального наполнения напрямую через веб-интерфейс. Реализован веб-интерфейс описания предметной области в дополнение к существующему описанию на языке JSON. Новый инструментарий позволяет как ускорить разработку и отладку микросервисов, так и сделать дальнейшие взаимодействие более удобным для пользователя-предметника.

Ключевые слова: булевы сети, метод булевых ограничений, качественный анализ, пакет прикладных микросервисов

Цитирование: Опарин Г.А. Инструментальные средства автоматизации разработки и применения пакета прикладных микросервисов / Г.А. Опарин, В.Г. Богданова, А.А. Пашинин // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 155-168. – DOI:10.25729/ESI.2024.34.2.015.

Введение. Спектр приложения двоичных динамических моделей необычайно широк, и с каждым годом количество объектов и задач, где требуется их использование, только возрастает. К современным приложениям двоичных динамических моделей относятся задачи искусственного интеллекта, биоинформатики, экологии, социологии, робототехники, криптографии, классификации, исследования отказоустойчивости компьютерных сетей, а также разнообразные задачи во многих других предметных областях. В этой связи актуальной является проблема создания эффективной технологии качественного исследования динамических свойств различных классов двоичных моделей высокого порядка на конечном интервале времени. Основой такой технологии является разработанный авторами метод булевых ограничений (Boolean Constraint Method, BCM) [1], позволяющий свести исходную задачу к задаче булевой выполнимости (SAT) или к задаче проверки истинности квантифицированной булевой формулы (QSAT). Обе задачи являются вычислительно

трудоемкими и требуют для своего решения дальнейшего развития микросервисной интеллектуальной вычислительной платформы HPCSOMAS-MSC [2, 3], обеспечивающей глубокое распараллеливание задачи и ориентированной на ее применение конечным пользователем в гибридной микросервисной вычислительной инфраструктуре (ГМВИ).

Платформа HPCSOMAS-MSC предоставляет следующие базовые средства:

- Программные средства, включающие библиотеку классов для создания агентов и микросервисов.
- Средства мультиагентного управления, включающие стандартный набор конфигурируемых системных агентов, который используется при настройке свойств уже скомпилированных агентов на конкретный вычислительный ресурс. В этот набор входят пользовательский агент, серверная программа-агент (менеджер-агент), вычислительные агенты (реализованные в виде микросервисов) и специализированные агенты организации двухуровневого параллельного выполнения вычислений [3].
- Инструментальные средства, включающие утилиты ABCSW (Agent Based Class Service Wizard) для автоматизации процессов создания агентов программным путем, конфигурирования агентов и MSCDT (Micro-Services Creating Deployment Testing) с набором мастеров для создания микросервисов на основе библиотеки классов, их развертывания и тестирования, а также подготовки вычислительных узлов.
- Языковые средства спецификации объектов ПО в текстовом формате JSON.

Основной целью работы является дальнейшее развитие инструментальных средств HPCSOMAS-MSC, автоматизирующих процессы создания и применения пакета прикладных микросервисов (applied microservices package, AMP, [2]) для исследования динамики и параметрического синтеза разнообразных систем управления, в том числе булевых сетей (двоичных динамических систем (ДДС)), в гибридной облачной среде с использованием методов и средств построения баз знаний. В качестве базы знаний (БЗ) выступает вычислительная модель предметной области (ПО), состав объектов которой и множество отношений между ними детально приводятся в [4]. Каждое функциональное отношение реализуется вычислительным микросервисом, созданным на основе программного модуля, вычисляющего значения выходных параметров по заданным значениям входных параметров. Набор вычислительных микросервисов составляет функциональное наполнение AMP. В состав системного наполнения входят сервисы постановки задачи, планирования и мониторинга вычислений, управления данными, обработки результатов. Управление вычислениями (децентрализованное, централизованное или иерархическое) осуществляется агентами AMP в зависимости от постановки задачи и способа организации базы знаний.

1. Краткий обзор методов и средств качественного исследования ДДС. Кроме ВСМ, к ДДС применимы и другие, также имеющие переборный характер, методы качественного анализа. В [1] рассматриваются достоинства и недостатки дедуктивного анализа [5], проверки модели [6], метода редукции [7, 8] и приводятся преимущества ВСМ, при использовании которого проблема сокращения перебора решается путем применения современных эффективных решателей SAT и QSAT задач. В [9] рассматриваются особенности качественного исследования свойства наблюдаемости ДДС с помощью метода полутензорного произведения [10], главным недостатком которого является необходимость оперировать с матрицами размерности $2^n \times 2^n$ (n – размерность вектора состояния булевой системы) при его использовании [11]. Публикации, посвященные вопросам распараллеливания этого метода, отсутствуют. В [12] приведена оценка допустимого значения размерности n (n<25). В работе [13] предложен подход к исследованию этого свойства, основанный на идее представления ДДС в полиномиальной форме. Как отмечено в [13],

используемое в этом методе вычисление базиса Грёбнера [14] приводит в общем случае к двойной экспоненциальной сложности.

Существующие в настоящее время инструментальные средства качественного анализа двоичных динамических моделей не обеспечивают требуемой глубины параллелизма, а также достаточной степени автоматизации исследования динамических моделей в ГМВИ. Так, веб-инструмент Analysis of Dynamic Algebraic Models (ADAM) [15] автоматически преобразует ряд типов дискретных моделей в полиномиальные динамические системы и анализирует их динамику, используя инструменты компьютерной алгебры. Интерактивный блокнот CoLoMoTo [16] предоставляет веб-интерфейс доступа к ряду консольных программных средств для качественного исследования моделей биологических сетей. Кроссплатформенный инструмент BoolSi, использующий высокопроизводительные вычисления для распределенного моделирования синхронных булевых сетей, является консольным приложением [17]. Сервис-ориентированные приложения, разработанные на основе платформы HPCSOMAS-MSC, предназначены для работы в высокопроизводительной гибридной среде, являются кроссплатформенными, имеют веб- и API- интерфейсы, а мультиагентное управление обеспечивает возможность проведения распределенных параллельных и конвейерно-параллельных вычислений. В состав узлов гибридной вычислительной среды могут входить как облачные, так и локальные ресурсы, в том числе кластеры, выделенные серверы, персональные компьютеры. В отличие от перечисленных выше средств [15-17], для исследования различных классов булевых сетей используется декларативный подход на основе BCM, позволяющий за счет естественного параллелизма по данным (путем расщепления булевой модели) существенно увеличить размерность вектора состояния булевой сети и промежуток времени ее функционирования при решении задач качественного анализа в высокопроизводительной вычислительной среде.

2. Новые компоненты HPCSOMAS-MSC. Научные исследования, в которых применяется предлагаемая технология создания АМР, проводятся нами в нескольких направлениях. Первое направление связано с построением области требуемой динамики в пространстве параметров регулятора замкнутой системы управления и задачами параметрического синтеза регулятора для различных классов линейных динамических объектов. Второе направление относится к качественному анализу ДДС на основе метода булевых ограничений. Третье направление связано с параметрическим синтезом обратной стабилизирующей связи для ДДС на основе логического подхода. На предыдущих этапах исследований для решения задач в рамках двух последних направлений был разработан АМР BCM-QABDS (BCM-based Qualitative Analysis of Boolean Dynamic Systems) [18], предназначенный для качественного анализа автономных синхронных ДДС. В настоящее время выполняется расширение этого пакета вычислительными микросервисами (обозначаемыми далее для краткости как CMS – Computational Micro-Service) для качественного анализа и параметрического синтеза новых классов булевых сетей (управляемых, сингулярных, неявных, асинхронных, двудольных и других). Соответственно расширяется и набор системных сервисов АМР, предоставляемых новыми компонентами инструментальных средств платформы HPCSOMAS-MSC, а именно, мастером спецификации ПО (рис. 1) и мастером композиции CMS (рис. 2).

2.1. Мастер спецификации ПО. Мастер спецификации ПО предоставляет следующие средства:

– системные веб-сервисы (в дополнение к вышеупомянутым языковым средствам) для содержательного описания объектов ПО, в частности параметров и микросервисов, с последующим сохранением этого описания в формате JSON в базе знаний АМР;

– системный веб-сервис для конфигурирования автономного режима работы CMS.

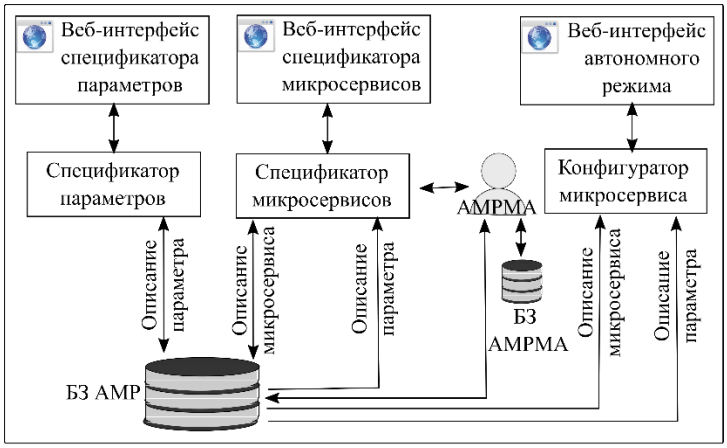


Рис. 1. Мастер спецификации ПО

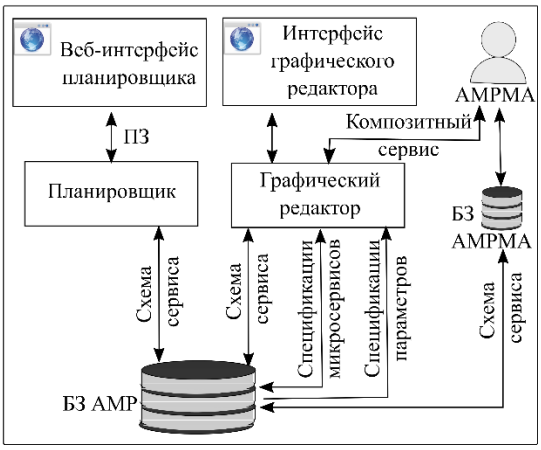


Рис. 2. Мастер композиции CMS

Параметрами ПО являются выделенные уникальные, возможно структурированные, переменные математической модели и алгоритмов ее исследования, посредством которых реализуются связи между вычислительными микросервисами. Спецификатор параметров автоматически поддерживает уникальность их ключевых имен, задаваемых разработчиком AMP. В перечень спецификаций параметра входят также тип параметра, его содержательное описание, и название разрабатываемого AMP, к которому он относится.

Интерфейс спецификатора микросервисов для разработчика (или администратора) приведен на рис. 3. В этом интерфейсе на вкладке «Main settings» вводятся ключевое имя, содержательное описание и базовые настройки вычислительного микросервиса, в частности, привязка к менеджеру-агенту AMP (AMPMA), разрешения на внешний доступ, пути к папкам. Набор входных и выходных параметров CMS указывается на вкладке «IO parameters». Для каждого параметра указывается ключевое имя («Unique id») из словаря параметров БЗ AMP, имя-идентификатор («Programmed id») из программного кода микросервиса, имя AMP («Package»), к которому относится параметр, тип параметра («Type») и содержательное описание параметра («Description»).

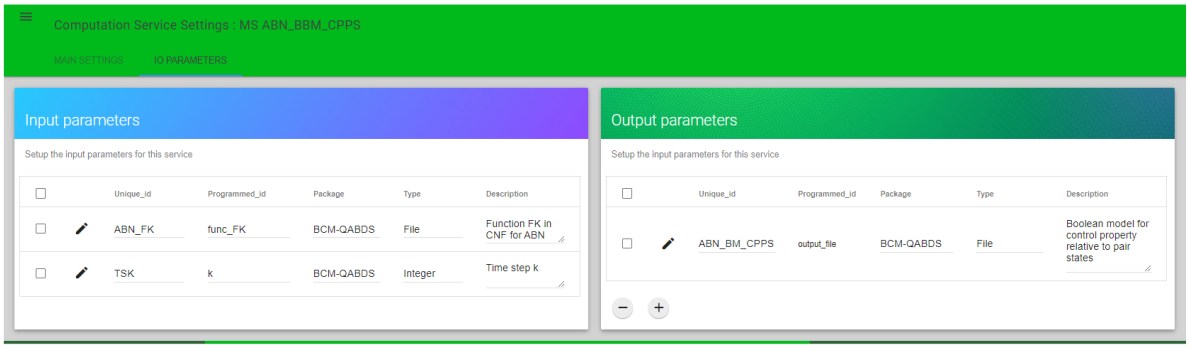


Рис. 3. Веб-интерфейс для ввода спецификаций микросервисов ПО

Интерфейс автономного режима работы микросервиса используется либо на этапе предварительного автономного тестирования уже разработанного и специфицированного вычислительного микросервиса, либо для обеспечения его автономного запуска с возможностью самостоятельного (без жесткой привязки к AMPMA) получения заданий от внешних программных систем. По существу, с помощью конфигурирования микросервиса создается упрощенный вариант менеджера-агента, способного получить задание, запустить микросервис и предоставить результаты работы.

На рис. 4 и 5 приведены базовая и модифицированная схемы взаимодействия пользователя и объектов АМР.

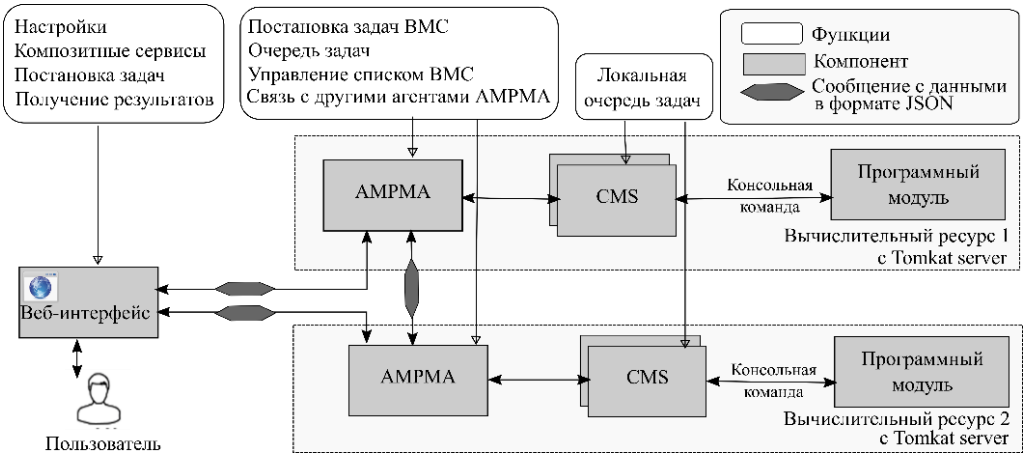


Рис. 4. Базовая схема взаимодействия

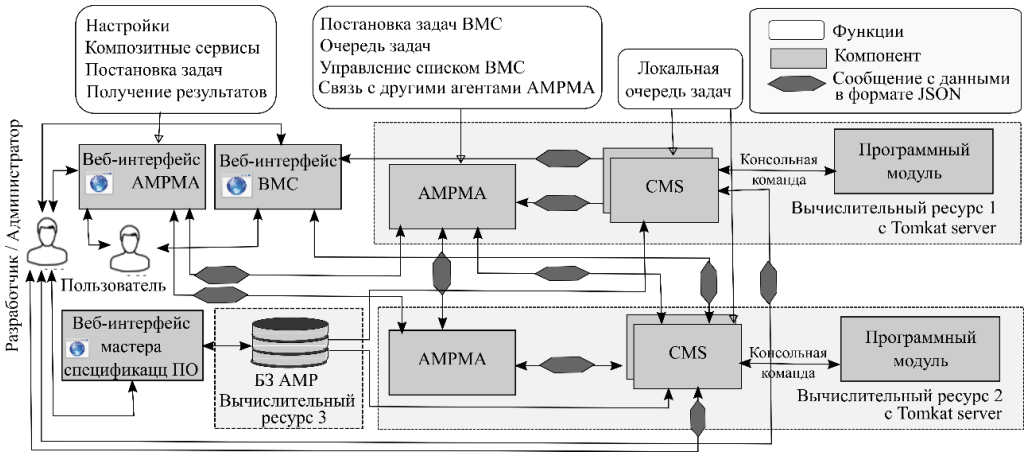


Рис. 5. Модифицированная схема взаимодействия

2.2. Мастер композиции CMS. Одним из преимуществ сервис-ориентированных вычислений является возможность объединять несколько сервисов для решения более сложных задач [19, 20]. Составной веб-сервис может быть создан либо про-активно, либо реактивно [21]. В первом случае сервисы-компоненты составного веб-сервиса собираются заранее. Во втором случае такой веб-сервис создается динамически по запросу пользователя. Консолидированная и целостная структура анализа моделей, языков, методов, платформ и инструментов композиции сервисов приведена в [22, 23].

Под композитным сервисом далее мы будем иметь в виду композицию микросервисов в статическом режиме и ансамбль микросервисов в динамическом. Композитные сервисы предназначены для решения рассмотренных во введении научных задач и ориентированы на использование в ГМВИ, в состав которой входят как локальные, так и облачные ресурсы.

Мастер композиции предоставляет два способа создания композитного сервиса:

- на основе построения плана вычислений по постановке задачи (ПЗ) с получением схемы сервиса, представляющей собой упорядоченную последовательность (композицию) или ансамбль CMS (в случае непроцедурной ПЗ на распределенной БЗ АМР [24]);
- с использованием графического редактора, позволяющего перестроить схему сервиса из БЗ АМР или создать новую схему на основе хранящихся в БЗ спецификаций микросервисов и параметров.

Получение схемы сервиса, в зависимости от ПЗ и способа организации БЗ АМР, выполняется следующим образом:

– для процедурной ПЗ на основе последовательности требуемых для решения задачи CMS и их семантического описания, хранящегося в БЗ АМР в виде фрагментов отношений, описывающих взаимосвязи микросервисов с параметрами, выполняется информационное планирование [24];
– для непроцедурной ПЗ и распределенной БЗ АМР на основе динамического подхода ансамбль CMS собирается группой распределенных вычислительных агентов на основе логического вывода [24];
– для непроцедурной ПЗ и централизованной БЗ АМР используется булева модель построения планов вычислительных действий в виде системы дифференциальных булевых уравнений, решения которых при заданных начальных условиях (согласованных с непроцедурной ПЗ) определяют разрешимость поставленной задачи и дают конструктивные планы ее решения (в том числе и параллельные) [25].

Графический редактор для построения схемы сервиса в виде графа реализован с использованием библиотеки baklavajs (https://github.com/newcat/baklavajs). В вершинах графа ставятся микросервисы (рис. 6), из спецификации автоматически добавляются параметры, связываются входы и выходы CMS, отмечаются параметры, выносимые на веб-интерфейс композитного сервиса.

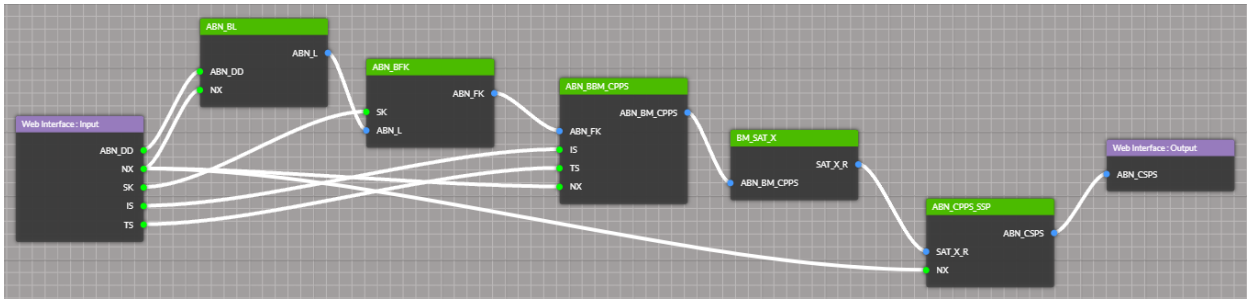


Рис. 6. Графическая визуализация схемы композитного сервиса

С помощью разработанных средств состав АМР BCM-QABDS дополнен CMS, требуемыми при решении следующих задач качественного анализа сингулярных, неявных и асинхронных булевых сетей:

– Построение булевых моделей (в виде SAT или QSAT задач) локальных динамических свойств, свойства периодичности траекторий и следующих свойств типа управляемости: основного свойства управляемости, свойства управляемости относительно пары состояний, свойства полной управляемости; свойства *k*-наблюдаемости. Для приведения булевых выражений в моделях динамических свойств к КНФ используются преобразование Цейтина [26], преобразование Плейстеда-Гринбаума [27] и преобразование булева уравнения АНФ=0 к виду КНФ=1 [28].
– Проверка выполнимости полученных булевых моделей осуществляется с помощью SAT и QSAT решателей, как разработанных ранее авторами в процессе исследований, так и внешних. Обзор SAT и QSAT решателей приведен в [29, 30].

В основе построения булевых моделей динамического свойства (первого этапа ее построения) для автономных, управляемых, сингулярных и асинхронных булевых сетей лежат функции одношагового и *k*-шагового переходов, соответственно L и Φ_k, отражающие динамику функционирования сети. На втором этапе булева модель дополняется спецификацией свойства. Булевы модели одного и того же динамического свойства различны для разных классов булевых сетей. В табл. 1 приведены для сравнения математические модели и функции одношагового перехода L для сингулярной [31] и асинхронной булевых сетей.

Таблица 1. Математическая модель и функция L сингулярной и асинхронной сетей

Класс булевой сети	Математическая модель	Функция одношагового перехода L
Сингулярная	$x^{t+1} = F(x^t, z^t)$, $H(x^t, z^t) = 0$	$L(x^0, x^1, z^0) = \vee_{i=1}^{n}(x_i^1 \oplus F_i(x^0, z^0)) \vee H(x^0, z^0) = 0$
Асинхронная	$x^t = F(x^{t-1}, u^{t-1})$	$L(x^0, x^1, g^0) = \vee_{i=1}^{n}(x_i^1 \oplus (f_i(x^0) \cdot g_i^0 \vee x_i^0 \cdot \bar{g}_i^0)) \vee C(g^0) = 0$

В приведенных в табл. 1 моделях общими являются следующие обозначения: $t \in T = \{1,2,..., k\}$ – дискретное время (номер шага), $x = col(x_1, x_2,..., x_n) \in B^n$ – n-мерный вектор состояния, $x_i \in B = \{0,1\}$, $x^0 \in B^n$ – начальное состояние.

Для математической модели сингулярной булевой сети, представленной в табл. 1, $z \in B^l$ – вектор вспомогательных переменных, $F(x, z)$, $H(x, z)$ – векторные функции алгебры логики, называемые функцией переходов и функцией ограничений соответственно:

$$(F : B^n \times B^l \to B^n,\ H : B^n \times B^l \to B).$$

Для математической модели асинхронной булевой сети, приведенной в табл. 1, $u \in U = \{1,2,..., n\}$ – управляющее воздействие (номер обновляемой переменной вектора состояния x), U – множество допустимых состояний управления u, $F(x,u) = col(F_1(x,u), F_2(x,u),..., F_n(x,u))$ – логическая вектор-функция, называемая глобальной функцией обновления ($F : B^n \times U \to B^n$). Координаты функции F определяются следующим образом:

$$\forall x \in B^n, \forall u \in U, F_i(x,u) = \begin{cases} f_i(x), \text{if } i = u, \\ x_i, \text{if } i \neq u \end{cases},$$

где $f_i(x)$ ($f_i : B^n \to B$, $i = 1,2,..., n$) – локальная булева функция обновления по i-той переменной вектора состояния. Математическая модель для k=1 имеет вид:

$$x_i^1 = \begin{cases} f_i(x^0), \text{if } i = u^0, \\ x_i^0, \text{if } i \neq u^0 \end{cases} i = 1,2,..., n. \quad (1)$$

Из (1) следует, что любые два смежных состояния x^{t-1}, x^t ($t \in T$) траектории различаются значением не более, чем по одной переменной вектора состояния. При формировании функции L каждый целочисленный элемент u из множества U представлен в виде характеристического вектора $g = col(g_1, g_2,..., g_n)$ так, что $g_i = 1$, если $i = u$ и $g_i = 0$ для всех $i \neq u$. Тогда с учетом замены множества U на множество G система уравнений (1) будет эквивалентна системе уравнений

$$x_i^1 = f_i(\mathrm{x}^0) \cdot \mathrm{g}_\mathrm{i}^0 \vee x_i^0 \cdot \bar{\mathrm{g}}_\mathrm{i}^0,\ i = 1,2,..., n \quad (2)$$

при условии, что вектор g^0 удовлетворяет ограничению

$$C(g^0) = \vee_{i=1}^{n-1} \vee_{j=i+1}^{n} g_i^0 \cdot g_j^0 \vee \bigwedge_{i=1}^{n} \bar{g}_i^0 = 0, \quad (3)$$

гарантирующему наличие в характеристическом векторе одной и только одной единицы, а функция L приобретает вид, представленный в табл. 1. Функция k-шагового перехода ($k > 1$) имеет следующий вид: $\Phi_k(x^0, x^1,..., x^k, g^0, g^1,..., g^{k-1}) = \vee_{t=1}^{k} L(x^t, x^{t-1}, g^{t-1}) = 0$.

Особый класс представляют собой неявные булевы сети – ДДС, функция переходов в которых определена неявным булевым уравнением относительно вектора переменных

текущего состояния и векторов переменных нескольких (следующих за текущим) состояний [32]. С точки зрения динамики неявной булевой сети основным является свойство продолжаемости ее локальных решений. В силу существенных различий в построении булевых моделей динамических свойств средства решения задач качественного анализа различных классов булевых сетей распределены по разным подсистемам АМР BCM-QABDS.

3. Демонстрационный пример. Рассмотрим более подробно разработку композитного сервиса для проверки свойства управляемости относительно пары состояний асинхронных булевых сетей, когда необходимо синтезировать такое управление g^*, которое переводит систему (2, 3) из заданного начального состояния $x^0 = c^0$ в целевое состояние c^* за k шагов, т.е. $x^k = c^*$. Необходимое управление – схема обновления переменных состояния $g^* = (g^0, g^1, ..., g^{k-1})$ – является решением булева уравнения:

$$\Phi_k(x^0, x^1, ..., x^k, g^0, g^1, ..., g^{k-1})\Big|_{\substack{x^0=c^0 \\ x^k=c^*}} = 0 \qquad (4)$$

относительно неизвестных векторов переменных $x^1, ..., x^{k-1}, g^0, g^1, ..., g^{k-1}$.

Для построения булевой модели свойства управляемости асинхронной булевой сети относительно пары состояний БЗ BCM-QABDS с помощью мастера спецификации дополнены следующими параметрами:

- ABN_DD – описание динамики сети;
- ABN_L и ABN_ФК – булевы ограничения в виде КНФ в формате DIMACS (https://logic.pdmi.ras.ru/~basolver/dimacs.html), соответствующие функциям L (табл. 1) и Φ_k;
- NX – размерность n вектора состояния x;
- IS, ST – начальное и целевое состояния, соответствующие c^0 и c^*;
- SK – дискретное время (номер шага k);
- ABN_BM_CPPS – булева модель рассматриваемого свойства в формате КНФ;
- SAT_X_R – результат работы ALLSAT решателя;
- ABN_CSPS – управляющая последовательность.

В функциональное наполнение АМР включены следующие CMS:

- ABN_BL – формирование функции L в виде КНФ в формате DIMACS;
- ABN_BFK – формирование функции Φ_k в виде КНФ в формате DIMACS;
- ABN_BBM_CPPS – формирование булевой модели ABN_BM_CPPS;
- ABN_CPPS_SSP – обработка результатов проверки выполнимости модели ABN_BM_CPPS.

Интерфейс для ввода параметров и микросервисов приведен при описании мастера спецификации ПО. Для проверки выполнимости булевой модели рассматриваемого свойства используется вычислительный микросервис BM_SAT_X, разработанный ранее на основе последовательного и параллельного ALLSAT решателей [33]. Требуемая компоновка микросервисов строится с помощью мастера композитных сервисов. Схему сервиса можно получить в планировщике по процедурной или непроцедурной постановке задачи, а также составить (или отобразить уже полученную) в графическом редакторе (см. рис. 6).

В качестве примера конкретной асинхронной булевой сети используется модель динамики одношаговых переходов вида (1, 2), описываемая следующей системой уравнений (n=3, $x = col(x_1, x_2, x_3)$):

$$\begin{aligned} x_1^1 &= (x_1^0 \vee \bar{x}_2^0 \vee \bar{x}_3^0) \cdot g_1^0 \vee x_1^0 \cdot \bar{g}_1^0 \\ x_2^1 &= \bar{x}_3^0 \cdot g_2^0 \vee x_2^0 \cdot \bar{g}_2^0 \\ x_3^1 &= x_2^0 \cdot g_3^0 \vee x_3^0 \cdot \bar{g}_3^0 \\ g_1^0 \cdot g_2^0 &\vee g_1^0 \cdot g_3^0 \vee g_2^0 \cdot g_3^0 \vee \bar{g}_1^0 \cdot \bar{g}_2^0 \cdot \bar{g}_3^0 = 0 \end{aligned} \quad . \qquad (5)$$

Проверим управляемость системы (5) для пары состояний $c^0 = 010$ и $c^* = 101$. По описанию динамики (5) с помощью микросервиса ABN_BL формируется функция L в виде КНФ в формате DIMACS (рис. 7). Переменным $x_1^0, x_2^0, x_3^0, g_1^0, g_2^0, g_3^0, x_1^1, x_2^1, x_3^1$ в этом формате соответствуют коды 1, 2, 3, 4, 5, 6, 7, 8, 9. Для переменных с отрицанием перед кодом ставится знак «-». Код «0» является разделителем дизъюнктов.

```
p cnf 9 21
7 -1 -4 0 7 2 -4 0 7 3 -4 0 7 -1 4 0 -7 1 -2 -3 0 -7 1 4 0 -7 1 -2 -3 -4
0 8 3 -5 0 8 -2 5 0 -8 -3 2 0 -8 2 5 0 -8 -3 -5 0 9 -2 -6 0 9 -3 6 0 -9 2
3 0 -9 3 6 0 -9 2 -6 0 -4 -5 0 -4 -6 0 -5 -6 0 4 5 6 0
```

Рис. 7. Функция L для системы (5) в формате DIMACS

Микросервис ABN_ВФК получает на основе L для заданного $k > 1$ функцию Φ_k, дублируя ограничения k раз с увеличением кодов очередного дубля на $6 \cdot (k-1)$. Микросервис ABN_BBM_PCCS завершает формирование булевой модели, добавляя к функции Φ_k спецификацию свойства (4). Для очередного k микросервисом BM_SAT_X проверяется выполнимость булевой модели. Для k=3 находятся два решения:

– -1 2 -3 -4 -5 6 -7 8 9 -10 11 -12 -13 -14 15 16 -17 -18 19 -20 21 0;
– -1 2 -3 4 -5 -6 7 8 -9 -10 -11 12 13 14 15 -16 17 -18 19 -20 21 0.

Микросервис ABN_CPPS_SSP обрабатывает этот результат и выдает окончательный ответ в следующем виде: $010(3) \to 011(2) \to 001(1) \to 101$; $010(1) \to 110(3) \to 111(2) \to 101$. В скобках указано значение управления u^0 (равное значению индекса единичного элемента вектора g^0). Таким образом, для начального состояния $c^0 = 010$ и целевого состояния $c^* = 101$ система (5) управляема для k=3, что подтверждается диаграммой переходов (рис. 8).

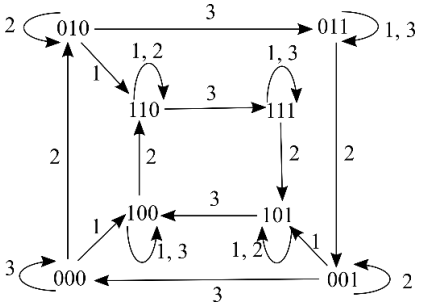


Рис. 8. Диаграмма переходов управляемой асинхронной булевой сети (4)

Технология решения задач с помощью разработанных инструментальных средств, представленная в демонстрационном примере, применялась для качественного анализа двоичных динамических моделей объектов в различных предметных областях. В частности, в [9] при решении практических задач в области биоинформатики проверялось наличие свойства k-наблюдаемости сети активации Т-клеток, двоичная динамическая модель которой

представлена в [13, пример 2]. Построены неявные булевы модели для решения задач искусственного интеллекта о планировании действий и о минимальном покрытии множества, имеющих большое количество приложений. Результаты вычислительных экспериментов для этих моделей приведены соответственно в [32] и [34].

Заключение. В статье представлены новые средства автоматизации создания и применения пакета прикладных микросервисов для решения задач качественного исследования разнообразных классов булевых сетей на основе метода булевых ограничений. С помощью этих средств функциональное наполнение пакета BCM-QADBS расширено микросервисами для проверки динамических свойств сингулярных, неявных и асинхронных булевых сетей. Практическая значимость разработанного инструментария обусловлена широким применением булевых сетей в качестве двоичных моделей объектов в разнообразных предметных областях и состоит в повышении эффективности исследования этих моделей путем автоматизации всех этапов его проведения.

Благодарности. Исследование выполнено при поддержке Министерства науки и высшего образования Российской Федерации, проект № 121032400051-9.

Список источников

1. Oparin G.A., Bogdanova V.G., Pashinin A.A. Qualitative analysis of autonomous synchronous binary dynamic systems. Mathematics in engineering, science and aerospace, 2019, vol. 10, no. 3, pp. 407-419.
2. Oparin G.A., Bogdanova V.G., Pashinin A.A. et al. Microservice-oriented approach to automation of distributed scientific computations. 42nd International convention on information and communication technology, electronics and microelectronics (MIPRO), IEEE, 2019, pp. 236-241, DOI: 10.23919/MIPRO.2019.8757053.
3. Опарин Г.А. Технология качественного исследования двоичных динамических систем в микросервисной инфраструктуре / Г.А. Опарин, В.Г. Богданова, А.А. Пашинин // Информационные и математические технологии, 2022. – № 4(28). – С. 137-148. DOI: 10.38028/ESI.2022.28.4.011.
4. Опарин Г.А. Управление конвейерно-параллельными вычислениями при решении задач качественного исследования двоичных динамических систем на основе метода булевых ограничений / Г.А. Опарин, В.Г. Богданова, А.А. Пашинин // Информационные и математические технологии в науке и управлении, 2019. – №3(15). – С. 79-90. – DOI: 10.25729/2413-0133-2019-3-07.
5. Маслов Ю.С. Теория дедуктивных систем и ее применения / Ю.С. Маслов. –М.: Радио и связь, 1986.
6. Jhala R., Majumdar R. Software model checking. ACM Comput. Surv, 2015, vol. 41, no. 4, pp. 21:1–21:54.
7. Vassilyev S.N. Method of reduction and qualitative analysis of dynamic systems: I. J. Comput. Syst. Sci. Int., 2006, vol. 45, pp. 17-25.
8. Vassilyev S. N. Method of reduction and qualitative analysis of dynamic systems: II. J. Comput. Syst. Sci. Int., 2006, vol. 45, pp. 167-179.
9. Oparin G.A., Bogdanova V.G., Pashinin A.A. Investigation of observability property of controlled binary dynamical systems: a logical approach. Proc. of the 2nd International Workshop on information, computation, and control systems for distributed environments, Irkutsk, Russia, July 6-7, 2020. CEUR-WS Proceedings, 2021, vol. 2913, pp. 129-138, DOI:10.47350/ICCS-DE.2021.11.
10. Cheng D., Qi H., Li Z. Analysis and control of Boolean networks: A semi-tensor product approach. Springer, London, 2011.
11. Hochma G., Margaliot M., Fornasini E. et al. Symbolic dynamics of Boolean control networks. Automatica, 2013, vol. 49, no. 8, pp. 2525-2530, DOI: 10.1016/j.automatica.2013.05.004.
12. Zhao Y., Qi H., Cheng D. Input-state incidence matrix of Boolean control networks and its applications. Sys. Contr. Lett., 2010, vol. 59, no. 12, pp. 767-774.
13. Li R., Yang M., Chu T. Controllability and observability of Boolean networks arising from biology. Chaos, 2015, vol. 25, no. 2, 023104 p.
14. Bardet M. Faugère C., Salvy B. On the complexity of the F5 Gröbner basis algorithm. Journal of symbolic computation, 2015, vol. 70, pp. 49-70.
15. Hinkelmann F., Brandon M., Guang B. et al. ADAM: Analysis of discrete models of biological systems using computer algebra. BMC bioinformatics, 2011, vol. 12:295, DOI: 10.1186/1471-2105-12-295.
16. Naldi A., Hernandez C., Levy N. et al. The CoLoMoTo Interactive notebook: accessible and reproducible computational analyses for qualitative biological networks. Frontiers in physiology, 2018, vol. 9, article 680, DOI: 10.3389/fphys.2018.00680.

17. Oles V., Kukushkin A. BoolSi: a tool for distributed simulations and analysis of Boolean networks. Journal of open research software, 2020, vol. 8, iss. 1, article 26, DOI:10.5334/jors.308.
18. Oparin G.A., Bogdanova V.G., Pashinin A.A. Microservice approach to the qualitative study of attractors of binary dynamic systems based on the Boolean constraint method. 43rd International Convention on information and communication technology, electronics and microelectronics, IEEE, 2020, pp. 1904-1909, DOI:10.23919/MIPRO48935.2020.9245300.
19. Huhns M.N., Sing M.P. Service-oriented computing: key concepts and principles. IEEE Internet computing, 2005, vol. 9, iss. 1, pp. 75-81.
20. Rao J., Su X. A survey of automated web service composition methods. In Cardoso J., Sheth A. (eds) Semantic Web services and Web process composition. SWSWPC 2004. Lecture Notes in computer science, Springer, Berlin, Heidelberg, 2005, vol. 3387, pp. 43-54, DOI: 10.1007/978-3-540-30581-1_5.
21. Rawat M.K., Upadhayay D.C. Invoking web services with software agents. IJERT, 2012, vol.1, iss.4. DOI:10.17577/IJERTV1IS4041.
22. Lemos A.L., Daniel F., Benatallah B. Web service composition: a survey of techniques and tools. ACM computing surveys, 2015, vol. 48, iss. 3, article 33, pp. 1-41, DOI: 10.1145/2831270.
23. Huf A., Siqueira F. Composition of heterogeneous web services: A systematic review. Journal of network and computer applications, 2019, vol. 143, pp. 89-110, DOI: 10.1016/j.jnca.2019.06.008.
24. Oparin G.A., Bogdanova V.G., Pashinin A.A. Automated tools for the development of microservice compositions for hybrid scientific computations. Proceedings of the 2nd International Workshop on information, computation, and control systems for distributed environments, CEUR-WS Proceedings, 2020, pp 201-213.
25. Oparin G.A., Bogdanova V.G., Pashinin A.A. Binary dynamic models of structural synthesis of programs. 46th International Convention on information and communication technology, electronics and microelectronics, 2023, pp. 1173-1178.
26. Tseitin G.S. On the complexity of derivation in propositional calculus. In Siekmann J.H., Wrightson G. (eds) Automation of reasoning. symbolic computation. Springer, Berlin, Heidelberg, 1983, pp. 466-483, DOI: 10.1007/978-3-642-81955-1_28.
27. Plaisted D.A., Greenbaum S. A structure preserving clause form translation. Symbolic computation, 1986, vol. 2, iss. 3, pp. 293-304.
28. Sage 9.8. Reference manual. An ANF to CNF Converter using a dense/sparse strategy. Available at: https://doc.sagemath.org/html/en/reference/sat/sage/sat/converters/polybori.html (accessed: 05/29/2023).
29. Gong W., Zhou X. A survey of SAT solver. AIP conference proceedings, 2017, vol. 1836, p. 020059.
30. Marin P., Massimo Narizzano M., Pulina L. et al. Twelve years of QBF evaluations: QSAT Is PSPACE-hard and it shows. Fundam. Informaticae, 2016, vol. 149, iss. 1-2, pp. 133-158.
31. Oparin G.A., Bogdanova V.G., Pashinin A.A. The Boolean constraint method application for qualitative analysis of the dynamical properties of singular Boolean networks. Computational technologies, 2023, vol. 28, no. 1. pp. 48-60, DOI: 10.25743/ICT.2023.28.1.006.
32. Oparin G. A., Bogdanova V. G., Pashinin A. A. Implicit Boolean network for planning actions. 45th Jubilee international convention on information, communication and electronic technology (MIPRO), IEEE, 2022, pp. 871-876, DOI: 10.23919/MIPRO55190.2022.9803418.
33. Toda T., Soh T. Implementing efficient all solutions SAT solvers. ACM Journal of experimental algorithmics, 2016, vol. 21, no. 1.12, pp. 1-44, DOI: 10.1145/2975585.
34. Oparin G.A., Bogdanova V.G., Pashinin A.A. Implicit Boolean networks and their application to combinatorial problems. Mathematics in engineering, science and aerospace, 2022, vol. 13, no. 1, pp. 25-35.

Опарин Геннадий Анатольевич. *Доктор технических наук, профессор, главный научный сотрудник лаборатории 5.1 Параллельных и распределенных вычислительных систем, заведующий отделением вычислительных и управляющих систем, Институт динамики систем и теории управления им. В.М. Матросова СО РАН, AuthorID: 1498, SPIN: 4544-2804, ORCID: 0000-0002-3757-6041, prn51@icc.ru, 664033, Россия, г. Иркутск, ул. Лермонтова, 134.*

Богданова Вера Геннадьевна. *Кандидат технических наук, доцент, старший научный сотрудник лаборатории 5.1 Параллельных и распределенных вычислительных систем, Институт динамики систем и теории управления им. В.М. Матросова СО РАН, AuthorID: 1536, SPIN: 9123-3563, ORCID: 0000-0002-5763-6849, bvg@icc.ru, 664033, Россия, г. Иркутск, ул. Лермонтова, 134.*

Пашинин Антон Алексеевич. *Младший научный сотрудник лаборатории 5.1 Параллельных и распределенных вычислительных систем, Институт динамики систем и теории управления им. В.М. Матросова СО РАН, AuthorID: 776425, SPIN: 9325-0914, ORCID: 0000-0002-1205-7595, apcrol@gmail.com, 664033, Россия, г. Иркутск, ул. Лермонтова, 134.*

UDC 004.421+004.4’2+004.771
DOI:10.25729/ESI.2024.34.2.015

Automating tools for the development and application of an applied microservices package

Gennady A. Oparin, Vera G. Bogdanova, Anton A. Pashinin

Matrosov institute for system dynamics and control theory SB RAS,
Russia, Irkutsk, *apcrol@gmail.com*

Abstract. Results related to the developed earlier Boolean constraint method for studying the dynamics and parametric synthesis of control systems, namely, Boolean networks, in a microservice infrastructure using knowledge base tools, are presented. The computational model of the subject area acts as a knowledge base, which is understood as a set of information about the subject area objects and the relationships between them. Each functional relationship is implemented by a computational microservice created based on a software module that calculates the values of the output parameters from the given values of the input parameters. A set of computational microservices constitutes the functional software of an applied microservices package. The system software consists of services for setting the task, planning and monitoring calculations, data management, and processing results. Agents of the applied microservices package, depending on the problem statement and the knowledge base organization, carry out computing management (decentralized, centralized or hierarchical). In connection with the studying of new classes of Boolean networks (controlled, singular, implicit, asynchronous, bipartite, and others), we are extending the earlier created applied microservices package based on the above approach and intended for solving the problems of qualitative analysis of autonomous Boolean networks. The aim of the study is to develop a new toolkit for the HPCSOMAS-MSC microservice intelligent computational platform for creating, configuring and accessing computational microservices, which allows interacting with package microservices directly through a web interface. A web interface for the subject area description is implemented in addition to the existing one based on the JSON language. The new toolkit allows both to speed up the development and debugging of microservices, and to make further interaction more convenient for the subject area specialist.

Keywords: boolean network, qualitative analysis, Boolean constraint method, applied microservices package

Acknowledgements: the study was supported by the Ministry of Science and Higher Education of the Russian Federation, project no. 121032400051-9.

References

1. Oparin G.A., Bogdanova V.G., Pashinin A.A. Qualitative analysis of autonomous synchronous binary dynamic systems. Mathematics in engineering, science and aerospace, 2019, vol. 10, no. 3, pp. 407-419.
2. Oparin G.A., Bogdanova V.G., Pashinin A.A. et al. Microservice-oriented approach to automation of distributed scientific computations. 42nd International convention on information and communication technology, electronics and microelectronics (MIPRO), IEEE, 2019, pp. 236-241, DOI: 10.23919/MIPRO.2019.8757053.
3. Oparin G.A., Bogdanova V.G., Pashinin A.A. Tehnologija kachestvennogo issledovanija dvoichnyh dinamicheskih sistem v mikroservisnoj infrastrukture [Technology of qualitative study of binary dynamic systems in microservice infrastructure]. Informatsionnye i matematicheskie tekhnologii v nauke i upravlenii [Information and mathematical technologies in science and management], 2022, vol. 4, no. 28, pp. 137-148, DOI: 10.38028/ESI.2022.28.4.011.
4. Oparin G.A., Bogdanova V.G., Pashinin A.A. Upravlenie konvejerno-parallel'nymi vychislenijami pri reshenii zadach kachestvennogo issledovanija dvoichnyh dinamicheskih sistem na osnove metoda bulevyh ogranichenij [Management of pipeline-parallel computations for solving problems of qualitative study of binary dynamic systems based on the Boolean constraint method]. Informatsionnye i matematicheskie tekhnologii v nauke i upravlenii [Information and mathematical technologies in science and management], 2019, vol. 3, no. 15, pp. 79-90, DOI: 10.25729/2413-0133-2019-3-07.
5. Maslov S. Ju. Teorija deduktivnyh sistem i ee primenenie [Deductive systems theory and its applications]. M., Radio i svjaz’ [Radio and communications], 1986.
6. Jhala R., Majumdar R. Software model checking. ACM Comput. Surv, 2015, vol. 41, no. 4, pp. 21:1–21:54.
7. Vassilyev S.N. Method of reduction and qualitative analysis of dynamic systems: I. J. Comput. Syst. Sci. Int., 2006, vol. 45, pp. 17-25.
8. Vassilyev S. N. Method of reduction and qualitative analysis of dynamic systems: II. J. Comput. Syst. Sci. Int., 2006, vol. 45, pp. 167-179.

9. Oparin G.A., Bogdanova V.G., Pashinin A.A. Investigation of observability property of controlled binary dynamical systems: a logical approach. Proc. of the 2nd International Workshop on information, computation, and control systems for distributed environments, Irkutsk, Russia, July 6-7, 2020. CEUR-WS Proceedings, 2021, vol. 2913, pp. 129-138, DOI:10.47350/ICCS-DE.2021.11.
10. Cheng D., Qi H., Li Z. Analysis and control of Boolean networks: A semi-tensor product approach. Springer, London, 2011.
11. Hochma G., Margaliot M., Fornasini E. et al. Symbolic dynamics of Boolean control networks. Automatica, 2013, vol. 49, no. 8, pp. 2525-2530, DOI: 10.1016/j.automatica.2013.05.004.
12. Zhao Y., Qi H., Cheng D. Input-state incidence matrix of Boolean control networks and its applications. Sys. Contr. Lett., 2010, vol. 59, no. 12, pp. 767-774.
13. Li R., Yang M., Chu T. Controllability and observability of Boolean networks arising from biology. Chaos, 2015, vol. 25, no. 2, 023104 p.
14. Bardet M. Faugère C., Salvy B. On the complexity of the F5 Gröbner basis algorithm. Journal of symbolic computation, 2015, vol. 70, pp. 49-70.
15. Hinkelmann F., Brandon M., Guang B. et al. ADAM: Analysis of discrete models of biological systems using computer algebra. BMC bioinformatics, 2011, vol. 12:295, DOI: 10.1186/1471-2105-12-295.
16. Naldi A., Hernandez C., Levy N. et al. The CoLoMoTo Interactive notebook: accessible and reproducible computational analyses for qualitative biological networks. Frontiers in physiology, 2018, vol. 9, article 680, DOI: 10.3389/fphys.2018.00680.
17. Oles V., Kukushkin A. BoolSi: a tool for distributed simulations and analysis of Boolean networks. Journal of open research software, 2020, vol. 8, iss. 1, article 26, DOI:10.5334/jors.308.
18. Oparin G.A., Bogdanova V.G., Pashinin A.A. Microservice approach to the qualitative study of attractors of binary dynamic systems based on the Boolean constraint method. 43rd International Convention on information and communication technology, electronics and microelectronics, IEEE, 2020, pp. 1904-1909, DOI:10.23919/MIPRO48935.2020.9245300.
19. Huhns M.N., Sing M.P. Service-oriented computing: key concepts and principles. IEEE Internet computing, 2005, vol. 9, iss. 1, pp. 75-81.
20. Rao J., Su X. A survey of automated web service composition methods. In Cardoso J., Sheth A. (eds) Semantic Web services and Web process composition. SWSWPC 2004. Lecture Notes in computer science, Springer, Berlin, Heidelberg, 2005, vol. 3387, pp. 43-54, DOI: 10.1007/978-3-540-30581-1_5.
21. Rawat M.K., Upadhayay D.C. Invoking web services with software agents. IJERT, 2012, vol.1, iss.4. DOI:10.17577/IJERTV1IS4041.
22. Lemos A.L., Daniel F., Benatallah B. Web service composition: a survey of techniques and tools. ACM computing surveys, 2015, vol. 48, iss. 3, article 33, pp. 1-41, DOI: 10.1145/2831270.
23. Huf A., Siqueira F. Composition of heterogeneous web services: A systematic review. Journal of network and computer applications, 2019, vol. 143, pp. 89-110, DOI: 10.1016/j.jnca.2019.06.008.
24. Oparin G.A., Bogdanova V.G., Pashinin A.A. Automated tools for the development of microservice compositions for hybrid scientific computations. Proceedings of the 2nd International Workshop on information, computation, and control systems for distributed environments, CEUR-WS Proceedings, 2020, pp 201-213.
25. Oparin G.A., Bogdanova V.G., Pashinin A.A. Binary dynamic models of structural synthesis of programs. 46th International Convention on information and communication technology, electronics and microelectronics, 2023, pp. 1173-1178.
26. Tseitin G.S. On the complexity of derivation in propositional calculus. In Siekmann J.H., Wrightson G. (eds) Automation of reasoning. symbolic computation. Springer, Berlin, Heidelberg, 1983, pp. 466-483, DOI: 10.1007/978-3-642-81955-1_28.
27. Plaisted D.A., Greenbaum S. A structure preserving clause form translation. Symbolic computation, 1986, vol. 2, iss. 3, pp. 293-304.
28. Sage 9.8. Reference manual. An ANF to CNF Converter using a dense/sparse strategy. Available at: https://doc.sagemath.org/html/en/reference/sat/sage/sat/converters/polybori.html (accessed: 05/29/2023).
29. Gong W., Zhou X. A survey of SAT solver. AIP conference proceedings, 2017, vol. 1836, p. 020059.
30. Marin P., Massimo Narizzano M., Pulina L. et al. Twelve years of QBF evaluations: QSAT Is PSPACE-hard and it shows. Fundam. Informaticae, 2016, vol. 149, iss. 1-2, pp. 133-158.
31. Oparin G.A., Bogdanova V.G., Pashinin A.A. The Boolean constraint method application for qualitative analysis of the dynamical properties of singular Boolean networks. Computational technologies, 2023, vol. 28, no. 1. pp. 48-60, DOI: 10.25743/ICT.2023.28.1.006.

32. Oparin G. A., Bogdanova V. G., Pashinin A. A. Implicit Boolean network for planning actions. 45th Jubilee international convention on information, communication and electronic technology (MIPRO), IEEE, 2022, pp. 871-876, DOI: 10.23919/MIPRO55190.2022.9803418.
33. Toda T., Soh T. Implementing efficient all solutions SAT solvers. ACM Journal of experimental algorithmics, 2016, vol. 21, no. 1.12, pp. 1-44, DOI: 10.1145/2975585.
34. Oparin G.A., Bogdanova V.G., Pashinin A.A. Implicit Boolean networks and their application to combinatorial problems. Mathematics in engineering, science and aerospace, 2022, vol. 13, no. 1, pp. 25-35.

Oparin Gennady Anatol'evich. *Doctor of technical sciences, professor, chief researcher of the laboratory 5.1. of Parallel and distributed computing systems, head of computing and control systems department, institute of system dynamics and control theory named after V.M. Matrosov SB RAS, AuthorID: 1498, SPIN: 4544-2804, ORCID: 0000-0002-3757-6041, prn51@icc.ru, 664033, Russia, Irkutsk, st. Lermontov, 134.*

Bogdanova Vera Gennad'evna. *PhD in technical sciences, associate professor, senior researcher of the laboratory 5.1. of parallel and distributed computing systems, institute of system dynamics and control theory named after V.M. Matrosov SB RAS, AuthorID: 1536, SPIN: 9123-3563, ORCID: 0000-0002-5763-6849, bvg@icc.ru, 664033, Russia, Irkutsk, st. Lermontov, 134.*

Pashinin Anton Alekseevich. *Junior researcher of the laboratory 5.1. of parallel and distributed computing systems, institute of system dynamics and control theory named after V.M. Matrosov SB RAS, AuthorID: 776425, SPIN: 9325-0914, ORCID: 0000-0002-1205-7595, apcrol@gmail.com, 664033, Russia, Irkutsk, st. Lermontov, 134.*

Статья поступила в редакцию 22.06.2023; одобрена после рецензирования 08.04.2024; принята к публикации 30.05.2024.

The article was submitted 06/22/2023; approved after reviewing 04/08/2024; accepted for publication 05/30/2024.

УДК 004.9:519.768
DOI:10.25729/ESI.2024.34.2.016

Оптимизационное предредактирование узкоспециальных русскоязычных текстов для их машинного перевода на английский язык

Животова Алена Анатольевна, Бердоносов Виктор Дмитриевич

Комсомольский-на-Амуре государственный университет,
Россия, Комсомольск-на-Амуре, *zhivotova.aa@gmail.com*

Аннотация. В статье исследуется процесс перевода текстов, а именно метод оптимизационного предредактирования, как способ повышения качества машинного перевода на английский язык при работе с русскоязычными узкоспециальными текстами. Авторы рассматривают математическую модель процесса перевода и постановку задачи машинного перевода, предлагают новую теорию вероятностной оценки сложности задачи перевода, приводят постановку и решение задачи оптимизационного предредактирования, описывают методику подготовки данных для обучения модели автоматического оптимизационного предредактирования. В результате исследования реализован программный комплекс оптимизационного предредактирования русскоязычных текстов. При разработке программного комплекса использованы ресурсы Центра коллективного пользования научным оборудованием «Центр обработки и хранения научных данных ДВО РАН». Данные для обучения и валидации моделей предоставлены ООО «Агентство переводов “ФИАС-Амур”». Тестирование программного комплекса показало эффективность предложенных методик для повышения качества машинного перевода узкоспециальных русскоязычных текстов на английский язык.

Ключевые слова: машинный перевод, оптимизационное предредактирование, сложность задачи перевода, качество перевода

Цитирование: Животова А.А. Оптимизационное предредактирование узкоспециальных русскоязычных текстов для их машинного перевода на английский язык / А.А. Животова, В.Д. Бердоносов // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 169-182. – DOI:10.25729/ESI.2024.34.2.016.

Введение. Перевод – рутинная необходимость во многих отраслях, включая науку, производство, медицину и т.д., и с ростом количества информации и скорости ее генерирования растет и потребность в повышении качества перевода наряду с сокращением затрат на него. Несмотря на выдающиеся прорывы нейросетевых, гибридных и больших языковых моделей машинного перевода (МП) в области семантической точности и гладкости перевода, вопрос качества перевода системами МП нельзя назвать решенным. Результат работы МП – черновик, который пользователь должен оценить и доработать самостоятельно. При этом пользователь без знания языка перевода не имеет инструментов для того, чтобы влиять на результат или хотя бы оценить качество полученного перевода [1, 2]. Эту проблему активно освещают зарубежные исследователи A. Lear, C. Quinci, C. Canfora, A. Ottman, D. Kenny, P. Sanchez-Gijon. Предоставляя пользователю средства обработки текста на языке, носителем которого он является, на любом из этапов перевода, можно повысить его качество. Именно на этот принцип опирается концепция интерактивного перевода, широко описанная в литературе [3]. Одним из направлений интерактивного перевода является перевод с предредактированием, когда исходный текст предварительно редактируется с целью его адаптации для более легкого «понимания» системой МП.

Значительный вклад в разработку теоретических и практических основ в области подготовки исходных текстов к переводу, предварительного редактирования и упрощения естественных языков для систем автоматической обработки текстов, в частности систем МП, внесли зарубежные авторы: V. Kumar, F. Azadi, M. Federico, V. Alabau – в области интерактивного перевода [4]; V. Sereton, P. Bouillon, J. Gerlach, A. Taufik, Y. Liang, W. Han, A. G. Arenas, C. Shei, Y. Hiraoka, M. Yamada, R. Miyata, A. Fujita – в области разработки подходов к пред-

редактированию [5, 6, 7]; L. O'Brien, D. Folaron, W. Aziz, M. Toledo – в области контролируемых и упрощенных языков [7]. Среди российских авторов и для русского языка данная тема освещена незначительно, однако известны работы И. В. Оборневой [8], А.Д Дмитриевой, А. Н. Лапошиной, М. Ю. Лебедевой [9] и др. в области оценки восприятия текста и упрощения русскоязычных текстов в соответствии с квалификацией реципиента.

Целью работы является разработка моделей и алгоритмов и их реализация для повышения качества МП узкоспециальных технических текстов путем автоматического оптимизационного предредактирования.

Основная задача заключается в том, чтобы, используя особенности работы алгоритмов систем МП и основы теории перевода, автоматизировать предварительное редактирование исходных текстов, оптимизировав их структуру, благодаря чему системы МП будут эффективнее переводить их на требуемый язык и допускать меньше стилистических ошибок, для распознания которых требуется более высокая компетенция пользователя в области языка перевода.

В основе проведенного исследования лежит применение апробированных математических методов, включая теорию множеств, численные методы оптимизации, такие, как метод наименьших квадратов и метод градиентного спуска, статистические методы, в том числе метод максимального правдоподобия.

В работе предложен новый алгоритм, позволяющий расширить область применения оптимизационного метода градиентного спуска путём использования элементов нечеткой логики в выражении функции правдоподобия через функцию принадлежности полученного текста низкой сложности задачи перевода для выбранной системы МП; предложен новый алгоритм, позволяющий расширить область применения метода наименьших квадратов для поиска весов значимости параметров исходного текста для вероятностной оценки ожидаемого качества его перевода на целевой язык; предложена новая архитектура и реализован программный комплекс для повышения качества МП текстов с русского языка на английский язык, отличающийся от существующих применением ансамбля моделей для оптимизационного предредактирования на основе вероятностной оценки сложности задачи перевода с целью повышения качества МП текстов с русского языка на английский язык.

1. Математическая модель процесса перевода. Перевод – передача смысла текста на языке оригинала в соответствии с языковыми и культурными традициями языка перевода согласно требованиями конечного реципиента.

Процесс перевода формально может быть представлен совокупностью множеств и операций над этими множествами. В дальнейшем слово «множество» опускается для удобства повествования. Пусть $txt_{i\text{ТХТ}}$ – исходный текст, характеризующийся множеством выраженных смысловых единиц СМ_i, $\text{яз}_{\text{вх}}$ – язык исходного текста, $\text{яз}_{\text{вых}}$ – язык перевода, $\text{дп}_{i\text{ДП}}$ – предметная область (домен приложения) исходного текста, $\text{ТР}|txt_{i\text{ТХТ}} : \exists\ \text{ОК}|\text{ТР} \in R$ – требования к переводу исходного текста на язык перевода, такие, что существует ОК – аналитически вычислимая оценка качества перевода, соответствующая критериям, зафиксированным в требованиях к переводу, перевод выполняется переводчиком $\text{пер}_{i\text{ПЕР}}$, обладающим компетенцией $\mathbf{К}_{\mathbf{пер}i\mathbf{ПЕР}}$ и специализацией $\mathbf{С}_{\mathbf{пер}i\mathbf{ПЕР,\ дп}i\mathbf{ДП}}$, тогда качественный перевод – это сгенерированный текст $txt_{j\text{ТХТ}}$ на языке $\text{яз}_{\text{вых}}$, характеризующийся множеством выраженных смысловых единиц СМ_j: $(\text{СМ}_j \rightarrow \text{СМ}_i)$. В общем виде соотношения множеств функция перевода $F_{\text{пер}}$ может быть описана как

$$F_{\text{пер}} : (txt_{i\text{ТХТ}}, \mathbf{К}_{\mathbf{пер}i\mathbf{ПЕР}}, \mathbf{С}_{\mathbf{пер}i\mathbf{ПЕР,\ дп}i\mathbf{ДП}}) \rightarrow txt_{j\text{ТХТ}} \,|\big(\text{СМ}_j \rightarrow \text{СМ}_i, \text{ТР}\big|txt_{i\text{ТХТ}}\big). \quad (1)$$

Обобщенно модель процесса перевода представлена на рисунке 1. Подробное описание полной модели было опубликовано авторами ранее [10].

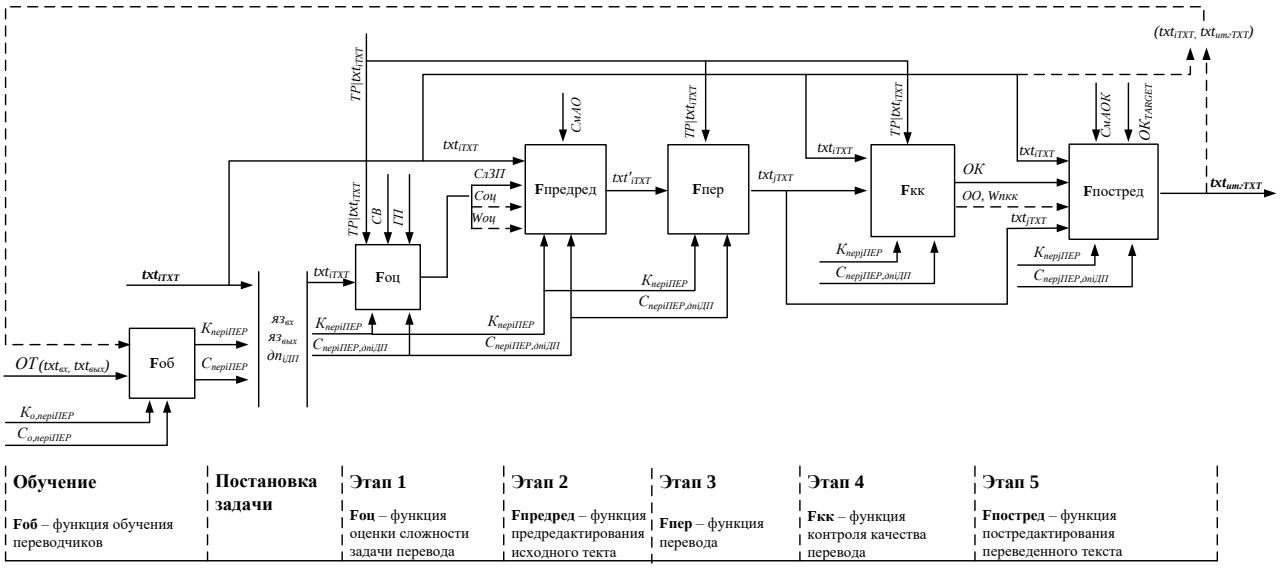


Рис. 1. Модель процесса перевода

Математически, для систем МП задачу перевода можно формализовать через метод максимизации функции правдоподобия [11] следующим образом. Опишем условия задачи, пусть:

$txt_{i\text{ТХТ}}$ – исходный текст на языке $\text{яз}_{\text{вх}}$.

$txt_{j\text{ТХТ}}$ – переведенный текст на языке $\text{яз}_{\text{вых}}$.

$TXT_{i\text{trg}}$ – множество всех возможных вариантов перевода текста txt_{iTXT} на язык $\text{яз}_{\text{вых}}$:

$$TXT_{i\text{trg}} = \{txt_0, txt_1, \dots, txt_j ,$$

где j – общее число вариантов перевода, $txt_{j\text{ТХТ}} \in TXT_{i\text{trg}}$.

$\text{ОК}_{i\text{trg}}$ – множество нормированных оценок качества перевода текста txt_{iTXT} в соответствии с требованиями к переводу $\text{ТР}|txt_{i\text{ТХТ}}$ для всех возможных вариантов перевода $TXT_{i\text{trg}}$:

$$\text{ОК}_{i\text{trg}} = \{\text{ОК}_0, \text{ОК}_1, \dots, \text{ОК}_j\},$$

где ОК_j – оценка качества для j-го варианта переведенного текста.

Каждому варианту перевода соответствует одна оценка качества перевода, то есть множества $TXT_{i\text{trg}}$ и $\text{ОК}_{i\text{trg}}$ биективны: $TXT_{i\text{trg}} \leftrightarrow \text{ОК}_{i\text{trg}}$.

$[min\text{ОК};\ max\text{ОК}]$ – диапазон значений оценок качества перевода $\text{ОК}_{i\text{trg}}$.

$\text{ОК}_{\text{доп}}$ – минимально допустимое значение критерия «Высокая оценка качества перевода» при допущении, что чем выше значение ОК_j, тем лучше.

КП_i – нечёткое подмножество множества ОК_{itrg}, определяющее принадлежность элементов множества $\text{ОК}_{i\text{trg}}$ классу «Высокая оценка качества перевода текста $txt_{i\text{ТХТ}}$»:

$$\text{КП}_i = \{(\text{ОК},\ \mu_{\text{КП}i}(\text{ОК}))|\text{ОК} \in \text{ОК}_{i\text{trg}}\}.$$

$\mu_{\text{КП}i}(\text{ОК})$ – функция принадлежности, указывающая, в какой степени текст txt с оценкой ОК принадлежит нечеткому множеству КП_i .

$\mu_{\text{КП}i}(\text{ОК}) \in [0; 1]$ и имеет вид логистической кривой:

$$\mu_{\text{КП}i}(\text{ОК}) = \frac{1}{1+e^{-\left(\frac{\text{ОК}-\text{ОК}_{\text{доп}}}{max\text{ОК}-\text{ОК}_{\text{доп}}}\right)2\pi}}. \quad (2)$$

Требуется максимизировать правдоподобие сгенерированного системой МП текста $txt_{j\text{ТХТ}}$, то есть вероятность того, что $txt_{j\text{ТХТ}}$ примет такое значение, при котором $\mu_{\text{КП}i}(\text{ОК})$ будет максимальна. Тогда логарифмическая функция правдоподобия МП $F_{\text{МП}}$ примет вид:

$$F_{\text{МП}}(\theta, \mu_{\text{КП}i}(\text{ОК})) = lnP_\theta(\max \mu_{\text{КП}i}(\text{ОК})) \to \max_\theta , \quad (3)$$

где θ – параметры системы МП из множества исполнителей перевода, или переводчиков: $\text{пер}_{i\text{ПЕР}} \in \text{ПЕР}$, максимизирующие вероятность P получить максимальное значение функции принадлежности $\mu_{\text{КП}i}(\text{ОК})$.

Решение поставленной задачи лежит в области оптимизации и совершенствования алгоритмов генерации переведенного текста.

В результате теоретического моделирования определено, что в системах МП не реализован этап переводческого процесса, который выполняется при «ручном переводе», а именно, оценка сложности задачи перевода. На этом этапе переводчик оценивает вероятность получения качественного перевода, то есть соответствующего требованиям заказчика, и, если эта вероятность низкая, выбирает стратегию оптимизации предредактирования исходного текста с целью повышения вероятности получения качественного перевода.

2. Сложность задачи перевода. Для разработки стратегии и методики оптимизационного предредактирования требуется определить критерий оптимизации исходного текста. В качестве такого критерия была выбрана оценка сложности задачи перевода. При оценке сложности задачи перевода переводчик обращает внимание на неизвестные ему слова и сочетания слов на языке $\text{яз}_{\text{вх}}$, для которых он не может идентифицировать значение смысловой единицы, либо смысловые единицы, для которых он не может найти аналог на языке перевода $\text{яз}_{\text{вых}}$ среди известных ему слов и сочетаний слов. Множества свойств и параметров исходного текста $\text{СВ}|txt_{i\text{ТХТ}}$ и $\text{ГП}|txt_{i\text{ТХТ}}$, и то, обладает ли переводчик достаточной компетенцией $\mathbf{К}_{\mathbf{пер}i\mathbf{ПЕР}}$ относительно языков $\text{яз}_{\text{вх}}$ и $\text{яз}_{\text{вых}}$ и специализацией $\mathbf{С}_{\mathbf{пер}i\mathbf{ПЕР},\ \mathbf{дп}i\mathbf{ДП}}$, т.е. навыками описания семантических единиц на языке перевода в рамках заданной предметной области исходного текста, определяет вероятность создания переводчиком переведенного текста на таком уровне качества, который определяется требованиями $\text{ТР}|txt_{i\text{ТХТ}}$.

Оценки сложности задачи перевода включает следующие шаги:

Шаг 1. Исходя из домена приложения текста $\text{дп}_{i\text{ДП}}$, формируется множество оценок текста $\text{ОЦ} = \text{СВ} \cup \text{ГП}$. Свойства текста СВ условно можно разделить на группы признаков: общие (количество символов/слов/строк и т.д., стиль, язык, домен приложения и пр.); лексические (процент покрытия текста лексическими минимумами, частотными списками и др.); морфологические (количество различных частей речи и грамматических форм); синтаксические (глубина глагольных и именных групп, связи между глаголами в предложениях); признаки, основанные на базовых подсчетах (средняя длина слов и предложений и пр.). Совокупность свойств и признаков определяет главные параметры текста ГП, к которым относится целостность, связность, удобочитаемость, сложность и другие.

Шаг 2. Для каждого значения $\text{св}_{i\text{СВ}}, \text{гп}_{i\text{ГП}} \in \text{ОЦ}$, на основе требований к переводу $\text{ТР}|txt_{iTXT}$, компетенций переводчика относительно языковой пары $\mathbf{К}_{\mathbf{пер}i\mathbf{ПЕР}}$ и специализации переводчика относительно домена приложения текста $\mathbf{С}_{\mathbf{пер}i\mathbf{ПЕР},\ \mathbf{дп}i\mathbf{ДП}}$ формируется значение значимости $w_{\text{оц}k}$, множество нормированных значений $w_{\text{оц}k}$ значимости формирует матрицу значимости оценок сложности $W_{\text{оц}}$ размерностью $1 \times k$, где k – общее число оценок, которые выступают коэффициентами уравнения поиска теоретического значения качества перевода.

Шаг 3. Для каждого i-го фрагмента текста при $i = \overline{1, N}$ формируется матрица оценок фрагмента исходного текста $C_{\text{оц}i}$ размерностью $1 \times k$, где k – общее число оценок.

Шаг 4. На основании оценок $C_{\text{оц}i}$ и значимости $W_{\text{оц}}$ формируется уравнение поиска теоретического результирующего фактора, т.е. качества перевода $\widehat{КП}$:

$$\widehat{\text{КП}}_i = w_0 + w_{\text{оц}1}\text{С}_{\text{оц}i1} + w_{\text{оц}2}\text{С}_{\text{оц}i2} + \cdots + w_{\text{оц}k}\text{С}_{\text{оц}ik}. \quad (4)$$

Для системы МП веса значимости оценок рассчитываются на основании тренировочных данных с использованием численного метода наименьших квадратов [12], при котором минимизируется сумма квадратов отклонений эмпирических значений результативного признака от теоретических, полученных по уравнению (4):

$$S(w) = \sum_{i=1}^{R}(\text{ОК}_i - \widehat{\text{ОК}}_i(\text{С}_{\text{оц}i}, w))^2,$$

$$S(w) = \sum_{i=1}^{R}(\text{ОК}_i - w_0 + w_{\text{оц}1}\text{С}_{\text{оц}i1} + w_{\text{оц}2}\text{С}_{\text{оц}i2} + \cdots + w_{\text{оц}k}\text{С}_{\text{оц}ik})^2 \to min, \quad (5)$$

где R – объем тренировочной выборки.

Для решения задачи минимизации необходимо найти стационарные точки функции $S(w)$, продифференцировав её по неизвестным параметрам w и приравняв производные к нулю

$$\sum_{i=1}^{R}(\text{ОК}_i - \widehat{\text{ОК}}_i(\text{С}_{\text{оц}i}, w))\frac{\partial \widehat{\text{ОК}}_r(\text{С}_{\text{оц}i}, w)}{\partial w} = 0. \quad (6)$$

Получаем систему нормальных уравнений с k неизвестными:

$$\begin{cases} \sum \text{ОК} = Rw_0 + w_{\text{оц}1}\sum \text{С}_{\text{оц}1} + w_{\text{оц}2}\sum \text{С}_{\text{оц}2} + \cdots + w_{\text{оц}k}\sum \text{С}_{\text{оц}k} \\ \sum \text{ОК}\cdot\text{С}_{\text{оц}1} = w_0\sum \text{С}_{\text{оц}1} + w_{\text{оц}1}\sum \text{С}_{\text{оц}1}{}^2 + w_{\text{оц}2}\sum \text{С}_{\text{оц}2}\text{С}_{\text{оц}1} + \cdots + w_{\text{оц}k}\sum \text{С}_{\text{оц}k}\text{С}_{\text{оц}1}. \\ \cdots \\ \sum \text{ОК}\cdot\text{С}_{\text{оц}k} = w_0\sum \text{С}_{\text{оц}k} + w_{\text{оц}1}\sum \text{С}_{\text{оц}1}\text{С}_{\text{оц}k} + w_{\text{оц}2}\sum \text{С}_{\text{оц}2}\text{С}_{\text{оц}k} + \cdots + w_{\text{оц}k}\sum \text{С}_{\text{оц}k}{}^2 \end{cases}$$

Решение этой системы уравнений дает нам общую формулу поиска весов значимости $W_{\text{оц}}$ в матричной форме*:*

$$W_{\text{оц}} = (\text{С}_{\text{оц}}{}^T\cdot\text{С}_{\text{оц}})^{-1}\cdot\text{С}_{\text{оц}}{}^T\cdot\text{ОК} = (\frac{1}{R}\text{С}_{\text{оц}}{}^T\cdot\text{С}_{\text{оц}})^{-1}\frac{1}{R}\text{С}_{\text{оц}}{}^T\cdot\text{ОК}. \quad (7)$$

Шаг 5. Для каждого *i-го* фрагмента текста рассчитывается вероятность получения переведенного текста на таком уровне качества, который определяется требованиями $\text{ТР}|txt_{iTXT}$, применив к уравнению (4) логит-преобразование:

$$p_i = \frac{1}{1+e^{-\widehat{\text{КП}}_i}}. \quad (8)$$

Шаг 6. Сложность задачи перевода i*-го* фрагмента текста оценивается по формуле:

$$\text{СлЗП}_i = \frac{1}{p_i}. \quad (9)$$

Шаг 7. Результирующая сложность задачи перевода текста – это наибольшее значение сложности задачи перевода СлЗП_i среди N фрагментов исходного текста, то есть

$$\text{СлЗП}_{txti\text{ТХТ}} = \max \text{СлЗП}_i \quad (10)$$

В зависимости от значения $\text{СлЗП}_{txti\text{ТХТ}}$ определяется стратегия дальнейшей обработки текста, в том числе необходимость применять оптимизационное предредактирование.

3. Задача оптимизационного предредактирования. Задача оптимизационного предредактирования состоит в том, чтобы максимизировать правдоподобие, то есть вероятность того, что при параметрах Ψ предредактора, текст $txt'_{i\text{ТХТ}}$ на языке $\text{яз}_{\text{вх}}$ будет эквивалентен $txt_{i\text{ТХТ}}$ по смыслу, понятен системе МП $\text{пер}_{i\text{ПЕР}}$ и оценка качества ОК_j перевода $txt_{j\text{ТХТ}}$ относительно txt_{iTXT} при генерации перевода из txt'_{iTXT} будет максимальной. Далее опишем задачу более подробно.

Опишем условия задачи, пусть:

1) $txt'_{i\text{ТХТ}}$ – текст на языке $\text{яз}_{\text{вх}}$, созданный системой автоматического оптимизационного предредактирования, такой, при котором $\text{СМ}'_i \to \text{СМ}_i$ и $txt'_{i\text{ТХТ}} \neq txt_{i\text{ТХТ}}$, где СМ – смысл или упорядоченный набор семантических единиц, описываемый текстом: $\text{СМ} = \{(ce_0, ce_1, \ldots, ce_{n\text{см}}): ce_{i\text{см}} \in \text{СЕ}\}$, $\text{СМ}'_i$ и СМ_i – смыслы $txt'_{i\text{ТХТ}}$ и $txt_{i\text{ТХТ}}$, соответственно;
2) $TXT_{i\text{src}}$ – множество всех возможных вариантов предредактированного текста, т.е. выражения смысла СМ_i текста $txt_{i\text{ТХТ}}$ на языке $\text{яз}_{\text{вх}}$*:*
3) $TXT_{i\text{src}} = \{txt_0, txt_1, \ldots, txt_k\}$,
4) где k – обще число вариантов предредактированного текста, причем $txt_{i\text{ТХТ}}$, $txt'_{i\text{ТХТ}} \in TXT_{i\text{src}}$*;*
5) $\text{мСлЗП}_{i\text{src}}$ – множество оценок сложности задачи перевода вариантов предредактирования текста $txt_{i\text{ТХТ}}$ в соответствии с компетентностью $\mathbf{К_{перЛПЕР}}$ и специализацией $\mathbf{С_{перЛПЕР,\ дпЙДП}}$ системы МП $\text{пер}_{i\text{ПЕР}}$ для всех возможных вариантов предредактированного текста $TXT_{i\text{src}}$*:*
6) $\text{мСлЗП}_{i\text{src}} = \{\text{СлЗП}_0, \text{СлЗП}_1, \ldots, \text{СлЗП}_k\}$;

7) Каждому варианту предредактированного текста соответствует одна оценка сложности задачи перевода для системы МП $\text{пер}_{i\text{ПЕР}}$, то есть множества $TXT_{i\text{src}}$ и $\text{мСлЗП}_{i\text{src}}$ биективны: $TXT_{i\text{src}} \leftrightarrow \text{мСлЗП}_{i\text{src}}$;
8) $\text{СлЗП}_k \in \text{мСлЗП}_{i\text{src}}$ – оценка сложности задачи перевода варианта предредактированного текста $txt'_{i\text{ТХТ}}$ для системы МП $\text{пер}_{i\text{ПЕР}}$;
9) $[min\text{СлЗП};\ max\text{СлЗП}]$ – диапазон значений оценок сложности задачи перевода $\text{мСлЗП}_{i\text{src}}$;
10) $\text{СлЗП}_{\text{доп}}$ – максимально допустимое значение критерия «Низкая сложность задачи перевода» при допущении, что чем ниже значение СлЗП_k, тем лучше;
11) нСлЗП_i – нечёткое подмножество множества $\text{мСлЗП}_{i\text{src}}$, определяющее принадлежность элементов множества $\text{мСлЗП}_{i\text{src}}$ и соответствующих элементов множества $TXT_{i\text{src}}$ классу «Низкая сложность задачи перевода»:
12) $\text{нСлЗП}_i = \{(\text{СлЗП},\ \mu_{\text{нСлЗП}i}(\text{СлЗП}))|\text{СлЗП} \in \text{мСлЗП}_{i\text{src}}\}$;
13) $\mu_{\text{нСлЗП}i}(\text{СлЗП})$ – функция принадлежности, указывающая в какой степени текст txt с оценкой СлЗП принадлежит нечеткому множеству нСлЗП_i;
14) $\mu_{\text{нСлЗП}i}(\text{СлЗП}) \in [0;\ 1]$ и имеет вид логистической кривой:

$$\mu_{\text{нСлЗП}i}\ (\text{СлЗП}) = \frac{1}{1+e^{\left(\frac{\text{СлЗП}-\text{СлЗП}_{\text{доп}}}{|min\text{СлЗП}-\text{СлЗП}_{\text{доп}}|}\right)2\pi}}. \quad (11)$$

Требуется максимизировать правдоподобие сгенерированного системой оптимизационного предредактора текста txt'_{iTXT}, то есть вероятность того, что txt'_{iTXT} примет такое значение, при котором $\mu_{\text{нСлЗП}i}(\text{СлЗП})$ будет максимальна.

В дискретном случае функция правдоподобия $F_{\text{АОПР}}(\Psi, \mu_{\text{нСлЗП}i}(\text{СлЗП}))$ – вероятность выборки $\mu_{\text{нСлЗП}i}(\text{СлЗП}) = \{\mu_0, \mu_1, \dots, \mu_l\}$ в рассматриваемой серии экспериментов будет равняться $\{\max \mu_{\text{нСлЗП}i}(\text{СлЗП})_0, \max \mu_{\text{нСлЗП}i}(\text{СлЗП})_1, \dots, \max \mu_{\text{нСлЗП}i}(\text{СлЗП})_l\}$. Эта вероятность меняется в зависимости от Ψ:

$$F_{\text{АОПР}}(\Psi, \mu_{\text{нСлЗП}i}(\text{СлЗП})) = \prod_{l=1}^{L} \text{F}_{\text{АОПР}}(\mu_{\text{нСлЗП}i}(\text{СлЗП})_l) =$$
$$= P_\Psi(\mu_0 = \max \mu_{\text{нСлЗП}i}(\text{СлЗП})_0) \cdot \dots \cdot P_\Psi(\mu_l = \max \mu_{\text{нСлЗП}i}(\text{СлЗП})_l) = \quad (12)$$
$$= P_\Psi(\mu_0 = \max \mu_{\text{нСлЗП}i}(\text{СлЗП})_0), \dots, \mu_l = \max \mu_{\text{нСлЗП}i}(\text{СлЗП})_l),$$

где l – номер экземпляра в обучающей выборке объемом L.

Тогда логарифмическая функция правдоподобия автоматического оптимизационного предредактирования $F_{\text{АОПР}}$ имеет вид:

$$L_{\text{АОПР}}(\Psi, \mu_{\text{нСлЗП}i}(\text{СлЗП})) = lnP_\Psi(\max \mu_{\text{нСлЗП}i}(\text{СлЗП})), \quad (13)$$

где Ψ – параметры системы автоматического оптимизационного предредактирования, максимизирующие вероятность P получить максимальное значение функции принадлежности $\mu_{\text{нСлЗП}i}(\text{СлЗП})$.

Поскольку $ln(y)$ монотонна, то точки максимума $F_{\text{АОПР}}(\Psi, \mu_{\text{нСлЗП}i}(\text{СлЗП}))$ и $L_{\text{АОПР}}(\Psi, \mu_{\text{нСлЗП}i}(\text{СлЗП}))$ совпадают, и оценкой максимального правдоподобия можно назвать точку максимума функции $L_{\text{АОПР}}(\Psi, \mu_{\text{нСлЗП}i}(\text{СлЗП}))$ по Ψ. Задача оптимизации, таким образом, заключается в поиске оценки максимального правдоподобия $\widehat{\Psi}$ вектора параметров Ψ, или:

$$\widehat{\Psi} = \arg\max_{\Psi} L_{\text{АОПР}}(\Psi, \mu_{\text{нСлЗП}i}(\text{СлЗП})) \quad (14)$$

Решение поставленной задачи оптимизации выполняется методом градиентного спуска (подъема) [13]. Для этого необходимо найти градиент логарифмической функции правдоподобия $L_{\text{АОПР}}(\Psi, \mu_{\text{нСлЗП}i}(\text{СлЗП}))$ – вектор, который показывает направление возрастания функции.

Учитывая, что Ψ – вектор параметров системы автоматического оптимизационного предредактирования и $\Psi = \{\psi_1, \psi_2, \dots, \psi_m\}$, где m – количество параметров модели, градиент функции $L_{\text{АОПР}}(\Psi, \mu_{\text{нСлЗП}i}(\text{СлЗП}))$ может быть найден по формуле:

$$\nabla L_{\text{АОПР}}(\Psi) = (\partial L_{\text{АОПР}}/\partial\psi_1, \partial L_{\text{АОПР}}/\partial\psi_2, \ldots, \partial L_{\text{АОПР}}/\partial\psi_m), \quad (15)$$

где $\partial L_{\text{АОПР}}/\partial\psi_m$ – частная производная функции правдоподобия по m-ному параметру.

Обновление параметров Ψ происходит итеративно для каждого $\psi_m \in \Psi$:

$$\Psi^{[s+1]} = \Psi^{[s]} + \alpha \cdot \nabla L_{\text{АОПР}}(\Psi^{[s]}), \quad (16)$$

где s – шаг оптимизации, $s \in [0; S]$ и S – общее число шагов оптимизации, а $\Psi^{[0]}$ – начальное приближение параметров модели; α – скорость обучения, т.е. положительное число, определяющее размер шага на каждой итерации.

Для оценки сходимости используется евклидова норма градиента функции $\nabla L_{\text{АОПР}}(\Psi)$:

$$\|\nabla L_{\text{АОПР}}(\Psi)\| = \sqrt{\left(\frac{\partial L_{\text{АОПР}}}{\partial\psi_1}\right)^2 + \left(\frac{\partial L_{\text{АОПР}}}{\partial\psi_2}\right)^2 + \ldots + \left(\frac{\partial L_{\text{АОПР}}}{\partial\psi_m}\right)^2}. \quad (17)$$

Уменьшение нормы градиента указывает на сходимость оптимизации. Если норма градиента не снижается, это свидетельствует о медленной сходимости и необходимости изменения параметров оптимизации, например, скорости обучения α.

Оптимизация выполняется, пока норма градиента не достигла заданной точности ε, критерий остановки:

$$\left\|\nabla L_{\text{АОПР}}(\Psi^{[s]})\right\| \le \varepsilon. \quad (18)$$

Используя описанную математическую модель, создадим модель автоматического оптимизационного предредактирования.

4. Обучающие данные для модели оценки сложности задачи перевода. Для обучения модели оценки сложности задачи перевода заданного текста $txt_{i\text{ТХТ}}$ с языка $\text{яз}_{\text{вх}}$ на язык $\text{яз}_{\text{вых}}$ переводчиком $\text{пер}_{i\text{ПЕР}}$ в соответствии с формализованными требованиями к переводу $\text{ТР}|txt_{i\text{ТХТ}}$ требуются исходные данные в виде корпуса параллельных текстов следующей структуры:

$TranslatorExpCor$: $[src;\ trg;\ ref]$, где src – это оригинал, т.е. текст на языке $\text{яз}_{\text{вх}}$, trg – это перевод (текст на языке $\text{яз}_{\text{вых}}$), выполненный переводчиком или системой МП $\text{пер}_{i\text{ПЕР}}$; ref – это контрольный перевод (текст на языке $\text{яз}_{\text{вых}}$), т.е. проверенный эталон.

Путем обработки корпуса $TranslatorExpCor$ модулями оценки качества МП и препроцессинга текстовых данных для взвешенной оценки параметров русскоязычного текста формируется база данных структурного анализа предложений, содержащая расчет вещественных параметров по морфологическим, синтаксическим, лексическим и прочим признакам [14].

5. Обучающие данные для модели автоматического оптимизационного предредактирования. Для обучения модели, которая будет преобразовывать текст на языке $\text{яз}_{\text{вх}}$ в текст требуемой структуры для повышения качества перевода, необходимо создать корпус тренировочных текстов в паре $\text{яз}_{\text{вх}} - \text{яз}_{\text{вх}}$. Для оптимизации временных затрат на подготовку исходных данных для тренировки модели предредактирования текста предлагается методика с использованием обратного перевода для генерирования эталонного предредактированного текста.

Структура параллельного корпуса исходных данных: $RefCor$: [src_ref; tgt_ref], где src_ref – это оригинал, т.е. текст на языке $\text{яз}_{\text{вх}}$, tgt_ref – это перевод (текст на языке $\text{яз}_{\text{вых}}$).

Методика сбора корпуса обучающих текстов для модели оптимизационного предредактирования включает следующие шаги:

1. Настраиваем системы МП MT: tgt − src, MT: src − tgt.

При помощи системы MT: tgt − src переводим текст tgt_ref на язык $\text{яз}_{\text{вх}}$, получим массив текстовых данных pre_src (массив условно предредактированных текстов).

При помощи системы MT: src − tgt переводим текст src_ref на язык $\text{яз}_{\text{вых}}$, получим массив текстовых данных tgt1 (src_ref → tgt1).

При помощи системы MT: src − tgt переводим текст pre_src на язык $\text{яз}_{\text{вых}}$, получим массив текстовых данных tgt2 (src_ref → tgt1).

Оцениваем качество выполненного перевода на язык $\text{яз}_{\text{вых}}$ tgt1 и tgt2 относительно эталона tgt_ref, получаем массивы оценок QC_score(tgt1) и QC_score(tgt2).

Для дальнейшей работы отберем тренировочный корпус TrainCor, включающий пары src_ref_i и pre_src_i, для которых наблюдается повышение оценки качества перевода на английский язык при применении предредактирования и при условии, что $\Delta\text{QC_score}_i$ является условно значимой $d_{\max}$ для выбранного типа оценки:

$$\text{TrainCor} = \{(\text{src}_{\text{ref}_i};\ \text{pre}_{\text{src}_i}) : \exists\ (\text{tgt1}_i, \text{tgt2}_i) | \text{QC_score}(\text{tgt2}_i) > \\ > \text{QC_score}(\text{tgt1}_i)\ \&\ \Delta\text{QC_score}_i \geq d_{\max}\} \quad (19)$$

Полученный корпус обучающих текстов TrainCor объемом выборки L будем использовать для обучения языковой модели LM: $\text{src} - \text{pre_src}$ для решения задачи автоматического оптимизационного редактирования текстов на языке $\text{яз}_{\text{вх}}$.

6. Оценка качества выполненного перевода. Для реализации программного комплекса определим критерий качества МП QC_score. Традиционно, для оценки качества перевода используются алгоритмы, которые сравнивают выполненный перевод с одним или несколькими эталонными переводами при помощи числовой метрики. В рамках исследования была выбрана метрика hLEPOR [15], которая имеет наивысший балл корреляции Пирсона с человеческими суждениями по языковой паре английский-русский.

$hLEPOR$ – гармоническое среднее между штрафом за длину перевода в сравнении с эталоном, штрафом за различие в позициях и степенью перекрытия n-грамм; рассчитывается в диапазоне от 0 до 1, где 0 – полое несовпадение; 1 – полное совпадение:

$$hLEPOR = \frac{(LP \cdot PP \cdot OR)^{\frac{1}{3}}}{(\frac{1}{LP} + \frac{1}{PP} + \frac{1}{OR})^{\frac{1}{3}}} \quad (20)$$

где LP — штраф за длину, который учитывает различие в длине между переводом и эталонным переводом; PP — штраф за различие в позициях, который учитывает различие в расположении слов в переводе и эталоне; OR — степень перекрытия, которая учитывает совпадение n-грамм в переводе и эталоне.

Для расчета каждого из компонентов используются следующие формулы:

$$LP = f(x) = \begin{cases} e^{1-\frac{ref}{cand}}, & cand > ref \\ 1, & cand \leq ref \end{cases}, \quad (21)$$

где $cand$ — длина переведенного текста, а ref — длина эталонного перевода;

$$PP = \frac{1}{1+p_{diff}}, \quad (22)$$

где p_{diff} — среднее арифметическое различий позиций одинаковых n-грамм в переводе и эталоне;

$$OR = BP \cdot ng_{prec} \cdot (1 - pen), \quad (23)$$

где BP — штраф за слишком короткий перевод, pen — дополнительный штраф за ошибки, ng_{prec} — точность n-грамм, которая рассчитывается как отношение количества совпадающих n-грамм в переводе и эталоне к общему количеству n-грамм в переводе.

6. Программный комплекс оптимизационного предредактирования русскоязычных текстов. Опираясь на описанные математические модели и методики, была разработана архитектура программного комплекса, который состоит из трех основных подсистем: подсистемы тренировки языковой модели, подсистемы оценки сложности задачи перевода, подсистемы оптимизационного предредактирования русскоязычного текста и генерации МП на английский язык. При реализации приняты следующие допущения:

1. Критерий качества перевода должен быть четко определен и формализован с возможностью получения вещественного нормированного значения. Могут применяться любые метрики оценки качества в зависимости от требований к качеству перевода.

2. Для тестирования МП необходим тренировочный корпус, включающий тексты на языке оригинала и перевод, принятый за эталон. В компаниях, внедривших ISO 17100 и CAT, процесс накопления тренировочных корпусов, включающих исходный текст, перевод, выполненный системой МП и проверенный перевод, утвержденный редактором, происходит автоматически в режиме реального времени.
   Схематично архитектура программного комплекса представлена на рисунке 2.

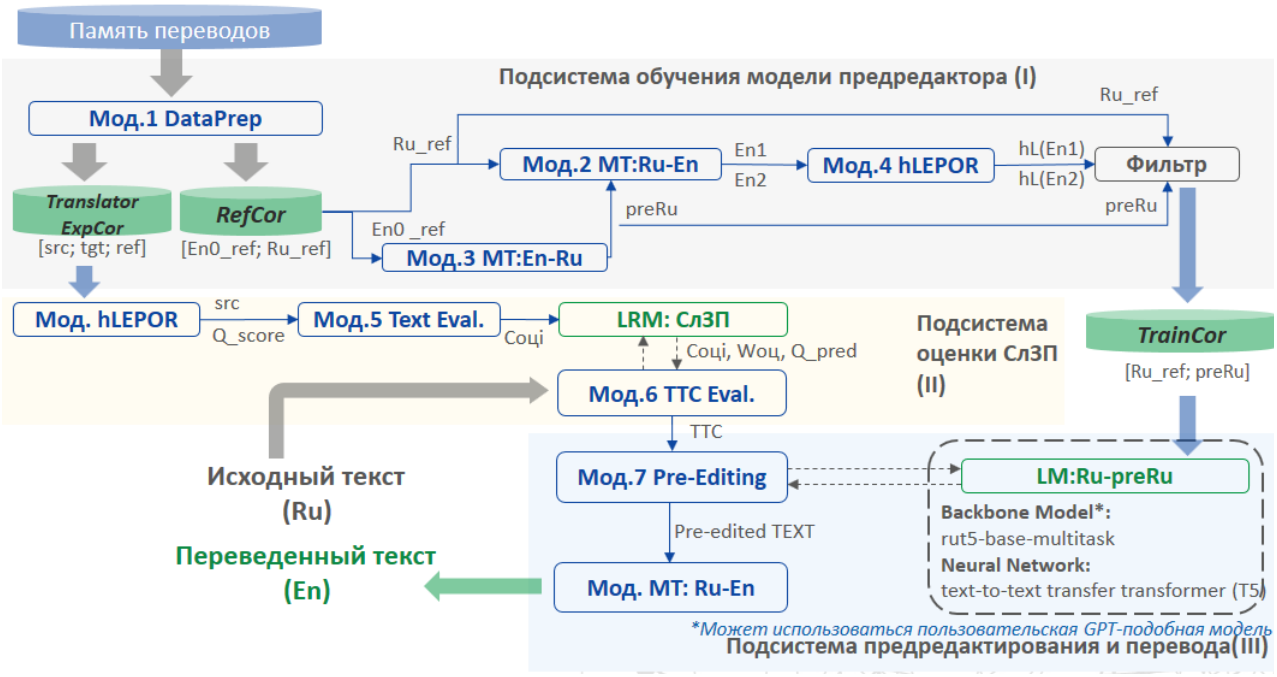


Рис. 2. Архитектура программного комплекса оптимизационного предредактирования русскоязычных текстов

Подсистема тренировки языковой модели оптимизационного предредактирования русскоязычных текстов (I) состоит из пяти программных компонентов: *Мод.1 DataPrep* – модуля подготовки «сырых данных», полученных из памятей переводов «Translation Memories» поставщика лингвистических услуг, который формирует корпуса TranslatorExpCor и RefCor для тренировки моделей оценки сложности задачи перевода и оптимизационного предредактирования соответственно; *Мод.2 MT:Ru-En, Мод.3 MT:En-Ru* – модулей МП (генератор перевода в языковой паре русский-английский и генератор в языковой паре английский-русский); *Мод.4 hLEPOR* – модуля оценки качества МП реализует алгоритм по метрике hLEPOR; модуля фильтрации данных, подходящих для тренировки модели, в котором производится отбор по условию повышения оценки качества после предредактирования. В результате обработки эталонного корпуса RefCor модулями системы 1 – 3 полученный тренировочный корпус TrainCor используется для обучения модели оптимизационного предредактирования *LM:Ru-preRu*.

Подсистема оценки сложности задачи перевода (II) состоит из трех модулей: модуля оценки качества перевода, выполненного системой МП, относительно эталонного по метрике hLEPOR; *Мод.5 Text Eval.* – препроцессора для взвешенной оценки свойств русскоязычного текста, включая морфологические, синтаксические, лексические и другие, всего 96 параметров; *Мод.6 TTC Eval.* – модуля оценки сложности задачи перевода с применением модели логистической регрессии *LRM:СлЗП*, которая определяет ожидаемое качество перевода полученного текста на основании взвешенной оценки его свойств и весов значимости оценок относительно системы МП.

Подсистема оптимизационного предредактирования и генерации МП (III) состоит из двух модулей: *Мод.7 Pre-editing* – модуля автоматического предредактирования русскоязычных текстов на основе модели *LM:Ru-preRu*, которая в качестве опорной использует модель

русского языка rut5-base-multitask на основе нейронной сети типа *text-to-text transfer transformer* (T5), представленной командой Google в 2020 г., дообученной на корпусе параллельных русскоязычных текстов подзадачу перефразирования [16]; модуля генерации МП с русского языка на английский язык на основе модели *Helsinki-NLP/opus-mt-ru-en*.

7. Тестирование программного комплекса. Программный комплекс [17] реализован на языке Python, модули обучения моделей *LRM:СлЗП* и *LM:Ru-preRu* развернуты в Центре коллективного пользования научным оборудованием «Центр обработки и хранения научных данных ДВО РАН», созданного и функционирующего на базе ВЦ ДВО РАН – обособленном подразделении ХФИЦ ДВО РАН. Тип используемой ЭВМ: компьютер с архитектурой x86, x86_64.

Исходные данные для обучения моделей и тестирования в виде русскоязычных узкоспециальных технических текстов, переведенных на английский язык, предоставлены ведущей компанией по оказанию лингвистических услуг в области технического перевода в Хабаровском крае ООО «Агентство переводов «ФИАС-Амур» в объеме ~60 000 ст. стр. текста (1 ст. стр. = 1800 знаков с пробелами). Из предоставленных данных сформированы корпуса RefCor объемом ~140 000 экземпляров и TranslatorExpCor объемом ~90 000 экземпляров. После обработки корпуса RefCor в подсистеме тренировки языковой модели оптимизационного предредактирования русскоязычных текстов в корпус TrainCor вошло ~84 000 экземпляров. Тестовая выборка для оценки работы системы TestCor составила ~17 000 экземпляров.

Тестирование программного комплекса включало следующие этапы:

1. Машинный перевод тестовой выборки и оценка его качества алгоритмом hLEPOR → получение оценки $hLEPOR(En1)$.
2. Оценка сложности задачи перевода тестовой выборки $\text{СлЗП}(Ru)$ → применение оптимизационного предредактирования только к тем семплам, для которых $\text{СлЗП}(Ru) > \text{СлЗПдоп}$ → оценка СлЗП после оптимизационного предредактирования $\text{СлЗП}(preRu)$ → машинный перевод → получение оценки $hLEPOR\ (TTC/PE)$.
3. Сравнение и анализ полученных результатов.

В рамках тестирования принято, что минимально допустимая сложность задачи перевода $\text{СлЗП}_{\text{доп}} = 1{,}43$. Так как сложность задачи перевода обратно пропорциональна вероятности получения перевода требуемого качества, при $\text{СлЗП}_{\text{доп}} = 1{,}43$ данная вероятность составляет 0,7.

С использованием оценки сложности задачи перевода из тестовой выборки было отобрано 5440 экземпляров для оптимизационного предредактирования (32,58% тестовой выборки).

Примеры оптимизационного предредактирования и его влияния на сложность задачи перевода для системы МП представлены в таблице 1. Результаты перевода текстов после оптимизационного предредактирования на английский язык представлены в таблице 2.

Таблица 1. Оптимизационное предредактирование русскоязычных текстов

Экз.	**Исходный текст (Ru_ref)**	**Текст после предредактирования (preRU)**	**ΔСлЗП**
1	Высоковольтные испытания проводятся по отдельно разрабатываемой и утверждаемой «Программе проведения высоковольтных испытаний кабеля 110 кВ».	Испытания на высоковольтные кабели проводятся в соответствии с отдельно разработанной и утвержденной программой испытаний высоковольтных кабелей 110 кВ.	-1,356
2	Оборудование должно быть рассчитано на двойные фидеры, а если такое оборудование отсутствует, в центральном шкафу предусматривают установку контроллера автоматического ввода резерва.	Оборудование должно быть способно управлять двумя фидерами, в случае отсутствия такого оборудования в центральном шкафу должен быть установлен переключатель ввода резерва.	-0,444

3	ТУ на поставку включают в себя, помимо прочего, следующее:	Спецификация покупки должна содержать и не ограничиваться:	-0,167
4	По результатам месяца подготовка отчета (10 число) по отклонениям от намеченного графика.	Отчет о ходе месяца (10 числа) об отклонениях от плана работы.	-0,271

Таблица 2. Результаты машинного перевода на английский язык

Экз.	**МП исходного текста (En1)**	**МП после предредактирования (En2)**	**ΔhLEPOR**
1	The high voltage tests are conducted on a separate design and approval of the 110 kV high voltage test programme.	High voltage cables shall be tested according to a separately developed and approved 110 kV high voltage cables programme.	0,271
2	The equipment shall be designed for double feeders, and if such equipment is not available, an automatic backup controller shall be installed in the central cabinet.	The equipment shall be capable of controlling two feeders, in the absence of such equipment, a standby switch shall be installed in the central cabinet.	0,146
3	TA for supply includes, inter alia, the following:	The purchase specification shall contain and not be limited to:	0,135
4	Based on the month's results, the report (10 times) is based on deviations from the schedule.	Monthly progress report (10th) on deviations from the workplan.	0,131

Результаты оценки качества после оптимизационного предредактирования представлены в таблице 3.

Таблица 3. Результаты применения оптимизационного предредактирования

	h_LEPOR (En1)	**h_LEPOR (En2)**	**ΔhLEPOR**	**СлЗП (Ru)**	**СлЗП (preRu)**
mean	0,508044336	0,583044424	0,075	2,111294	1,075499
std	0,140363512	0,128616684	0,0739	0,75958	0,052124
min	0	0,10348	1E-05	1,431731	1
25%	0,4253	0,506915	0,02056	1,692549	1,035022
50%	0,52027	0,59196	0,05472	1,928669	1,067745
75%	0,60549	0,67087	0,104235	2,314225	1,107176
max	0,99834	1	0,58203	20,19173	1,310526

Принятые обозначения: mean – математическое ожидание; std – среднеквадратичное отклонение; min – минимальное значение выборки; 25% – значение, меньше которого 25% значений выборки; 50% – медиана, т.е. значение, меньше и больше которого 50% значений выборки; 75% – значение, меньше которого 75% значений выборки; max – максимальное значение в выборке.

Показано, что качество перевода отдельных сегментов, подвергшихся оптимизационному предредактированию, в среднем, возросло на 15%. Максимальное повышение качества перевода составило 30% в отдельных сегментах.

Использование полученных результатов исследования и внедрение программного комплекса в работу агентства переводов «ФИАС-Амур» позволило повысить эффективность использования систем МП и производительность труда редакторов переводов, а также оптимизировать затраты на оказание услуг перевода узкоспециальной технической документации.

В ходе внедрения за время мониторинга при помощи программного комплекса было переведено на английский язык 677 стандартных страниц текста. При этом, средняя производительность редакторов переводов при работе с программным комплексом увеличилась с 3,8 до 4,3 стандартных страницы в час. Таким образом, внедрение программного комплекса позволило увеличить производительность редакторов переводов на 13,16%, что является значимым показателем при переводе большого объема документации в условиях дефицита квалифицированных кадров.

Описанные модели и методы могут быть масштабированы на различные языковые пары и способы перевода, включая ручной перевод, они намечают подходы к управлению рисками, связанными с качеством перевода в зависимости от компетенции выбранных исполнителей, и предоставит индустрии инструмент объективной оценки исполнителей в рамках поставленной задачи на перевод.

Программный комплекс может быть внедрен в компаниях, генерирующих от 1000 страниц перевода в месяц, предоставив инструментарий повышения качества перевода, в том числе для редакторов без знания языка перевода.

Заключение. В ходе работы впервые предложена методика оценки сложности переводческой задачи для переводчика на основе его компетенции и специализации и параметров исходного текста, которая позволяет прогнозировать риски некачественного и/или несвоевременного решения задачи перевода; предложена новая методика для повышения качества машинного перевода текстов с русского языка на английский язык, отличающаяся от существующих применением обратного перевода для сбора тренировочных данных и оптимизационного предредактирования на основе вероятностной оценки сложности задачи перевода.

Программный комплекс представляет собой ансамбль алгоритмов и моделей, включая модель классификации и генеративную модель русского языка, каждая из которых имеет потенциал к доработке с целью повышения точности, что позволит улучшить эффективность программного комплекса в целом.

Результаты исследования подтверждают эффективность применения оптимизационного предредактирования русскоязычных узкоспециальных текстов с целью повышения качества МП на английский язык. Разработанный авторами программный комплекс оптимизационного предредактирования имеет потенциал к доработке и повышении точности. Интеграция программного комплекса в контур автоматизации процессов перевода технической документации позволяет снизить затраты на постредактирование МП и организацию переводческих процессов.

Благодарности. Работа выполнена в рамках соглашения о стратегическом научно-технологическом сотрудничестве между ФГБОУ ВО «КнАГУ» (г. Комсомольск-на-Амуре), Центром обработки и хранения научных данных ДВО РАН на базе ВЦ ДВО РАН – обособленного подразделения ХФИЦ ДВО РАН (г. Хабаровск) и ООО «Агентство переводов «ФИАС-Амур» (г. Комсомольск-на-Амуре).

Список источников

1. Quinci C., Pontrandolfo G. Testing neural machine translation against different levels of specialization. Transkom, 2023, vol.1, pp.174-209.
2. Canfora C., Ottmann A. Risks in neural machine translation. Translation spaces, 2020, vol. 9(1), pp. 58–77.
3. Kumar V., Kulkarni A., Singh P., Ramakrishnan G. A machine assisted human translation system for technical documents. Miami MT Summit XV, 2015, vol.2, pp. 259-272.
4. Hiraoka Y., Yamada M. Pre-editing plus neural machine translation for subtitling: effective pre-editing rules for subtitling of TED Talks. MT Summit XVII, Dublin, Ireland, 2019, vol.2, pp. 64-74.
5. Miyata R., Fujita A. Dissecting human pre-editing toward better use of off-the-shelf machine translation systems. Proceedings of the 20th Annual Conference of the European association for machine translation (EAMT), User studies papers, Prague, Czech Republic, 2017.
6. Taufik A. Pre-editing of Google neural machine translation. Journal of English language and culture, 2020, vol. 10, no. 2, pp. 64-74.

7. O'Brien, S. Controlling controlled English: an analytical of several controlled language rule sets. Proceedings of EAMT-CLAW, Dublin, Ireland, 2003, pp. 105-114.
8. Оборнева И. В. Автоматизация оценки качества восприятия текста / И. В. Оборнева // ВЕСТНИК Московского городского педагогического университета, 2015. – №2(5). – С. 221–233.
9. Дмитриева А.Д. Квантитативное исследование стратегий упрощения на материале адаптированных текстов для изучающих РКИ / А.Д. Дмитриева, А.Н. Лапошина, М.Ю. Лебедева // Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог», 2021. – С. 191-204.
10. Zhivotova A.A., Berdonosov V.D., Gordin S.A. Mathematical modeling of the translation process and its optimization by the criterion of quality maximization. Information Technologies and intelligent decision-making systems: communications in computer and information science, 2023, vol. 1821, pp. 1-15.
11. Barber D. Bayesian Reasoning and machine learning. Cambridge: Cambridge university press, 2012, DOI:10.1017/CBO9780511804779.
12. MacKay D.J.C. Information theory, inference, and learning algorithms. Cambridge: Cambridge University Press, 2003, DOI: 10.2277/0521642981
13. Bishop, C.M. Pattern Recognition and Machine Learning. Springer, Berlin, 2006.
14. Животова А.А. Регрессионный анализ корреляции качества машинного перевода и параметров исходного текста / Животова А.А., Бердоносов В.Д. // Информатика и системы управления, 2023. – №2(76). – С.121-133.
15. Li-Feng Han A., Wong D.F., Chao L.S. et al. Language-independent model for machine translation evaluation with reinforced factors. Proceedings of machine translation Summit XIV: Posters, Nice, France, 2013, pp. 215-222.
16. Дале Д. Многозадачная модель T5 для русского языка. – URL: https://habr.com/ru/articles/581932/ (дата обращения: 20.08.2023).
17. Свидетельство о государственной регистрации программы для ЭВМ № 2023682260 Российская Федерация. Программный комплекс для предредактирования и машинного перевода узкоспециальных русскоязычных текстов на английский язык: № 2023680875: заявл. 09.10.2023: опубл. 24.10.2023 / А. А. Животова, В. Д. Бердоносов; заявитель Животова А.А. – 1 с.

Животова Алена Анатольевна. *Аспирант, ассистент кафедры Прикладная математика факультета компьютерных технологий Комсомольского-на-Амуре государственного университета, ORCID: 0000-0003-1037-5503, SPIN: 3142-9290, РИНЦ AuthorID: 771774, zhivotova.aa@gmail.com, г. Комсомольск-на-Амуре, пр. Ленина, д. 27.*

Бердоносов Виктор Дмитриевич. *К.т.н., доцент кафедры Прикладная математика факультета компьютерных технологий Комсомольского-на-Амуре государственного университета, ORCID: 0000-0003-4093-779X, SPIN: 8730-2226, РИНЦ AuthorID: 644399, berd1946@gmail.com, г. Комсомольск-на-Амуре, пр. Ленина, д. 27.*

UDC 004.9:519.768
DOI:10.25729/ESI.2024.34.2.016

Optimizational pre-editing of highly specialized Russian-language texts for its machine translation into English

Alena A. Zhivotova, Victor D. Berdonosov

Komsomolsk-na-Amure State University,
Russia, Komsomolsk-na-Amure, *zhivotova.aa@gmail.com*

Abstract. The authors study the process of text translation, particularly the method of optimizing pre-editing as a way to improve the quality of machine translation into English for Russian-language highly specialized texts. The paper considers the mathematical model of translation process and machine translation task formulation, proposes a new theory for probabilistic estimation of translation task complexity, provides the formulation and solution of optimizing pre-editing task, describes data preparation methodology for training automatic optimizing pre-editing model. As the research result the software package for optimizing pre-editing of Russian-language texts is developed. The software package has been developed using resources of the Center for Scientific Equipment Collective Use "Center for Processing and Storage of Scientific Data of the Far Eastern Branch of the Russian Academy of Sciences". Data for models training and validation are provided by Translation Agency FIAS-Amur Co., Ltd. Software package testing has proved the effectiveness of the proposed methods for improving the quality of machine translation of highly specialized Russian-language texts into English.

Keywords: machine translation, optimizational pre-editing, translation task complexity, translation quality

Acknowledgements: the research was conducted within the agreement on strategic scientific and technological cooperation between Komsomolsk-na-Amure State University (Komsomolsk-na-Amure), Center for Processing and Storage of Scientific Data of the Far Eastern Branch of the Russian Academy of Sciences (Khabarovsk), and Translation Agency FIAS-Amur Co., Ltd. (Komsomolsk-na-Amure).

References

1. Quinci C., Pontrandolfo G. Testing neural machine translation against different levels of specialization. Transkom, 2023, vol.1, pp.174-209.
2. Canfora C., Ottmann A. Risks in neural machine translation. Translation spaces, 2020, vol. 9(1), pp. 58–77.
3. Kumar V., Kulkarni A., Singh P., Ramakrishnan G. A machine assisted human translation system for technical documents. Miami MT Summit XV, 2015, vol.2, pp. 259-272.
4. Hiraoka Y.,Yamada M. Pre-editing plus neural machine translation for subtitling: effective pre-editing rules for subtitling of TED Talks. MT Summit XVII, Dublin, Ireland, 2019, vol.2, pp. 64-74.
5. Miyata R., Fujita A. Dissecting human pre-editing toward better use of off-the-shelf machine translation Systems. Proceedings of the 20th Annual Conference of the European association for machine translation (EAMT), User studies papers, Prague, Czech Republic, 2017.
6. Taufik A. Pre-editing of Google neural machine translation. Journal of English language and culture, 2020, vol. 10., no. 2, pp. 64-74.
7. O'Brien, S. Controlling controlled English: an analytical of several controlled language rule sets. Proceedings of EAMT-CLAW, Dublin, Ireland, 2003, pp. 105-114.
8. Oborneva, I. V. Avtomatizaciya ocenki kachestva vospriyatiya teksta [Automation of text perception quality assessment]. VESTNIK Moskovskogo gorodskogo pedagogicheskogo universiteta [Herald of Moscow City Pedagogical University], 2015, no.2(5), pp. 221-233.
9. Dmitrieva A.D., Laposhina A., Lebedeva M. Kvantitativnoe issledovanie strategij uproshheniya na materiale adaptirovanny`x tekstov dlya izuchayushhix RKI [A Quantitative Study of Simplification Strategies in Adapted Texts for L2 Learners of Russian]. Komp`yuternaya lingvistika i intellektual`ny`e texnologii: po materialam mezhdunarodnoj konferencii "Dialog" [Computer Linguistics and Intelligent Technologies: Proceedings of the International Conference "Dialog"], 2021, pp. 191-204.
10. Zhivotova A.A., Berdonosov V.D., Gordin S.A. Mathematical modeling of the translation process and its optimization by the criterion of quality maximization. Information Technologies and intelligent decision-making systems: communications in computer and information science, 2023, vol. 1821, pp. 1-15.
11. Barber D. Bayesian Reasoning and machine learning. Cambridge: Cambridge university press, 2012, DOI:10.1017/CBO9780511804779.
12. MacKay D.J.C. Information theory, inference, and learning algorithms. Cambridge: Cambridge University Press, 2003, DOI: 10.2277/0521642981
13. Bishop, C.M. Pattern Recognition and Machine Learning. Springer, Berlin, 2006.
14. Zhivotova A.A., Berdonosov V.D. Regressionnyj analiz korrelyacii kachestva mashinnogo perevoda i parametrov iskhodnogo teksta [Regression analysis of the correlation between machine translation quality and source text features]. Informatika i sistemy upravleniya [Information science and control systems], 2023, no.2(76), pp.121-133.
15. Li-Feng Han A., Wong D. F., Chao L.S. et al. Language-independent model for machine translation evaluation with reinforced factors. Proceedings of machine translation Summit XIV: Posters, Nice, France, 2013, pp. 215-222.
16. Dale, D. Mnogozadachnaya model` T5 dlya russkogo yazy`ka [T5 multitasking model for Russian language]. Available at: https://habr.com/ru/articles/581932/ (accessed: 08/20/2023).
17. Zhivotova A.A., Berdonosov V.D. Programmny`j kompleks dlya predredaktirovaniya i mashinnogo perevoda uzkospecial`ny`x russkoyazy`ch-ny`x tekstov na anglijskij yazy`k [Program complex for pre-editing and machine translation of highly specialized Russian-language texts into English]. Software Registration Certificate of the Russian Federation no. 2023680875 (2023).

Alena Anatolievna Zhivotova. *Postgraduate student, assistant of Applied Mathematics department, faculty of computer technologies, Komsomolsk-na-Amure state university, ORCID: 0000-0003-1037-5503, SPIN: 3142-9290, AuthorID: 771774, zhivotova.aa@gmail.com, Russia, Komsomolsk-na-Amure, 27 Lenin Ave.*

Victor Dmitrievich Berdonosov. *PhD in technical sciences, associate professor of Applied Mathematics department, faculty of computer technologies, Komsomolsk-na-Amure state university, ORCID: 0000-0003-4093-779X, SPIN: 8730-2226, AuthorID: 644399, berd1946@gmail.com, Russia, Komsomolsk-na-Amure, 27 Lenin Ave.*

Статья поступила в редакцию 08.12.2023; одобрена после рецензирования 29.03.2024; принята к публикации 06.06.2024.

The article was submitted 12/08/2023; approved after reviewing 03/29/2024; accepted for publication 06/06/2024.