

# Программные продукты и системы

DOI 10.15827/0236-235X

/ Об эффективности  
алгоритма саранчи

/ Сравнительный  
анализ  
методов сжатия  
табличных данных

/ Реализация  
программного  
комплекса  
психодиагностики

/ О нейросетевом  
методе оценки  
потерь  
электроэнергии



№ 2

ТОМ 37

2024

---

Научно-исследовательский институт  
«Центрпрограммсистем»

---

# Программные продукты и системы

НАУЧНЫЙ ЖУРНАЛ

2024, том 37, № 2  
(год издания тридцать седьмой)

**Главный редактор**  
**Г.И. САВИН**, академик РАН

## SOFTWARE & SYSTEMS

Research Journal

2024, vol. 37, no. 2

**Editor-in-Chief**  
**G.I. SAVIN**, Academician of the Russian Academy of Sciences

---

Research Institute CENTERPROGRAMSYSTEM

---

## © ПРОГРАММНЫЕ ПРОДУКТЫ И СИСТЕМЫ

Научный журнал

2024. Т. 37. № 2

DOI: 10.15827/0236-235X.142

### Главный редактор

**Г.И. САВИН**, академик РАН

### Научные редакторы номера:

**Б.М. ШАБАНОВ**, д.т.н., чл.-корр. РАН

**С.В. УЛЬЯНОВ**, д.ф.-м.н., профессор

Издатель НИИ «Центрпрограммсистем»  
(г. Тверь, Россия)

Учредитель **В.П. Куприянов**

Журнал зарегистрирован в Роскомнадзоре  
3 марта 2020 г.

Регистрационное свидетельство ПИ № ФС 77-77843

Подписной индекс в каталоге

Урал-Пресс **70799**

ISSN 0236-235X (печатн.)

ISSN 2311-2735 (онлайн)

## РЕДАКЦИОННАЯ КОЛЛЕГИЯ

**Семенов Н.А.** – заместитель главного редактора, д.т.н., профессор Тверского государственного технического университета (г. Тверь, Россия)

**Сотников А.Н.** – заместитель главного редактора, д.ф.-м.н., профессор, заместитель директора Межведомственного суперкомпьютерного центра РАН (г. Москва, Россия)

**Афанасьев А.П.** – д.ф.-м.н., профессор Московского физико-технического института (технического университета), заведующий Центром распределенных вычислений Института проблем передачи информации РАН (г. Москва, Россия)

**Баламетов А.Б.** – д.т.н., профессор Азербайджанского научно-исследовательского и проектно-изыскательского института энергетики (г. Баку, Азербайджан)

**Борисов В.В.** – д.т.н., профессор филиала Национального исследовательского университета «МЭИ» в г. Смоленске (г. Смоленск, Россия)

**Голенков В.В.** – д.т.н., профессор Белорусского государственного университета информатики и радиоэлектроники (г. Минск, Беларусь)

**Елизаров А.М.** – д.ф.-м.н., профессор Института математики и механики им. Н.И. Лобачевского Казанского федерального университета (г. Казань, Россия)

**Еремеев А.П.** – д.т.н., профессор Национального исследовательского университета «МЭИ» (г. Москва, Россия)

**Кузнецов О.П.** – д.т.н., профессор Института проблем управления РАН (г. Москва, Россия)

**Мамросенко К.А.** – к.т.н., доцент Московского авиационного института (Национального исследовательского университета), руководитель Центра визуализации и спутниковых информационных технологий НИИСИ РАН (г. Москва, Россия)

**Палюх Б.В.** – д.т.н., профессор Тверского государственного технического университета (г. Тверь, Россия)

**Сулейманов Д.Ш.** – академик АН Республики Татарстан, д.т.н., профессор Казанского государственного технического университета (г. Казань, Россия)

**Татарникова Т.М.** – д.т.н., доцент, профессор Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В.И. Ульянова (Ленина) (г. Санкт-Петербург, Россия)

**Ульянов С.В.** – д.ф.-м.н., профессор, ведущий научный сотрудник Объединенного института ядерных исследований (г. Дубна, Россия)

**Хорошевский В.Ф.** – д.т.н., профессор Московского физико-технического института (технического университета) (г. Москва, Россия)

**Шабанов Б.М.** – д.т.н., чл.-корр. РАН, директор Межведомственного суперкомпьютерного центра РАН (г. Москва, Россия)

**Язенин А.В.** – д.ф.-м.н., профессор Тверского государственного университета (г. Тверь, Россия)

## АССОЦИИРОВАННЫЕ ЧЛЕНЫ РЕДАКЦИИ

*Национальный исследовательский университет «МЭИ», г. Москва, Россия*

*Технологический институт Южного федерального университета, г. Таганрог, Россия*

*Тверской государственный технический университет, г. Тверь, Россия*

### АДРЕС ИЗДАТЕЛЯ И РЕДАКЦИИ

Россия, 170024,

г. Тверь, просп. Николая Корыткова, д. 3а

Телефон (482-2) 39-91-49

Факс (482-2) 39-91-00

E-mail: [red@cps.tver.ru](mailto:red@cps.tver.ru)

Сайт: [www.swsys.ru](http://www.swsys.ru)

Дата выхода в свет 16.06.2024 г.

Отпечатано ИПП «Фактор и К»

Россия, 170100, г. Тверь, ул. Крылова, д. 26

Выпускается один раз в квартал

Год издания тридцать седьмой

Формат 60×84 1/8. Объем 164 стр.

Заказ № 5. Тираж 1000 экз. Цена 550,00 руб.

© SOFTWARE & SYSTEMS

Research Journal

2024, vol. 37, no. 2

DOI: 10.15827/0236-235X.142

**Editor-in-chief**

**G.I. SAVIN**, Academician of RAS

**Science editors of the issue:**

**B.M. Shabanov**, Dr.Sc. (Engineering),

Corresponding Member of the RAS

**S.V. Ulyanov**, Dr.Sc. (Physics and Mathematics), Professor

**Publisher Research Institute**

**CENTERPROGRAMSYSTEM** (Tver, Russian Federation)

**Founder V.P. Kupriyanov**

The journal is registered with the Federal Service  
for Supervision of Communications, Information Technology  
and Mass Communications (Roskommnadzor)  
March 3<sup>rd</sup>, 2020

Registration certificate III № ФС 77-77843

**ISSN 0236-235X** (print)

**ISSN 2311-2735** (online)

## EDITORIAL BOARD

**Semenov N.A.** – Deputy Editor-in-Chief, Dr.Sc. (Engineering), Professor of the Tver State Technical University (Tver, Russian Federation)

**Sotnikov A.N.** – Deputy Editor-in-Chief, Dr.Sc. (Physics and Mathematics), Professor, Deputy Director of the Joint Supercomputer Center of the Russian Academy of Sciences (Moscow, Russian Federation)

**Afanasiev A.P.** – Dr.Sc. (Physics and Mathematics), Professor of the Moscow Institute of Physics and Technology, Head of Centre for Distributed Computing of Institute for Information Transmission Problems (Moscow, Russian Federation)

**Balametov A.B.** – Dr.Sc. (Engineering), Professor of the Azerbaijan Scientific-Research & Design-Prospecting Power Engineering Institute (Baku, Azerbaijan)

**Borisov V.V.** – Dr.Sc. (Engineering), Professor of the MPEI Branch in Smolensk (Smolensk, Russian Federation)

**Golenkov V.V.** – Dr.Sc. (Engineering), Professor of the Belarusian State University of Informatics and Radioelectronics (Minsk, Republic of Belarus)

**Elizarov A.M.** – Dr.Sc. (Physics and Mathematics), Professor of the N.I. Lobachevsky Institute of Mathematics and Mechanics of the Kazan Federal University (Kazan, Russian Federation)

**Eremeev A.P.** – Dr.Sc. (Engineering), Professor of the National Research University "Moscow Power Engineering Institute" (Moscow, Russian Federation)

**Kuznetsov O.P.** – Dr.Sc. (Engineering), Professor of the Institute of Control Sciences of the Russian Academy of Sciences (Moscow, Russian Federation)

**Mamrosenko K.A.** – Ph.D. (Engineering), Associate Professor of the Moscow Aviation Institute (National Research University), Head of the Center of Visualization and Satellite Information Technologies SRISA RAS (Moscow, Russian Federation)

**Palyukh B.V.** – Dr.Sc. (Engineering), Professor of the Tver State Technical University (Tver, Russian Federation)

**Suleimanov D.Sh.** – Academician of TAS, Dr.Sc. (Engineering), Professor of the Kazan State Technical University (Kazan, Russian Federation)

**Tatarnikova T.M.** – Dr.Sc. (Engineering), Associate Professor, Professor of the St. Petersburg Electrotechnical University "LETI" (St. Petersburg, Russian Federation)

**Ulyanov S.V.** – Dr.Sc. (Physics and Mathematics), Professor of the Dubna International University for Nature, Society and Man (Dubna, Russian Federation)

**Khoroshevsky V.F.** – Dr.Sc. (Engineering), Professor of the Moscow Institute of Physics and Technology (Moscow, Russian Federation)

**Shabanov B.M.** – Dr.Sc. (Engineering), Corresponding Member of the RAS, Director of the Joint Supercomputer Center of the Russian Academy of Sciences (Moscow, Russian Federation)

**Yazenin A.V.** – Dr.Sc. (Physics and Mathematics), Professor of the Tver State University (Tver, Russian Federation)

## ASSOCIATED EDITORIAL BOARD MEMBERS

**National Research University "Moscow Power Engineering Institute", Moscow, Russian Federation**  
**Technology Institute at Southern Federal University, Taganrog, Russian Federation**  
**Tver State Technical University, Tver, Russian Federation**

### EDITORIAL BOARD AND PUBLISHER OFFICE ADDRESS

Nikolay Korytkov Ave, 3a, Tver, 170024, Russian Federation

**Phone:** (482-2) 39-91-49 **Fax:** (482-2) 39-91-00

**E-mail:** red@cps.tver.ru

**Website:** www.swsys.ru

Release date 16.06.2024

Printed in printing-office "Faktor i K"

Krylova St. 26, Tver, 170100, Russian Federation

Published quarterly. 37th year of publication

Format 60×84 1/8. Circulation 1000 copies

Prod. order № 5. Wordage 164 pages. Price 550,00 rub.

## Вниманию авторов

Журнал «Программные продукты и системы» публикует материалы научного и научно-практического характера по новым информационным технологиям, результаты академических и отраслевых исследований в области использования средств вычислительной техники. Практикуются выпуски тематических номеров по искусственному интеллекту, системам автоматизированного проектирования, по технологиям разработки программных средств и системам защиты, а также специализированные выпуски, посвященные научным исследованиям и разработкам отдельных вузов, НИИ, научных организаций.

Журнал «Программные продукты и системы» внесен в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней кандидата и доктора наук.

Информация об опубликованных статьях по установленной форме регулярно предоставляется в систему РИНЦ, в CrossRef и в другие базы и электронные библиотеки.

Журнал «Программные продукты и системы» включен в ядро коллекции РИНЦ, размещенное на платформе Web of Science в виде базы данных RSCI.

Автор статьи отвечает за подбор, оригинальность и точность приводимого фактического материала. При перепечатке ссылка на журнал обязательна. Статьи публикуются бесплатно.

### Условия публикации

К рассмотрению принимаются оригинальные материалы, отвечающие редакционным требованиям и соответствующие тематике журнала. Группы научных специальностей:

#### **1.2. Компьютерные науки и информатика**

1.2.1. Искусственный интеллект и машинное обучение (физико-математические науки).

1.2.2. Математическое моделирование, численные методы и комплексы программ (физико-математические науки, технические науки).

#### **2.3. Информационные технологии и телекоммуникации**

2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки, физико-математические науки).

2.3.2. Вычислительные системы и их элементы (технические науки).

2.3.3. Автоматизация и управление технологическими процессами и производствами (технические науки).

2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (технические науки, физико-математические науки).

2.3.6. Методы и системы защиты информации (технические науки, физико-математические науки).

2.3.7. Компьютерное моделирование и автоматизация (технические науки, физико-математические науки).

2.3.8. Информатика и информационные процессы (технические науки).

Работа представляется в электронном виде в формате Word. Объем статьи вместе с иллюстрациями – не менее 10 000 знаков. Диаграммы, схемы, графики должны быть доступными для редактирования (Word, Visio, Excel). Заголовок должен быть информативным; сокращения, а также терминологию узкой тематики желательно в нем не использовать. Количество авторов на одну статью – не более четырех, количество статей одного автора в номере, включая соавторство, – не более двух. Список литературы, наличие которого обязательно, должен включать не менее 10 пунктов.

Необходимы также содержательная структурированная аннотация (не менее 200 слов), ключевые слова (7–10) и индекс УДК. Название статьи, аннотация и ключевые слова должны быть переведены на английский язык (машинный перевод недопустим), а фамилии авторов, названия и юридические адреса организаций (если нет официального перевода) – транслитерированы по стандарту BGN/PCGN.

Вместе со статьей следует прислать экспертное заключение о возможности открытого опубликования материала и авторскую справку. Обязательно соблюдение автором договора (публичной оферты).

### Порядок рецензирования

Все статьи, поступающие в редакцию (соответствующие тематике и оформленные согласно требованиям к публикации), подлежат двойному слепому рецензированию в течение месяца с момента поступления, рецензия отправляется авторам.

В редакции сформирован устоявшийся коллектив рецензентов, среди которых члены редколлегии журнала, эксперты из числа крупных специалистов в области информатики и вычислительной техники ведущих вузов страны, а также ученые и специалисты НИИСИ РАН, МСЦ РАН (г. Москва) и НИИ «Центрпрограмм-систем» (г. Тверь).

Редакция журнала «Программные продукты и системы» в своей работе руководствуется сводом правил Кодекса этики научных публикаций, разработанным и утвержденным Комитетом по этике научных публикаций (Committee on Publication Ethics – COPE).

## Вычисления обратных функций распределений: алгоритмы и программы

Л.В. Агамиров<sup>1,2</sup>, В.Л. Агамиров<sup>1,2</sup>✉, В.А. Вестяк<sup>1</sup>

<sup>1</sup> Московский авиационный институт (Национальный исследовательский университет), г. Москва, 125993, Россия

<sup>2</sup> Московский технический университет связи и информатики (МТУСИ), г. Москва, 111024, Россия

### Ссылка для цитирования

Агамиров Л.В., Агамиров В.Л., Вестяк В.А. Вычисления обратных функций распределений: алгоритмы и программы // Программные продукты и системы. 2024. Т. 37. № 2. С. 137–145. doi: 10.15827/0236-235X.142.137-145

### Информация о статье

Группа специальностей ВАК: 1.2.2

Поступила в редакцию: 06.10.2023

После доработки: 13.11.2023

Принята к публикации: 21.11.2023

**Аннотация.** Точный расчет ряда функций распределения случайных переменных в прикладных задачах математической статистики вызывает существенные вычислительные трудности, обусловленные наличием бесконечных пределов интегрирования, необходимостью минимизации целевых функций и отсутствием удовлетворительных аппроксимаций. Для решения поставленных задач авторами данной статьи получены точные аналитические соотношения, позволяющие производить численное интегрирование функций распределения коэффициента вариации и нецентрального распределения Стьюдента, сводящееся к вычислению однократных интегралов. Обратные функции распределения определяются минимизацией на основе симплекс-метода Нелдера–Мида. Аналогично решается задача точного вычисления числовых характеристик порядковых статистик. Разработаны алгоритмы и программы на JavaScript с открытым кодом для реализации указанных вычислительных задач. Расчеты иллюстрируются графиками и таблицами, в которых представлены результаты вычисления квантилей нецентрального  $t$ -распределения Стьюдента в диапазоне объемов выборки от 3 до 50, вероятностей от 0,01 до 0,99 и доверительных вероятностей 0,9, 0,95 и 0,99. Время расчета в полном диапазоне всех параметров составляет не более 10–15 секунд на компьютере средней производительности, точность расчета – порядка  $10^{-5}$ . Отмечается, что основные временные затраты составляет численное интегрирование в связи с наличием бесконечных пределов интегрирования, в то время как минимизация осуществляется довольно быстро (не более 20–30 итераций). В статье также представлены результаты вычисления квантилей относительных коэффициентов вариации для объемов выборки 3–10, генеральных коэффициентов вариации 0,05, 0,3, 0,5 и вероятностей в диапазоне от 0,01 до 0,99. Выполнены сравнительные расчеты числовых характеристик нормальных и weibullовских порядковых статистик, полученных прямым интегрированием и соответствующими аппроксимациями. В рассматриваемых программах применяются лишь простейшие и достаточно точные аппроксимации стандартных распределений: нормального распределения, гамма-функции, неполной гамма-функции. Разработанные алгоритмы пригодны для широкого класса непрерывных распределений, обратные функции которых не имеют приемлемых аппроксимаций.

**Ключевые слова:** алгоритмы и программы, JavaScript, нецентральное распределение Стьюдента, точное распределение коэффициента вариации, распределение порядковых статистик, нормальное распределение, распределение Вейбулла

**Введение.** В прикладных задачах математической статистики существует ряд случайных переменных, точный расчет функций распределения которых вызывает значительные вычислительные трудности из-за наличия бесконечных пределов интегрирования, необходимости минимизации целевых функций, отсутствия удовлетворительных аппроксимаций и т.п. К таким распределениям относятся нецентральное распределение Стьюдента, распределения коэффициента вариации, порядковых статистик и другие. Кроме задач вычисления мощности статистических критериев, значения квантилей нецентрального распределения Стьюдента необходимы в задачах надежности для обоснования толерантных интервалов и гарантированных ресурсных характеристик техни-

ческих систем [1]. Область применения коэффициента вариации в инженерных задачах рассматривается в работах [1, 2]. Доказательство точного распределения коэффициента вариации для нормальной выборки содержится в классической книге Лемана [3], все последующие работы – это различной точности аппроксимации этого распределения. Математические ожидания и ковариации порядковых статистик широко применяются в прикладных задачах математической статистики.

Анализ литературных источников по теме данного исследования, связанного с разработкой алгоритмов и программ для точного вычисления функций статистических распределений, показал, что их количество весьма ограничено. Так, например, в [4, 5] рассматривается

наиболее распространенное применение обратных функций – моделирование случайных величин методом обратных функций, а также разработка аналитических приближений для обращения функций вероятностных распределений [6]. Многие исследования посвящены новым методам представления аналитических аппроксимаций для прямых и обратных функций сложных нецентральных статистических распределений [7–9]. В то же время непосредственно связанные с темой данной работы источники практически отсутствуют. По мнению авторов, причина в том, что в настоящее время специалисты в области прикладной статистики используют в основном аппроксимации, созданные в математических пакетах с закрытым кодом типа Boost [10], Matlab, Statistica, Mathcad и в других. В свою очередь, как показал анализ документации по этим пакетам, они в значительной степени основаны на алгоритмах и программах прикладной статистики в рамках проекта Royal Statistical Society “Applied Statistics algorithms” (<http://lib.stat.cmu.edu/apstat/>), содержащего около 250 алгоритмов начиная с 1968 г. (проект завершился в 1997 г.), переведенных с языка Algol на Fortran, а затем на C++ ([https://people.sc.fsu.edu/~jburkardt/cpp\\_src/cpp\\_src.html](https://people.sc.fsu.edu/~jburkardt/cpp_src/cpp_src.html)). Точные вычисления содержатся в уникальных статистических таблицах [11] и их русскоязычных аналогах (особенно в части нецентральных распределений) [12], которые, возможно, выполнены методами численного интегрирования в специализированных математических институтах, но недоступны для анализа кода и использования в задачах компьютерного моделирования. Кроме того, приведенные в табличном виде процентные точки обладают дискретностью, что вызывает необходимость интерполяции или экстраполяции, а это, в свою очередь, снижает точность расчетов. Очевидно, что при современном уровне развития информационных технологий использование таблиц является анахронизмом и может служить лишь для контроля точности численных расчетов, а также в учебных целях. Следует также упомянуть статистические программы, встроенные в популярный язык Python, которые полностью базируются на динамических библиотеках Boost с закрытым кодом. В отличие от Python эти библиотеки в стандарт C++ пока не встроены и требуют отдельной, часто трудоемкой установки. Тем не менее анализ некоторых открытых кодов статистических функций Boost показывает, что они основаны на многоуровневых аппроксима-

циях, выполненных многими исследователями в разное время с множеством условий. Преимуществами аппроксимаций Boost являются их высокая точность и практически мгновенное действие. Аппроксимации, встроенные в специализированный язык R (The R Project for Statistical Computing, <https://www.r-project.org/>), уступают по точности функциям Boost. Все сказанное касается, разумеется, сложных распределений, стандартные распределения в данной работе не рассматриваются.

Таким образом, исследования, направленные на разработку ПО для реализации точных алгоритмов вычисления сложных функций статистических распределений, являются актуальной задачей. При этом точность решений определяется точностью не аппроксимаций, а численного интегрирования и методов минимизации, которые можно неограниченно увеличивать с учетом лишь временных затрат.

Следует также отметить, что приемлемых по точности расчетов параметров порядковых статистик авторы не обнаружили, поэтому преимущество предлагаемых методов расчета прямым интегрированием сложных функций с последующей минимизацией целевых функций является не только точность вычислений (зачастую аппроксимации дают не меньшую точность), но и сам подход. Он позволяет распространять данную методику практически на любую схожую задачу с небольшими ограничениями типа непрерывности и дифференцируемости исходных функций, в то время как аппроксимацию следует полностью модифицировать для каждой конкретной задачи с тщательным последующим тестированием. Исключения, разумеется, встречаются, например, при численном интегрировании двойных интегралов с бесконечными пределами. В этих случаях альтернативы аппроксимациям нет из-за неоправданно больших временных затрат.

Целью настоящего исследования является разработка алгоритмов и программ расчета обратных функций сложных статистических распределений. Описываемые алгоритмы пригодны для широкого класса непрерывных распределений. В рассматриваемых программах применяются лишь простейшие и достаточно точные аппроксимации стандартных распределений. В настоящей работе это аппроксимации нормального распределения, гамма-функции, неполной гамма-функции, точность которых оказалась выше алгоритмов на Фортране (библиотека SSP) [13]. Авторы предлагают алгоритмы и программы на языке Javascript [14], ко-

торые выложены по адресу <http://g2.plzvpn.ru:3000/Rank/rank.html> (исходный программный код реализации по ссылке [https://github.com/AVL095/Inverse\\_distribution\\_functions](https://github.com/AVL095/Inverse_distribution_functions)). Выбор языка обусловлен его общедоступностью и быстродействием.

Вышеуказанные задачи основаны на распределении отношения независимых случайных величин  $x = \zeta_1/\zeta_2$ :

$$F(x) = \int_0^\infty F_1(tx)f_2(t)dt \quad (1)$$

или

$$F(x) = 1 - \int_0^\infty f_1(t)F_2\left(\frac{t}{x}\right)dt, \quad (2)$$

где  $f_1(t)$ ,  $F_1(t)$  – функция плотности и функция распределения случайной величины  $\zeta_1$ ;  $f_2(t)$ ,  $F_2(t)$  – функция плотности и функция распределения случайной величины  $\zeta_2$ .

В соответствии с этим в выборке объема  $n$  из нормального распределения  $N(a, \sigma)$  с выборочными средним  $\hat{a} = \bar{x} = \sum_{i=1}^n x_i / n$  и дисперсией

$\hat{\sigma}^2 = s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)$  случайная величина  $t' = (\bar{x} - a + \delta)\sqrt{n} / s$  имеет нецентральное распределение Стьюдента с функцией распределения [3]

$$F(t') = \beta = \int_0^\infty f(x)\Phi(t'x, \delta, 1)dx, \quad (3)$$

откуда определяется квантиль распределения уровня  $\beta$ . Здесь

$$\Phi(x, a, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left[-\frac{(t-a)^2}{2\sigma^2}\right] dt \quad (4)$$

функция распределения нормального закона;  $\delta$  – параметр нецентральности;  $f = n - 1$  – число степеней свободы. Распределение выборочного среднего подчиняется нормальному закону с параметрами  $N(a, \delta/n^{0.5})$ . Переменная  $y = s^2(n - 1)/\delta^2$  имеет  $\chi^2$  распределение с  $f$  степенями свободы и плотностью

$$\varphi(y) = \frac{2^{-f/2}}{\Gamma(f/2)} y^{f/2-1} \exp(-y/2),$$

где  $\Gamma(x)$  – гамма-функция (см. модуль *stat.js*). Плотность выборочного распределения стандартного отклонения  $x = (s^2/f)^{0.5}$  в формуле (3) определяется на основании теоремы о плотности монотонной функции случайной величины:

$$f(x) = \varphi(fx^2)(fx^2)' = \frac{2(f/2)^{f/2}}{\Gamma(f/2)} x^{f-1} \exp(-fx^2/2). \quad (5)$$

Таким образом, распределение коэффициента вариации  $\gamma = \sigma/a$  подчиняется нецентраль-

ному распределению Стьюдента с параметром нецентральности  $\delta = (n)^{0.5}/\gamma$  и числом степеней свободы  $f$  [3], то есть  $v_\beta = \delta/t'_\beta = \sqrt{n}/(t'_\beta \gamma)$ .

Поэтому для вычисления квантилей коэффициента вариации достаточно иметь квантили нецентрального распределения Стьюдента. Прямое вычисление функции распределения выборочного коэффициента вариации сводится к уравнению

$$F(v) = \beta = 1 - \int_0^\infty f(x)\Phi\left(x\frac{v}{\gamma}, 1, \gamma/\sqrt{n}\right)dx, \quad (6)$$

откуда определяется квантиль распределения коэффициента вариации уровня  $\beta$ ,  $\gamma$  – генеральное значение коэффициента вариации.

В настоящей работе анализировались оба подхода, по точности и быстродействию они показали идентичные результаты.

Для вычисления функций нецентрального распределения Стьюдента и коэффициента вариации необходимо численное интегрирование уравнений (3) или (6). Квантили распределения, соответствующие заданной вероятности, определялись последовательными приближениями. С этой целью в настоящей работе использовался симплекс-метод Нелдера–Мида (деформируемого многогранника).

Алгоритм расчета обратной функции нецентрального распределения Стьюдента реализован путем интегрирования уравнения (3) с последующей минимизацией квадратичной функции  $q = (\beta - \beta_0)^2$ , представляющей собой квадрат разности заданной  $\beta_0$  и расчетной  $\beta$  доверительных вероятностей. Параметры функции *simpl* (модуль *stat.js*):  $x[nx]$  – вектор размерности  $nx$ , на входе содержащий начальные приближения, на выходе – точки минимума;  $nx$  – число переменных минимизируемой функции (в рассматриваемом случае  $nx = 1$ ); *step* – начальный шаг минимизации; *eps* – относительная точность выхода; *lim* – максимальное число итераций, функция *simpl* возвращает число выполненных итераций *iter*; *funx* – имя минимизируемой функции. Вызов функции: *iter = simpl(x, nx, stepx, eps, lim, funx)*.

Численное интегрирование уравнения (3) осуществлялось методом Буля (модуль *stat.js*). Параметры функции *IntegrateFunction*:  $k$  – номер выбранного метода, соответствующий размерности аппроксимирующего полинома (3, 4 – метод Симпсона, 5 – метод Буля); *nstep* – количество шагов интегрирования; *xl, xu* – нижний и верхний пределы интегрирования; *fpolinom* – имя интегрируемой функции, функция возвращает значение интеграла. Вызов функции: *beta =*

= *IntegrateFunction(k, nstep, xl, xu, fpolinom)*. Вспомогательные программы содержатся в модуле *stat.js*. С целью исключения возможных ошибок в программах используются лишь простейшие и проверенные конструкции языка JavaScript, включая ввод данных и вывод результатов на экран.

Аналогичный алгоритм реализован и для распределения коэффициента вариации, он отличается только типом интегрируемой функции (в данном случае (6)).

Как показали расчеты, быстрый и успешный поиск точек минимума в программе *simpl* существенно зависит от заданных на входе начальных приближений. С этой целью использовалась следующая нормальная аппроксимация квантиля [1] (функция *invnontap* в модуле *stat.js*):

$$t'(\beta, \delta, f) = \frac{\left(1 - \frac{1}{4f}\right)\delta + z_{\beta}\sqrt{\left(1 - \frac{1}{4f}\right)^2 - \frac{z_{\beta}^2}{2f} + \frac{\delta^2}{2f}}}{\left(1 - \frac{1}{4f}\right)^2 - \frac{z_{\beta}^2}{2f}}, \quad (7)$$

где  $\delta$  – параметр нецентральности;  $f$  – число степеней свободы;  $z_{\beta}$  – квантиль нормированного нормального распределения уровня  $\beta$ ;  $\beta$  – доверительная вероятность.

Применительно к распределению коэффициента вариации в качестве первого приближения использовалась аппроксимация

$$\frac{v_p}{\gamma} = \sqrt{\frac{\chi_{p,f}^2}{(1+1/\gamma^2)f - \chi_{p,f}^2/(f+1)}}, \quad (8)$$

где  $v_p$  – квантиль распределения коэффициента вариации уровня вероятности  $p$ ;  $\gamma$  – генеральное значение коэффициента вариации;  $\chi_{p,f}^2$  – квантиль распределения хи-квадрат [1].

Аппроксимация (8) дает существенные погрешности при значениях  $\gamma$ , больших 0,3.

Менее сложной с точки зрения программирования представляется задача точного вычисления числовых характеристик (математических ожиданий  $E(x_r)$  и дисперсий  $D(x_r)$ ) порядковых статистик  $x_r$  [2, 4, 5] в выборке объема  $n$  прямым интегрированием:

$$E(x_r) = \frac{\int_{-\infty}^{\infty} \varphi(x)x[1-F(x)]^{n-r}[F(x)]^{r-1} dx}{B(r, n-r+1)}, \quad (9)$$

$$D(x_r) = \frac{\int_{-\infty}^{\infty} \varphi(x)x^2[1-F(x)]^{n-r}[F(x)]^{r-1} dx}{B(r, n-r+1)} -$$

$$-E^2(x_r), \quad (10)$$

где  $r = 1 - n$  – номер порядковой статистики;  $\varphi(x)$  – плотность распределения;  $F(x)$  – функция распределения;  $B(a, b)$  – бета-функция. Численное интегрирование осуществляется с помощью функции *dataorder* в модуле *cvar.js*.

В таблице 1 показаны результаты вычисления квантилей нецентрального распределения Стьюдента. В программе варьировались объем испытаний  $n$  от 3 до 50, вероятности  $p$  от 0,01 до 0,99 и доверительные вероятности 0,9, 0,95 и 0,99. В таблице представлен фрагмент расчетов для  $n$ , равных от 3 до 10. Время расчета в полном диапазоне всех параметров составляет не более 10–15 секунд на компьютере невысокой производительности (индекс производительности – 1,0, ОЗУ – 4Гб, процессор – Intel® Pentium® 1,9 GHz). Точность расчета – порядка  $10^{-5}$ . Сопоставления по быстродействию с аналогичными расчетами, выполняемыми приближенными методами, например, в библиотеках Boost, не анализируются, так как вычисления аппроксимаций составляют доли секунды.

Необходимо отметить, что изменение требуемой точности расчетов (*eps*), то есть относительной точности выхода в программе минимизации, практически не влияет на машинное время, поскольку при надлежащем задании начальных приближений программа очень быстро выходит на минимум. В результате вариация этого параметра в широком диапазоне значений приводит к одним и тем же точечным оценкам с незначительным увеличением числа итераций. Несколько большее значение имеют задаваемые предельное число итераций (*lim*) и шаг минимизации (*step*), но только тогда, когда программа вообще не находит минимума, что считается неудовлетворительным и подлежит исключению. В этом случае необходимо уточнить начальные приближения. Применительно к задачам, рассматриваемым в данной работе, такого не наблюдалось. Следует также отметить, что скорость вычислений на языке C++ в десятки раз выше, чем на Javascript и Python.

Основные временные затраты приходятся именно на численное интегрирование, а программа минимизации работает достаточно быстро (не более 20–30 итераций), что вполне объяснимо, так как в связи с наличием бесконечных пределов интегрирования для достижения требуемой точности необходимо охватить максимальный диапазон переменной независимо от применяемого метода. В рассматриваемых программах верхний предел интегрирования задавался равным 5, но может изменяться пользователем в зависимости от требуемой точности расчетов.

Таблица 1

**Квантили нецентрального распределения Стьюдента для объемов выборки  $n = 3-10$ , доверительной вероятности  $\beta$  и вероятностей  $p$  (параметр нецентральности  $\delta = z_p n^{0.5}$ )**

Table 1

**Quantiles of non-central Student distribution for sample sizes  $n = 3-10$ , confidence probability  $\beta$ , probabilities  $p$  (noncentrality parameter  $\delta = z_p n^{0.5}$ )**

$\beta = 0,99$									
$n$	$P = 0,01$	$0,05$	$0,1$	$0,3$	$0,5$	$0,7$	$0,9$	$0,95$	$0,99$
3	-1,3536792	-0,5105654	0,1245945	2,8786357	6,9645544	13,0712125	24,2407483	30,0860653	41,3883239
4	-1,8472004	-0,8857286	-0,2451913	1,8985941	4,5407048	8,223549	14,7597818	18,1669008	24,7745658
5	-2,296835	-1,214473	-0,5315038	1,4986883	3,7469478	6,7518861	11,9891732	14,7096064	19,9882608
6	-2,7131195	-1,5144931	-0,7810481	1,2576921	3,3649308	6,1001286	10,8048983	13,2408281	17,9659398
7	-3,1028759	-1,7934666	-1,0080186	1,0830341	3,1426678	5,7580231	10,210297	12,5087421	16,9644092
8	-3,4708345	-2,0558281	-1,2188556	0,9431022	2,9979522	5,5626035	9,8915944	12,1205258	16,43825
9	-3,8204406	-2,3045162	-1,4171848	0,8241254	2,8964601	5,4469441	9,7212346	11,9167803	16,166635
10	-4,1543004	-2,5416339	-1,6053317	0,7190867	2,8214376	5,3788426	9,6383305	11,8215914	16,0445284
$\beta = 0,95$									
3	-1,9566387	-1,1070312	-0,5793325	1,0051572	2,919986	5,6762818	10,6612588	13,2604081	18,2778621
4	-2,4923255	-1,4866088	-0,8877785	0,7109377	2,3533635	4,5308156	8,3238654	10,2877502	14,0847225
5	-2,9760095	-1,8286089	-1,1600269	0,5300989	2,1318469	4,1607664	7,6174653	9,3974797	12,8374608
6	-3,4205601	-2,1427448	-1,4080225	0,3909637	2,0150488	4,0125171	7,3637937	9,0819335	12,3992868
7	-3,8343448	-2,4350796	-1,6378358	0,2733724	1,9431808	3,9544098	7,2901881	8,9941472	12,2808368
8	-4,2231165	-2,7097237	-1,8532134	0,169149	1,8945787	3,9403276	7,3027406	9,0150293	12,3145682
9	-4,5910212	-2,9696243	-2,0567122	0,0742025	1,8595475	3,9502649	7,3612672	9,093709	12,4290667
10	-4,9411662	-3,2169811	-2,2501852	-0,0138155	1,8331132	3,974285	7,4460279	9,205277	12,5893981
$\beta = 0,90$									
3	-2,3570178	-1,4550122	-0,9262696	0,4518083	1,8856184	3,8590338	7,3753589	9,1997526	12,7140114
4	-2,9103929	-1,844764	-1,2341415	0,2596394	1,6377444	3,3860116	6,3756891	7,9131314	10,8764665
5	-3,4089776	-2,196217	-1,5099056	0,1169958	1,5332068	3,2551829	6,1320779	7,6022592	10,4334531
6	-3,8660398	-2,5186127	-1,7621781	-0,0031124	1,4758845	3,2272904	6,1082676	7,5735243	10,3920429
7	-4,2904886	-2,8181408	-1,996231	-0,1097249	1,4397552	3,2413693	6,1716037	7,6562707	10,5089838
8	-4,6884925	-3,0990964	-2,215594	-0,2070261	1,414924	3,2754144	6,2751304	7,7902943	10,6986604
9	-5,0645002	-3,3645879	-2,4227766	-0,2973447	1,3968153	3,3197823	6,3986238	7,9497059	10,9243219
10	-5,4218393	-3,6169423	-2,6196388	-0,3821411	1,3830284	3,369699	6,5322147	8,1219114	11,1680831

В таблице 2 представлены результаты вычислений квантилей относительных коэффициентов вариации  $v_p/\gamma$  для объемов выборки  $n = 3-10$ , генеральных коэффициентов вариации  $\gamma = 0,05, 0,3$  и  $0,5$  и вероятностей  $p$  в диапазоне от  $0,01$  до  $0,99$ . Быстродействие и точность примерно эквивалентны приведенным в предыдущем примере.

На рисунке отображены графики функций распределения коэффициента вариации при  $\gamma = 0,3$  для  $n = 5, 10$  и  $20$ , построенные в соответствии с приведенной методикой.

Сравнение результатов расчета математических ожиданий и дисперсий нормальных порядковых статистик, полученных аппроксима-

цией Дэйвида–Джонсона (функция *ordern*, модуль *stat.js*), с порядком разложения в ряд  $(n + 2)^3$  с табличными значениями показывают расхождение лишь в 4-м, 5-м знаке после запятой. В функции *ordern* содержатся производные функции распределения вплоть до 6-й, а также комментарии. Быстродействие настолько высокое, что в данной работе не анализируется.

Предлагаемое авторами прямое интегрирование функций распределения позволяет вычислять также числовые характеристики порядковых статистик для законов распределения, отличных от нормального. Функция распределения Вейбулла, например, с параметрами  $b$  и  $c$  имеет следующий вид:

Таблица 2

**Квантили выборочных коэффициентов вариации  $v_p/\gamma$  для объемов выборки  $n = 3-10$ , коэффициентов вариации  $\gamma$  и вероятностей  $p$**

Table 2

**Quantiles of sample variation coefficients  $v_p/\gamma$  for sample sizes  $n = 3-10$ , variation coefficients  $\gamma$  and probabilities  $p$**

$\gamma = 0,05$									
$n$	$P = 0,01$	<b>0,05</b>	<b>0,1</b>	<b>0,3</b>	<b>0,5</b>	<b>0,7</b>	<b>0,9</b>	<b>0,95</b>	<b>0,99</b>
3	0,1002102	0,2263956	0,3244862	0,5971515	0,8326885	1,0979006	1,51971	1,7344276	2,1533418
4	0,1955307	0,3422886	0,4411568	0,6887522	0,8881656	1,1058458	1,4454562	1,616912	1,9503398
5	0,2723549	0,421281	0,5154112	0,7405768	0,9161456	1,1048769	1,3962501	1,5426097	1,8265554
6	0,3327173	0,4783546	0,5672227	0,7744235	0,9329624	1,1017842	1,3606271	1,4901799	1,7410467
7	0,3809654	0,521763	0,6057963	0,7985351	0,9441737	1,0982481	1,3332994	1,4506201	1,6774501
8	0,4204085	0,5561038	0,6358691	0,816735	0,9521778	1,0947712	1,3114685	1,4193896	1,6277746
9	0,4533285	0,5841021	0,6601271	0,8310509	0,9581766	1,0915088	1,2935028	1,3939159	1,587587
10	0,4812971	0,6074751	0,6802103	0,8426637	0,9628401	1,0884974	1,278376	1,3726192	1,5542125
$\gamma = 0,3$									
3	0,0988091	0,2234869	0,3208023	0,5945781	0,8371579	1,1205464	1,6048203	1,8737959	2,4638789
4	0,1916642	0,3363389	0,4345316	0,6843881	0,8914472	1,1257984	1,5156774	1,7275832	2,1801959
5	0,266242	0,413217	0,507012	0,7353899	0,9187356	1,122733	1,4564944	1,6353082	2,0107654
6	0,3248317	0,468889	0,5577623	0,7688484	0,935102	1,1180122	1,4137562	1,5705593	1,8958781
7	0,3717281	0,5113646	0,5956947	0,7927851	0,9459959	1,1131818	1,3810865	1,5220013	1,8117507
8	0,4101458	0,5450793	0,6253817	0,8109232	0,9537644	1,1086514	1,3550862	1,4838882	1,7468857
9	0,442285	0,5726575	0,649414	0,8252398	0,959582	1,1045128	1,3337646	1,4529646	1,6949849
10	0,4696559	0,5957507	0,6693761	0,8368895	0,9641011	1,1007598	1,3158731	1,427234	1,6522852
$\gamma = 0,5$									
3	0,0963902	0,2184256	0,3143173	0,5895348	0,844215	1,1621726	1,7900129	2,2137372	3,5256371
4	0,1853166	0,3264466	0,4233633	0,6764026	0,8962664	1,1613484	1,6601829	1,9763649	2,8335787
5	0,2564419	0,4000839	0,4931283	0,7261626	0,9223899	1,1540322	1,5763381	1,8337191	2,4843834
6	0,312353	0,4536427	0,5422898	0,7590871	0,9380435	1,1461675	1,5170257	1,7369937	2,2692031
7	0,3572225	0,4947248	0,5792782	0,7828211	0,9484576	1,138908	1,4724104	1,6662827	2,1213024
8	0,3941057	0,527509	0,6084066	0,800924	0,9558808	1,132438	1,4373606	1,6118752	2,0123565
9	0,4250757	0,5544631	0,6321201	0,8152957	0,9614377	1,1267084	1,4089261	1,5684337	1,9281692
10	0,4515489	0,5771415	0,6519189	0,8270491	0,9657532	1,1216241	1,3852797	1,5327653	1,8607923

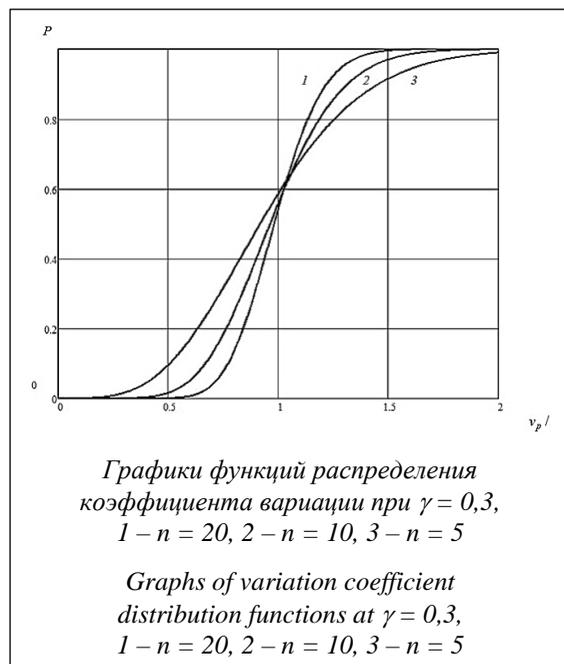
$$F(x) = 1 - \exp\left(-\frac{x}{c}\right)^b. \tag{11}$$

Приводя распределение к виду с параметрами сдвига  $a_w = lnc$  и масштаба  $\sigma_w = 1/b$ ,  $z = (\ln x - a_w)/\sigma_w$ , получим соотношения для функции и плотности распределения:

$$F(z) = 1 - e^{-e^z}; \varphi(z) = \frac{\partial F(z)}{\partial z} = ze^{z-e^z}, \tag{12}$$

которые подставляются в (9) и (10). Сравнительные расчеты числовых характеристик по-

рядковых статистик распределения Вейбулла по интегралам (9), (10) (функция *dataorder* в модуле *cvar.js*) и приближенных, основанных на аппроксимации Дэйвида–Джонсона (функция *orderw*, модуль *stat.js*), показали расхождение в 3-м, 4-м знаке после запятой. Пользователям рекомендуется самостоятельно принимать решение о применении той или иной модели в зависимости от требуемой точности расчетов. Предлагаемая аппроксимация позволяет также вычислять ковариации порядковых статистик.



Точное решение для ковариаций здесь не рассматривается, так как требует вычисления двойных интегралов. Поскольку распределение Вейбулла является несимметричным, следует вычислять все порядковые статистики от 1 до  $n$  в отличие от нормального закона.

### Заключение

Разработанные алгоритмы и программы вычисления обратных функций нецентрального распределения Стьюдента и распределения коэффициента вариации основаны на сочетании численного интегрирования и симплекс-метода минимизации, позволяющих получать быстрые

и точные решения для применения данных функций в задачах прикладной статистики.

Приведенные в таблицах сравнительные расчеты показали высокую точность вычислений и особенности существующих аппроксимаций. Так, например, алгоритмы аппроксимации функции нецентрального распределения Стьюдента дают весьма точные приближения во всем диапазоне вероятностей, объемов выборки и параметров нецентральности и вполне пригодны для последующей минимизации с целью получения процентных точек. В то же время известная аппроксимация МсКау для коэффициента вариации дает погрешности до 15 % при значениях коэффициента вариации, больших 0,3.

Представленные в полном объеме и в открытом доступе модифицированные программы и аппроксимации на языке Javascript позволяют без труда распространять полученные решения для других статистических задач, связанных с точным распределением отношений двух независимых случайных переменных.

Разработанные программы интегрирования функций с бесконечными пределами для получения числовых характеристик порядковых статистик позволяют производить необходимые вычисления для большого класса непрерывных функций распределения. В работе рассмотрены нормированные нормальное и Вейбулла распределения, а также представлены аппроксимации методом Дэйвида-Джонсона. Расчеты показали достаточно высокую точность разложения по сравнению с интегральными решениями (расхождения наблюдаются в 3-м, 4-м знаке после запятой).

### Список литературы

1. Агамиров Л.В., Вестяк В.А. Вероятностные методы расчета показателей надежности авиационных конструкций при переменных нагрузках. М.: МАИ, 2022. 256 с.
2. Tothfalusi L., Endrenyi L. An exact procedure for the evaluation of reference-scaled average bioequivalence. AAPS J., 2016, vol. 18, no. 2, pp. 476–489. doi: 10.1208/s12248-016-9873-6.
3. Lehmann E.L. Testing Statistical Hypothesis. NY, John Wiley & Sons Publ., 1986, 388 p.
4. Холкина А.В. Моделирование случайных величин методом обратных функций // Математическое и программное обеспечение информационных, технических и экономических систем: матер. VII Междунар. молодежн. научн. конф. 2019. С. 73–75.
5. Царев А.Д. Программные генераторы случайных блужданий // Вестн. КРАУНЦ. Физ.-мат. науки. 2020. Т. 31. № 2. С. 226–235. doi: 10.26117/2079-6641-2020-31-2-226-235.
6. Попов Г.А. Формула обращения для рациональных характеристических функций вероятностных распределений // Вестн. АГТУ. 2018. Т. 2018. № 2. doi: 10.24143/1812-9498-2018-2-7-22.
7. Gil A., Segura J., Temme N.M. New asymptotic representations of the noncentral t-distribution. Studies in Applied Math., 2023, vol. 151, no. 3, pp. 857–882. doi: 10.1111/sapm.12609.
8. Gil A., Segura J., Temme N.M. On the computation and inversion of the cumulative noncentral beta distribution. Appl. Math. Comput., 2019, vol. 361, pp. 74–86.
9. Gil A., Segura J., Temme N.M. A new asymptotic representation and inversion method for the Student's t distribution. Integral Transforms and Special Functions, 2022, vol. 33, no. 8, pp. 597–608. doi: 10.1080/10652469.2021.2007906.

10. Полухин А. Разработка приложений на С++ с использованием Boost. М.: ДМК Пресс, 2020. 346 с.
11. Pearson E.S., Hartley H.O. Biometrika Tables for Statisticians. Cambridge University Press, 1976, 286 p.
12. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983. 416 с.
13. 1130 Sci. Subroutine Package. Programmer's Manual, Program Number 1130-CM-02X. IBM Corporation, 1967, 191 p.
14. Агамиров Л.В., Вестяк В.А. Программа вычисления обратных функций сложных статистических распределений: Свид. о регистр. ПрЭВМ № 2022612358. Рос. Федерация, 2022.

Software &amp; Systems

doi: 10.15827/0236-235X.142.137-145

2024, 37(2), pp. 137–145

### Calculating inverse distribution functions: Algorithms and programs

Levon V. Agamirov <sup>1,2</sup>, Vladimir L. Agamirov <sup>1,2</sup>✉, Vladimir A. Vestyak <sup>1</sup><sup>1</sup> Moscow Aviation Institute (National Research University),  
Moscow, 125993, Russian Federation<sup>2</sup> Moscow Technical University of Communications and Informatics,  
Moscow, 111024, Russian Federation

#### For citation

Agamirov, L.V., Agamirov, V.L., Vestyak, V.A. (2024) 'Calculating inverse distribution functions: Algorithms and programs', *Software & Systems*, 37(2), pp. 137–145 (in Russ.). doi: 10.15827/0236-235X.142.137-145

#### Article info

Received: 06.10.2023

After revision: 13.11.2023

Accepted: 21.11.2023

**Abstract.** In applied problems of mathematical statistics, the exact calculation of some distribution functions of random variables causes significant computational difficulties due to infinite integration limits, the need to minimize objective functions, and the lack of satisfactory approximations. To solve these problems, the authors of the paper have obtained exact analytical relations that allow numerical integration of the distribution functions of a variation coefficient and the non-central Student distribution, which are reduced to calculating single integrals. Inverse distribution functions are determined by minimization based on the Nelder–Mead simplex method. The problem of accurately calculating the numerical characteristics of order statistics is solved in a similar way. The paper describes the developed algorithms and open source JavaScript programs to implement these computational tasks. The calculations are illustrated by graphs and tables that present the results of calculating non-central Student t-distribution quantiles in a range of sample sizes from 3 to 50, probabilities from 0.01 to 0.99 and confidence probabilities of 0.9, 0.95 and 0.99. The calculation time in the full range of all parameters is no more than 10–15 seconds on an average-performance computer. The calculation accuracy is about  $10^{-5}$ . It is noted that the main time expenditure is numerical integration due existing infinite integration limits, while minimization is quick (no more than 20–30 iterations). The paper also presents calculations results of relative variation coefficient quantiles for sample sizes of 3–10, general variation coefficients of 0.05, 0.3 and 0.5 and probabilities in the range from 0.01 to 0.99. There are also comparative calculations of the numerical characteristics of normal and Weibull order statistics obtained by direct integration and corresponding approximations. The programs under consideration use only the simplest and fairly accurate approximations of standard distributions: normal distribution, gamma function, incomplete gamma function. It is noted that the developed algorithms are suitable for a wide class of continuous distributions with their inverse functions that have no acceptable approximations.

**Keywords:** algorithms and programs, JavaScript, non-central Student distribution, variation coefficient exact distribution, order statistics distribution, normal distribution, Weibull distribution

#### References

1. Agamirov, L.V., Vestyak, V.A. (2023) *Probabilistic Methods for Calculating Reliability Indicators of Aircraft Structures under Variable Loads*. Moscow, 256 p. (in Russ.).
2. Tothfalusi, L., Endrenyi, L. (2016) 'An exact procedure for the evaluation of reference-scaled average bioequivalence', *AAPS J.*, 18(2), pp. 476–489. doi: 10.1208/s12248-016-9873-6.
3. Lehmann, E.L. (1986) *Testing Statistical Hypothesis*. NY: John Wiley & Sons Publ., 388 p.
4. Kholkina, A.V. (2019) 'Modeling random variables using the inverse function method', *Proc. VII Int. Youth Sci. Conf. Mathematical Support and Software for Information, Technical and Economic Systems*, pp. 73–75 (in Russ.).
5. Tsarev, A.D. (2020) 'Random walk software generation', *Bull. KRASEC. Phys. and Math. Sci.*, 31(2), pp. 226–235 (in Russ.). doi: 10.26117/2079-6641-2020-31-2-226-235.

6. Popov, G.A. (2018) 'Formula of inversion for rational characteristic functions of probability distributions', *Bull. of ASTU*, 2018(2), pp. 226–235 (in Russ.). doi: 10.26117/2079-6641-2020-31-2-226-235.
7. Gil, A., Segura, J., Temme, N.M. (2023) 'New asymptotic representations of the noncentral t-distribution', *Studies in Applied Math.*, 151(3), pp. 857–882. doi: 10.1111/sapm.12609.
8. Gil, A., Segura, J., Temme, N.M. (2019) 'On the computation and inversion of the cumulative noncentral beta distribution', *Appl. Math. Comput.*, 361, pp. 74–86.
9. Gil, A., Segura, J., Temme, N.M. (2022) 'A new asymptotic representation and inversion method for the Student's t distribution', *Integral Transforms and Special Functions*, 33(8), pp. 597–608. doi: 10.1080/10652469.2021.2007906.
10. Polukhin, A. (2020) *Development of Applications in C++ Using Boost*. Moscow, 346 p. (in Russ.).
11. Pearson, E.S., Hartley, H.O. (1976) *Biometrika Tables for Statisticians*. Cambridge University Press, 286 p.
12. Bolshev, L.N., Smirnov, N.V. (1983) *Tables of Mathematical Statistics*. Moscow, 416 p. (in Russ.).
13. *1130 Sci. Subroutine Package. Programmer's Manual, Program Number 1130-CM-02X* (1967) IBM Corporation, 191 p.
14. Agamirov, L.V., Vestyak, V.A. (2022) *A Program for Calculating Inverse Functions of Complex Statistical Distributions*. Pat. RF, № 2022612358.

**Авторы**

**Агамиров Левон Владимирович**<sup>1,2</sup>, д.т.н., профессор, itno\_agamirov@mail.ru  
**Агамиров Владимир Леонович**<sup>1,2</sup>, к.т.н., доцент, avl095@mail.ru  
**Вестяк Владимир Анатольевич**<sup>1</sup>, д.ф.-м.н., доцент, заведующий кафедрой, kaf311@yandex.ru

<sup>1</sup> Московский авиационный институт (Национальный исследовательский университет), г. Москва, 125993, Россия

<sup>2</sup> Московский технический университет связи и информатики (МТУСИ), г. Москва, 111024, Россия

**Authors**

**Levon V. Agamirov**<sup>1,2</sup>, Dr.Sc. (Engineering), Professor, itno\_agamirov@mail.ru  
**Vladimir L. Agamirov**<sup>1,2</sup>, Cand. of Sci. (Engineering), Associate Professor, avl095@mail.ru  
**Vladimir A. Vestyak**<sup>1</sup>, Dr.Sc. (Physics and Mathematics), Associate Professor, Head of Chair, kaf311@yandex.ru

<sup>1</sup> Moscow Aviation Institute (National Research University), Moscow, 125993, Russian Federation

<sup>2</sup> Moscow Technical University of Communications and Informatics, Moscow, 111024, Russian Federation

## Глобальная оптимизация на основе гибридизации алгоритмов роя саранчи и колонии пауков

С.И. Родзин <sup>1</sup>✉<sup>1</sup> Южный федеральный университет, г. Таганрог, 347922, Россия

### Ссылка для цитирования

Родзин С.И. Глобальная оптимизация на основе гибридизации алгоритмов роя саранчи и колонии пауков // Программные продукты и системы. 2024. Т. 37. № 2. С. 146–154. doi: 10.15827/0236-235X.142.146-154

### Информация о статье

Группа специальностей ВАК: 1.2.1

Поступила в редакцию: 08.01.2024

После доработки: 22.02.2024

Принята к публикации: 06.03.2024

**Аннотация.** Перспективным решением задач глобальной оптимизации являются метаэвристики, вдохновленные природой. Они представляют собой недетерминированные алгоритмы, исследующие пространство поиска решений, обучающиеся в процессе поиска, не привязанные к конкретной задаче, хотя и не гарантирующие точного решения. Целью данного исследования является разработка эффективного алгоритма для решения прикладных проблем глобальной оптимизации многомерных мультиэкстремальных функций, встречающихся в задачах вычислительной филогенетики, при проектировании электрических схем, расчетах инженерной безопасности зданий, калибровке моделей распространения радиосигналов и в других. Для достижения этой цели предложен гибридный алгоритм, моделирующий паттерны поведения роя саранчи и колонии пауков. Основное внимание уделено снижению вероятности преждевременной сходимости гибридного алгоритма, поддержке баланса между скоростью сходимости алгоритма и диверсификацией пространства поиска решений (интенсификация/диверсификация). Приводятся этапы модифицированных алгоритмов колонии пауков и роя саранчи, моделирующих различные паттерны их поведения, что позволяет снизить влияние очень хороших или плохих решений на процесс поиска. Гибридизация алгоритмов осуществляется путем их последовательной комбинации (препроцессор/постпроцессор). Алгоритм экспериментально протестирован на семи известных многомерных функциях. Результаты сопоставлены с конкурирующими алгоритмами роя частиц, дифференциальной эволюции, колонии пчел. Предлагаемый алгоритм обеспечивает лучшие результаты для всех рассмотренных функций. Проверка полученных результатов с использованием Т-критерия суммы рангов Уилкоксона для независимых выборок показала, что результаты по алгоритму являются статистически значимыми. Разработанное программное приложение предназначено для использования в рамках университетского курса по машинному обучению и биоинспирированной оптимизации, а также для решения широкого круга научно-прикладных задач поисковой оптимизации.

**Ключевые слова:** алгоритм, глобальный оптимум, агент, рой саранчи, колония пауков, паттерн поведения, интенсификация поиска, диверсификация поиска, тестовая функция, критерий Уилкоксона

**Благодарности.** Исследование выполнено в Южном федеральном университете за счет гранта РФФИ № 23-21-00089, <https://rscf.ru/project/23-21-00089/>

**Введение.** Эвристические алгоритмы, вдохновленные природой, вместе с нейронными сетями и другими методами образуют класс алгоритмов машинного обучения. Для них характерны исследование пространства решений, поиск точек в области определения функции, на которых она достигает минимального или максимального значения, их оценка и селективный отбор. Алгоритм обучается нахождению областей определения функции, которые содержат наилучшие решения. Поиск по БД Web of Science, Scopus, Google Scholar, цифровой библиотеке IEEE Xplore, российским библиотекам eLibrary.Ru и «КиберЛенинка» позволяет говорить о существовании более 400 алгоритмов, вдохновленных природой [1]. Их число и показатели цитируемости растут: 85 % статей, в которых предлагаются алгоритмы, вдохновленные природой, цитируются в среднем около 20 раз, в то время как

среднегодовая норма цитирования статей в целом в области Artificial Intelligence составляет около 5 [2]. С 2010 г. по настоящее время лидерами по количеству публикаций являются роевые алгоритмы. Они применяются практически во всех областях науки, техники и экономики. В этих областях многие задачи оптимизации и проектирования нелинейные, зачастую NP-трудные. Между тем такие алгоритмы, как оптимизация роя частиц (PSO), дифференциальная эволюция (DE), колония муравьев (ACO), пчел (ABC), светлячков (FA), поиска кукушки (CS) и многие другие продемонстрировали большой потенциал при решении сложных задач поисковой оптимизации за разумное время [3].

Чтобы алгоритм, вдохновленный природой, был эффективным, он должен обладать некоторыми особыми возможностями. Например, генерировать новые решения, улучшающие существующие путем интенсивного ис-

следования локального пространства вокруг найденного текущего хорошего решения (интенсификация). С другой стороны, алгоритм должен быть способен избегать преждевременной сходимости в локальном оптимуме путем расширения области поиска решений и обнаруживать области поиска, где может находиться глобальный оптимум (диверсификация). Удачная комбинация интенсификации и диверсификации поиска способствует высокой эффективности метаэвристики. Поиск баланса этих двух важных компонент любого алгоритма, инспирированного природой, является открытой исследовательской проблемой. Каждая биоэвристика использует различный баланс между ними, зачастую далекий от оптимального.

В данной работе предлагается гибридный алгоритм глобальной оптимизации, основанный на модифицированных алгоритмах роя саранчи и колонии пауков, а также исследуется его эффективность на ряде известных тестовых задач.

### Модифицированный алгоритм роя саранчи

Инспирированные природой алгоритмы, описывающие коллективные разумные паттерны поведения в мире животных и насекомых, привлекают внимание исследователей с конца прошлого века. Стереотипные поведенческие паттерны, наблюдаемые в группах животных и скоплениях насекомых, дают преимущества для выживания. Группа особей может решать достаточно сложные задачи, хотя отдельные особи, составляющие группу, относительно просты, используют ограниченную локальную информацию, допускают специализацию задач.

Саранча, являясь своего рода крупным кузнечиком, представляет пример таких групп насекомых. Саранча демонстрирует как одиночный, так и роевой паттерны поведения с их четкими поведенческими различиями [4]. При одиночном поведении, когда много пищи и места, саранча избегает контакта, поэтому рой исследует пространство, распределившись по всей площади. Паттерн роевого поведения, наоборот, означает стремление саранчи к агрегации вокруг перспективных источников пищи [5]. Использование в алгоритме оптимизации такого рода специализированных паттернов поведения позволяет поддерживать необходимый баланс между интенсификацией и диверсифи-

кацией поиска, избегать преждевременной сходимости к неоптимальным решениям.

В алгоритме это достигается путем включения операторов, моделирующих роевые и одиночные паттерны поведения агентов популяции. Речь идет о тех, кто отвечает за изменение позиций особей роя саранчи на текущей итерации алгоритма.

В природе налетевший рой саранчи, механизм образования которого достаточно сложен, способен уничтожить огромную площадь посевов за несколько часов. В работе [6] представлен алгоритм оптимизации, моделирующий поиск пищи роем саранчи, а в [5] – биологическая модель одиночного и роевого поведения саранчи, которая применяется при разработке модифицированного алгоритма роя саранчи.

Рассмотрим паттерн одиночного поведения особи саранчи и связанное с ним изменение ее позиции. Обозначим через  $x_i^k$  позицию  $i$ -й особи в рое из  $N$  особей на  $k$ -м шаге алгоритма. Позиция  $i$ -й особи на следующем шаге алгоритма  $x_i^{k+1}$  определяется как

$$x_i^{k+1} = x_i^k + \Delta x_i,$$

где величина  $\Delta x_i$  отражает изменение позиции  $i$ -й саранчи.

Паттерн одиночного поведения предполагает, что пара особей избегает сближения при условии короткой дистанции между ними. Если же дистанция между особями велика, то они начинают сближаться, поддерживая сплоченность роя. Следуя [7], определим силу сближения/удаления саранчи следующим образом:

$$s(r) = kar e^{-r/lim} - e^{-r}, \quad (1)$$

где  $r$  – дистанция между парой особей;  $kar$  – коэффициент сближения/удаления;  $lim$  – пороговая дистанция между особями. Сила удаления превышает силу сближения, если  $kar < 1$  и  $lim > 1$ , то есть дистанция между особями небольшая. Иными словами, сила взаимодействия между парой особей зависит от дистанции между ними:

$$s_{ij} = s(r_{ij})d_{ij},$$

где  $r_{ij} = |x_j - x_i|$  – дистанция между  $i$ -й и  $j$ -й особями роя;  $d_{ij} = (x_j - x_i)/r_{ij}$  – единичный вектор.

Тогда сумма всех сил взаимодействия с роем для  $i$ -й особи дает общую силу сближения/удаления роя и определяет изменение позиции  $i$ -й саранчи:

$$S_i = \sum_{i=1, j \neq i}^N s_{ij}. \quad (2)$$

Триггером перехода от одиночного поведения саранчи к роевому являются изменения

в окружающей среде. Рой в поисках пищи ведет себя очень сплоченно. Характерная черта паттерна роевого поведения заключается в стремлении роя сконцентрироваться вблизи особей, которым удалось найти пищу. Эту особенность роевого поведения можно смоделировать путем введения для каждой особи роя саранчи индикатора пищи  $f_i (f_i \in [0, 1])$ . После сортировки  $N$  особей роя производится отбор  $b$  доминирующих особей ( $b \ll N$ ) с наибольшим значением индикатора  $f_i$ , вблизи которых концентрируются остальные особи.

Рой саранчи из  $N$  особей образует популяцию решений  $L^k(\{l_1^k, l_2^k, \dots, l_N^k\})$  оптимизационной задачи, где число итераций  $k = 0, 1, 2, \dots, gen$ . Особь саранчи  $l_i^k (i = 1..N)$  характеризуется через множество  $\{l_{i1}^k, l_{i2}^k, \dots, l_{in}^k\}$ . Элемент этого множества отображает отдельную переменную оптимизационной задачи. Ограничения в задаче образуют множество допустимых решений:

$$S = \{l_i^k \in R^n \mid lb_d \leq l_{id}^k \leq ub_d\},$$

где  $lb_d$  – нижняя граница размерности  $d$ ;  $ub_d$  – верхняя граница размерности  $d$ . Фитнесс-функция  $f_i(l_i^k)$  характеризует уровень индикатора пищи каждой саранчи.

Таким образом, *модифицированный алгоритм роя саранчи* (МАРС) использует оператор  $O_{од}$ , характеризующий паттерн одиночного поведения саранчи, и  $O_{рой}$ , характеризующий паттерн роевого поведения саранчи. Оператор  $O_{од}$  способствует диверсификации области поиска решений и обнаружению точки глобального оптимума, а оператор  $O_{рой}$  – интенсивному исследованию локального пространства вокруг найденного текущего хорошего решения.

С помощью оператора  $O_{од}$  позиция  $l_i^k$   $i$ -й особи изменяется на величину  $\Delta l_i^k$  и становится равной

$$p_i = l_i^k + \Delta l_i^k.$$

При этом учитывается позиция доминирующих особей роя с наибольшим значением фитнесс-функции. Тогда сила сближения/удаления между парой особей ( $i, j$ ) вычисляется следующим образом:

$$s_{ij}^m = \rho(l_i^k, l_j^k) s(r_{ij}) d_{ij} + rand(1, -1),$$

где  $s(r_{ij})$  вычисляется согласно (1);  $d_{ij}$  направлен от  $l_i^k$  к  $l_j^k$ ; функция  $rand(-1, 1)$  генерирует случайное число из интервала  $(-1, 1)$ ; функция  $\rho(l_i^k, l_j^k)$  указывает на доминирование между парой особей ( $i, j$ ).

Особи с наибольшим значением фитнесс-функции присваивается ранг 0, а с наименьшим –

ранг  $N-1$ . Функция доминирования  $\rho$  определяется следующим образом:

$$\rho(l_i^k, l_j^k) = \begin{cases} e^{-\frac{5rank(l_i^k)}{N}}, & \text{если } rank(l_i^k) < rank(l_j^k), \\ e^{-\frac{5rank(l_j^k)}{N}}, & \text{если } rank(l_i^k) > rank(l_j^k), \end{cases} \quad (3)$$

где функция  $rank(\alpha)$  указывает на ранг особи.

Согласно (3), функция  $\rho$  принимает значения из интервала  $(0, 1)$ . Значение  $\rho = 1$  в случае, если особь имеет наилучшее значение фитнесс-функции в популяции. Значение  $\rho$ , близкое к 0, свидетельствует о низком значении фитнесс-функции.

После обновления множества позиций  $\{p_1, p_2, \dots, p_N\}$  особей колонии  $L^k$  корректируются значения множества фитнесс-функций  $F(f_1, f_2, \dots, f_N)$ . Причем результаты поиска должны улучшаться: если  $f_i(p_i) > f_i(l_i^k)$ , то особь перемещается в новую позицию  $p_i$ , в противном случае позиция  $l_i^k$  остается неизменной.

С помощью оператора  $O_{рой}$  производится интенсивное исследование локального пространства вокруг найденного текущего хорошего решения. Это происходит путем упорядочения фитнесс-функций особей по убыванию. Затем результаты упорядочения запоминаются в массиве  $B = \{b_1, b_2, \dots, b_N\}$ . Из массива  $B$  выбираются  $g$  особей с наилучшими значениями фитнесс-функции, которые составляют множество  $E$  наиболее перспективных для дальнейшего исследования решений. Исследование производится в окрестности каждой  $i$ -й особи из множества  $E$ . Радиус окрестности вычисляется по формуле

$$e_d = \frac{\sum_{q=1}^n (ub_q - lb_q)}{n} \beta,$$

где  $ub_q$  и  $lb_q$  – верхняя и нижняя границы в  $q$ -м измерении соответственно;  $n$  – размерность переменных в задаче оптимизации;  $\beta \in [0, 1]$  – параметр алгоритма. Верхняя граница  $uss^q_j$  и нижняя граница  $lss^q_j$  окрестности  $C_j$  в  $q$ -м измерении вычисляются по следующим формулам:

$$uss_j^q = b_{j,q} + e_d,$$

$$lss_j^q = b_{j,q} - e_d.$$

В окрестности  $C_j$  создаются  $h (h < 4)$  новых особей. Победителем является особь с наилучшим значением фитнесс-функции.

Таким образом, управляющими параметрами алгоритма МАРС являются  $kar$  – коэффи-

циент сближения/удаления,  $lim$  – пороговая дистанция между особями,  $g$  – число особей с наилучшими фитнес-функциями,  $N$  – размер популяции и  $gen$  – число поколений.

В процессе инициализации алгоритма МАРС устанавливается начальное значение счетчика поколений ( $k = 0$ ) и случайно генерируется начальная популяция  $L^0(\{I_1^0, I_2^0, \dots, I_N^0\})$ . При этом для измерения  $d$  предварительно задаются верхняя начальная граница параметра ( $ub_d$ ) и нижняя начальная граница параметра ( $lb_d$ ), в которых значения ( $\{li_1^0, li_2^0, \dots, li_n^0\}$ ) распределены случайно и равномерно:

$$li_j^0 = lb_d + rand(ub_d - lb_d),$$

где  $i = 1..N, d = 1..n$ .

Далее в алгоритме выполняются операторы  $O_{од}$ , характеризующий паттерн одиночного поведения саранчи, и  $O_{рой}$ , характеризующий паттерн роевого поведения саранчи для поддержки баланса интенсификации/диверсификации.

Останов алгоритма происходит при достижении числа итераций  $k = gen$ .

Представим псевдокод МАРС:

1: Ввод:  $kar, lim, g, N, gen$

2: Инициализация  $L^0 (k = 0)$

3: **until** ( $k = gen$ )

4:  $F \leftarrow$  оператор  $O_{од} (L^k)$

5:  $L^{k+1} \leftarrow$  оператор  $O_{рой} (L^k, F)$

6:  $k = k + 1$

7: **end until**

### Модифицированный алгоритм колонии пауков

В работе [8] представлен алгоритм оптимизации, моделирующий кооперативное поведение колонии социальных пауков, а в [9] – биологическая модель их взаимодействия. Используем эту модель для представления паттернов поведения в распределенной самоорганизующейся сети и построения *модифицированного алгоритма колонии пауков* (МАКП).

Основными функциями, характеризующими поведение пауков в колонии, являются строительство сетевой паутины, размножение и охота. При этом инструментом взаимодействия агрегации пауков является сеть. Она представляет собой канал связи через вибрации и постукивания. Вибрации используются пауками для синхронизации и декодирования своих действий, а также для авторизации особи, передающей сообщение. По интенсивности вибраций и постукиваний определяются вес паука и расстояние до него. Каждый шаг паука вызывает вибрацию. Поэтому, например, при охоте коло-

ния перестает двигаться, чтобы по сети почувствовать добычу и понять правильное направление движения. Колония использует определенную тактику и командную работу для защиты своих ресурсов и потомства с дифференциацией ролей.

Согласно биологической модели определим паттерны поведения пауков – размножение и кооперация [10]. Альфа-самцы пауков отличаются большим весом в сравнении с остальными самцами. При размножении они стремятся двигаться по сети к ближайшей самке. В отличие от альфа-самцов недоминирующие самцы в основном сосредоточиваются в центре колонии, чтобы воспользоваться ресурсами альфа-самцов.

С учетом этого МАКП включает следующую последовательность шагов.

*Шаг 1.* Инициализация популяции пауков  $S = B \cup M$  размером  $N$ . Поскольку, согласно биологической модели, в колонии преобладают самки, то их число  $N_b$  определяется по формуле

$$N_b = floor [(0,9 - rand(0,25)N),$$

где  $rand$  – случайное число в интервале  $[0, 1]$ ;  $floor(\cdot)$  – функция преобразования действительных чисел в целые. Общее число самцов  $N_m = N - N_b$ .

В целом колония социальных пауков  $S$  размером  $N$  состоит из подмножеств  $B = \{b_1, b_2, \dots, b_{N_b}\}$  и  $M = \{m_1, m_2, \dots, m_{N_m}\}$ .

*Шаг 2.* Случайная инициализация позиций пауков. Позиции самцов  $b_i$  и самок  $m_j$  генерируются случайно и равномерно. Для этого задаются нижний начальный параметр  $q_k^{ниж}$  и верхний начальный параметр  $q_k^{верх}$ :

$$b_{i,k}^0 = q_k^{ниж} + rand(0,1)(q_k^{верх} - q_k^{ниж}),$$

$$m_{j,k}^0 = q_k^{ниж} + rand(0,1)(q_k^{верх} - q_k^{ниж}),$$

где  $i = 1..N_b, j = 1..N_m, k = 1..n$ ; функция  $rand(0, 1)$  генерирует случайное число в интервале от 0 до 1.

*Шаг 3.* Диапазон взаимодействия в колонии социальных пауков определяется размерами пространства поиска по формуле

$$r = \frac{\sum_{k=1}^n (q_k^{верх} - q_k^{ниж})}{2n}.$$

*Шаг 4.* Вероятность участия в процессе размножения зависит от веса паука и вычисляется по методу рулетки:

$$P_{s_i} = \frac{w_i}{\sum_{k \in T^s} w_k},$$

где  $T^s$  – множество пауков в радиусе  $r$ . При этом сохраняется первоначальная пропорция между самками и самцами в популяции  $S$ .

**Шаг 5.** Позиция самки  $b_i$  на итерации алгоритма ( $t + 1$ ) изменяется по формулам:

$$b_i^{t+1} = b_i^t + \alpha Vib_{i,u}(m_u - b_i^t) + \beta Vib_{i,maxw}(m_{maxw} - b_i^t) + \delta(rand - 1/2)$$

с вероятностью PF,

$$b_i^{t+1} = b_i^t + \alpha Vib_{i,u}(m_u - b_i^t) + \beta Vib_{i,maxw}(m_{maxw} - b_i^t) + \delta(rand - 1/2)$$

с вероятностью  $1 - PF$ .

Здесь моделируются вибрации паутины. Вибрации зависят от веса и расстояния до паука, который их породил. Иными словами, особи рядом с пауком, создающим вибрации, улавливают их как более сильные, нежели особи, находящиеся вдали. Вибрация, которую  $i$ -й паук улавливает от  $j$ -го паука, определяется как

$$Vib_{i,j} = w_j e^{-d_{i,j}^2},$$

где  $d_{i,j} = \|s_i - s_j\|$  – евклидово расстояние между  $i$ -м и  $j$ -м пауками.

Модификация алгоритма заключается в использовании в МАКП следующих вариантов вибраций:

- $Vibc_i = w_c e^{-d_{ci}^2}$  – вибрации, улавливаемые  $i$ -м пауком от ближайшего к нему паука  $c$ , обладающего большим весом ( $w_c > w_i$ );

- $Vibb_{i,maxw} = w_{maxw} e^{-d_{bi,maxw}^2}$  – вибрации, улавливаемые  $i$ -м пауком от паука  $maxw$  с максимальным весом в колонии;

- $Vibf_i = w_f e^{-d_{fi}^2}$  – вибрации, улавливаемые  $i$ -м пауком от ближайшей самки  $f$ .

**Шаг 6.** Выполняется оператор, моделирующий движение самца паука:

$$m_i^{t+1} = \begin{cases} m_i^t + \alpha Vib_{i,b}(s_b - m_i^t) + \delta\left(rand - \frac{1}{2}\right), & \text{если } w_{N_{cp}} > w_i; \\ m_i^t + \alpha \left( \frac{\sum_{h=1}^{N_m} m_h^t w_{N_{b+h}}}{\sum_{h=1}^{N_m} w_{N_{b+h}}} - m_i^t \right), & \text{если } w_i > w_{N_{cp}} \end{cases}$$

где  $s_b$  – ближайшая к пауку самка, а величина

$$\left( \frac{\sum_{h=1}^{N_m} m_h^t w_{N_{b+h}}}{\sum_{h=1}^{N_m} w_{N_{b+h}}} - m_i^t \right)$$

представляет средневзвешенный вес самцов  $M$  в колонии.

Оператор определяет различные поведенческие паттерны, что способствует диверсификации поиска оптимума. К тому же уменьшается влияние очень хороших и плохих решений на результаты поиска.

**Шаг 7.** Выполняется оператор размножения в определенном диапазоне  $r$ , который вычисляется как

$$P_{s_i} = \frac{w_i}{\sum_{k \in T^r} w_k},$$

где  $T^r$  – множество пауков в радиусе  $r$ , которые участвуют в размножении.

**Шаг 8.** Останов алгоритма происходит при достижении заданного максимального числа итераций. Иначе – возврат к шагу 4 алгоритма.

### Гибридизация алгоритмов МАРС и МАКП

Комбинирование нескольких алгоритмов, инспирированных природой, при решении целого класса сложных задач глобальной оптимизации может оказаться более эффективным, нежели использование одного алгоритма. Перспективным направлением здесь является гибридизация различных алгоритмов глобальной оптимизации. Существуют различные способы гибридизации алгоритмов. Одним из них является последовательная комбинация алгоритмов (препроцессор/постпроцессор). Основная идея этого способа гибридизации заключается в том, чтобы на начальных этапах обеспечить широкий обзор всевозможных решений, а на последующих сузить области поиска.

Сценарий предлагаемой гибридизации включает поиск решений с помощью алгоритма МАКП в качестве препроцессора, а затем проверку и выявление лучшего решения с помощью алгоритма МАРС. Триггером перехода с алгоритма препроцессора является отсутствие улучшения глобального оптимума. Алгоритм постпроцессора стартует, используя в основном решения, полученные на заключительном шаге алгоритма препроцессора с исключением решений, близких к области одного и того же экстремума.

### Результаты экспериментальных исследований эффективности гибридного алгоритма

Для оценки эффективности предлагаемого гибридного алгоритма глобальной оптимизации, основанного на МАКП и МАРС, были проведены эксперименты с использованием тестовых многомерных функций.

Эксперименты проводились в программной среде на языке программирования C# с использованием среды программирования приложе-

ний Microsoft Visual Studio. Это позволило воспользоваться возможностями, предоставляемыми объектно-ориентированным подходом в разработке ПО, а также максимально использовать фреймворк Windows Forms (зарегистрирована программа для ЭВМ, запись № 2023681642 от 17.10.2023 г.). При отладке и тестировании использован компьютер IBM PC с процессором Core i7 и ОЗУ 8 Гб.

Набор тестовых функций, на которых проводилось программное моделирование, включал как одноэкстремальные, так и мультиэкстремальные функции:

- функция Розенброка

$$f_1(X) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2];$$

$$x_i \in [-30, 30]^n, x^* = (1, \dots, 1), f_1(x^*) = 0;$$

- сферическая функция

$$f_2(X) = \sum_{i=1}^n x_i^2, x_i \in [-100, 100]^n, x^* = (0, \dots, 0),$$

$$f_2(x^*) = 0;$$

- функция Экли

$$f_3(X) = -20 \exp\left(\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}\right) -$$

$$- \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right) + 20;$$

$$x_i \in [-32, 32]^n, x^* = (0, \dots, 0), f_3(x^*) = 0;$$

- функция Швевеля

$$f_4(X) = \sum_{i=1}^n \left(\sum_{j=1}^i x_j\right)^2,$$

$$x_i \in [-100, 100]^n, x^* = (0, \dots, 0), f_4(x^*) = 0;$$

- сумма квадратов

$$f_5(X) = \sum_{i=1}^n ix_i^2, x_i \in [-10, 10]^n, x^* = (0, \dots, 0),$$

$$f_5(x^*) = 0;$$

- функция Растригина

$$f_6(X) = \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i) + 10],$$

$$x_i \in [-5, 12; 5, 12]^n, x^* = (0, \dots, 0), f_6(x^*) = 0;$$

- функция Соломона

$$f_7(X) = -\cos\left(2\pi \sqrt{\sum_{i=1}^n x_i^2}\right) + 0,1 \sqrt{\sum_{i=1}^n x_i^2} + 1,$$

$$x_i \in [-100, 100]^n, x^* = (0, \dots, 0), f_7(x^*) = 0,$$

где  $n$  – размерность функции;  $x^*$  – оптимальное решение;  $f_i(x^*)$  – минимальное значение функции.

Например, график функции Растригина представляет собой сочетание сферической и синусоидной функций с множеством локальных и одним глобальным минимумом, «банановая» функция Розенброка – большое медленно убывающее плато с глобальным минимумом внутри параболической сильно вытянутой поверхности, функция Швевеля является мультиэкстремальной с «непредсказуемым» глобальным минимумом; функция Соломона – непрерывная, мультиэкстремальная, дифференцируемая. Подробное описание множества тестовых функций содержится в [11].

Результаты тестирования гибридного алгоритма МАКП-МАРС сопоставлялись с алгоритмами пчелиной колонии (Artificial Bee Colony, ABC), дифференциальной эволюции (Differential Evolution, DE), роя частиц (Particle Swarm Optimization, PSO) [12–15], МАРС, МАКП.

В экспериментах использовались следующие настройки параметров алгоритмов: ABC – параметр  $limit = 100$ ; DE – параметр  $F \in [0,4; 1,0]$ , параметр  $p_m \in [0,0; 1,0]$ ; PSO – коэффициенты обучения  $c_1 = 2$  и  $c_2 = 2$ , вес  $w$  линейно уменьшается с 0,9 до 0,2 при увеличении числа итераций; МАКП – параметр  $PF = 0,7$ ; МАРС – параметры  $kar = 0,6$ ,  $lim = 1$ ,  $N = 50$ ,  $g = 20$ ,  $gen = 1000$ .

В экспериментах использовались следующие настройки параметров алгоритмов: ABC – параметр  $limit = 100$ ; DE – параметр  $F \in [0,4; 1,0]$ , параметр  $p_m \in [0,0; 1,0]$ ; PSO – коэффициенты обучения  $c_1 = 2$  и  $c_2 = 2$ , вес  $w$  линейно уменьшается с 0,9 до 0,2 при увеличении числа итераций; МАКП – параметр  $PF = 0,7$ ; МАРС – параметры  $kar = 0,6$ ,  $lim = 1$ ,  $N = 50$ ,  $g = 20$ ,  $gen = 1000$ .

По каждой функции и каждому алгоритму осуществлялось 30 прогонов. Затем полученные результаты усреднялись.

Показателями, по которым проводилось сравнение, являлись среднее значение по лучшим решениям, медианное лучшее решение и стандартное отклонение от лучшего решения [16]. Усредненные результаты по 30 отдельным запускам приведены в таблице.

В каждой клетке таблицы указаны среднее, медианное решение и стандартное отклонение от лучшего решения соответственно.

Лучшие результаты выделены жирным шрифтом. Из таблицы видно, что гибридный алгоритм МАКП-МАРС обеспечивает лучшие результаты, нежели алгоритмы ABC, DE, PSO, МАКП, МАРС, практически для всех указанных функций.

Также оценивалась статистическая значимость полученных результатов. С этой целью применялся T-критерий суммы рангов Уилкоксона [17] для независимых выборок, найденных каждым из сравниваемых алгоритмов на 30 тестовых запусках, при уровне значимости 5%. Значение  $T < 0,05$  рассматривалось как адекватное доказательство против нулевой гипотезы, которая отвергается. Предложенный гибридный алгоритм МАКП-МАРС превосходит конкурирующие, а экспериментальные результаты по нему являются статистически значимыми.

**Результаты сравнения гибридного алгоритма МАКП-МАРС с алгоритмами ABC, DE, PSO, МАРС, МАКП**

**Results of comparing MAKP-MARS hybrid algorithm with the ABC, DE, PSO, MARS, MAKP algorithms**

Функция	ABC	DE	PSO	МАРС	МАКП	МАКП-МАРС
$f_1(X)$ роз	$9,26 \cdot 10^{-02}$	$2,27 \cdot 10^{-02}$	$2,73 \cdot 10^{-02}$	$1,47 \cdot 10^{-02}$	$8,77 \cdot 10^{-02}$	<b><math>1,02 \cdot 10^{-02}</math></b>
	$3,24 \cdot 10^{-01}$	$2,23 \cdot 10^{-02}$	$2,61 \cdot 10^{-02}$	$1,86 \cdot 10^{-02}$	$5,13 \cdot 10^{-02}$	<b><math>8,15 \cdot 10^{-03}</math></b>
	$1,81 \cdot 10^{-01}$	$5,03 \cdot 10^{-03}$	<b><math>5,83 \cdot 10^{-03}</math></b>	$6,33 \cdot 10^{-03}$	$2,68 \cdot 10^{-02}$	$7,17 \cdot 10^{-03}$
$f_2(X)$ сф	$7,08 \cdot 10^{-03}$	$1,97 \cdot 10^{-05}$	$8,44 \cdot 10^{-03}$	$9,23 \cdot 10^{-06}$	$5,12 \cdot 10^{-05}$	<b><math>3,38 \cdot 10^{-07}</math></b>
	$5,16 \cdot 10^{-03}$	$5,43 \cdot 10^{-05}$	$3,12 \cdot 10^{-02}$	<b><math>7,95 \cdot 10^{-06}</math></b>	$4,12 \cdot 10^{-05}$	$9,49 \cdot 10^{-06}$
	$1,86 \cdot 10^{-03}$	$1,03 \cdot 10^{-05}$	$1,60 \cdot 10^{-03}$	$2,17 \cdot 10^{-06}$	$8,82 \cdot 10^{-06}$	<b><math>1,12 \cdot 10^{-06}</math></b>
$f_3(X)$ экл	$6,26 \cdot 10^{-02}$	$7,01 \cdot 10^{-04}$	$3,57 \cdot 10^{-02}$	<b><math>3,43 \cdot 10^{-05}</math></b>	$4,22 \cdot 10^{-05}$	$4,05 \cdot 10^{-05}$
	$5,95 \cdot 10^{-02}$	$7,20 \cdot 10^{-04}$	$4,82 \cdot 10^{-02}$	$2,65 \cdot 10^{-05}$	$1,52 \cdot 10^{-05}$	<b><math>2,02 \cdot 10^{-06}</math></b>
	$1,33 \cdot 10^{-03}$	$2,21 \cdot 10^{-04}$	$1,15 \cdot 10^{-03}$	$3,71 \cdot 10^{-06}$	$8,16 \cdot 10^{-06}$	<b><math>1,15 \cdot 10^{-06}</math></b>
$f_4(X)$ шв	$2,34 \cdot 10^{-01}$	$8,26 \cdot 10^{-01}$	$8,62 \cdot 10^{-01}$	$7,98 \cdot 10^{-02}$	$8,77 \cdot 10^{-02}$	<b><math>3,20 \cdot 10^{-03}</math></b>
	$6,77 \cdot 10^{-01}$	$7,35 \cdot 10^{-01}$	$5,25 \cdot 10^{-01}$	$6,62 \cdot 10^{-02}$	$5,13 \cdot 10^{-02}$	<b><math>7,57 \cdot 10^{-03}</math></b>
	$2,72 \cdot 10^{-01}$	$1,66 \cdot 10^{-01}$	$1,12 \cdot 10^{-01}$	$2,76 \cdot 10^{-02}$	$2,68 \cdot 10^{-02}$	<b><math>2,82 \cdot 10^{-03}</math></b>
$f_5(X)$ ква	$2,47 \cdot 10^{-03}$	$2,47 \cdot 10^{-03}$	$6,96 \cdot 10^{-02}$	$9,99 \cdot 10^{-04}$	$8,98 \cdot 10^{-04}$	<b><math>1,46 \cdot 10^{-05}</math></b>
	$5,72 \cdot 10^{-03}$	$5,70 \cdot 10^{-03}$	$5,49 \cdot 10^{-02}$	<b><math>5,17 \cdot 10^{-04}</math></b>	$6,62 \cdot 10^{-04}$	$7,38 \cdot 10^{-04}$
	$6,63 \cdot 10^{-04}$	<b><math>1,58 \cdot 10^{-04}</math></b>	$3,33 \cdot 10^{-02}$	$1,76 \cdot 10^{-04}$	$1,76 \cdot 10^{-04}$	$1,76 \cdot 10^{-04}$
$f_6(X)$ рас	$9,58 \cdot 10^{-01}$	$8,26 \cdot 10^{-01}$	$8,48 \cdot 10^{-01}$	$8,14 \cdot 10^{-02}$	$7,88 \cdot 10^{-02}$	<b><math>2,82 \cdot 10^{-03}</math></b>
	$9,18 \cdot 10^{-01}$	$7,40 \cdot 10^{-01}$	$8,28 \cdot 10^{-01}$	$6,96 \cdot 10^{-02}$	$6,74 \cdot 10^{-02}$	<b><math>8,44 \cdot 10^{-03}</math></b>
	$2,19 \cdot 10^{-01}$	$2,28 \cdot 10^{-01}$	$1,16 \cdot 10^{-01}$	$5,39 \cdot 10^{-03}$	$2,13 \cdot 10^{-03}$	<b><math>1,28 \cdot 10^{-03}</math></b>
$f_7(X)$ сал	$4,21 \cdot 10^{-02}$	$6,52 \cdot 10^{-01}$	$9,35 \cdot 10^{-01}$	$7,56 \cdot 10^{-02}$	$8,24 \cdot 10^{-02}$	<b><math>4,81 \cdot 10^{-03}</math></b>
	$5,64 \cdot 10^{-01}$	$7,46 \cdot 10^{-01}$	$8,04 \cdot 10^{-01}$	$6,28 \cdot 10^{-02}$	$5,79 \cdot 10^{-03}$	<b><math>3,62 \cdot 10^{-03}</math></b>
	$3,49 \cdot 10^{-02}$	$3,35 \cdot 10^{-01}$	$2,48 \cdot 10^{-01}$	$2,24 \cdot 10^{-03}$	$3,20 \cdot 10^{-03}$	<b><math>2,03 \cdot 10^{-04}</math></b>

**Заклучение**

Результаты, полученные гибридным алгоритмом МАКП-МАРС на одноэкстремальных и многоэкстремальных многомерных тестовых функциях, превосходят результаты конкурирующих алгоритмов и являются статистически значимыми. Это связано с достигаемым балансом между скоростью сходимости алгоритма и диверсификацией пространства поиска решений.

Заслуживает внимания применение гибридного алгоритма МАКП-МАРС для решения задач динамической и стохастической оптимизации, многокритериальной оптимизации, мульти-модальной оптимизации, многомерной оптимизации, меметической оптимизации, в которой комбинируется множество поисковых алгоритмов, оптимизации и адаптации настроек параметров метаэвристик для достижения баланса между скоростью сходимости и диверсификацией пространства поиска решений.

**Список литературы**

1. Родзин С.И. Современное состояние биоэвристик: классификация, бенчмаркинг, области применения // Изв. ЮФУ. Технич. науки. 2023. № 2. С. 280–298. doi: 10.18522/2311-3103-2023-2-280-298.
2. Rajwar K., Deep K., Das S. An exhaustive review of the metaheuristic algorithms for search and optimization: Taxonomy, applications, and open challenges. Artificial Intelligence Review, 2023, vol. 56, pp. 13187–13257. doi: 10.1007/s10462-023-10470-y.
3. Molina D., Poyatos J., Ser J.D. et al. Comprehensive taxonomies of nature- and bio-inspired optimization: Inspiration versus algorithmic behavior, critical analysis recommendations. Cognitive Computation, 2020, vol. 12, pp. 897–939. doi: 10.1007/s12559-020-09730-8.
4. Cuevas E., Fausto F., Gonzalez A. The locust swarm optimization algorithm. In: New Advancements in Swarm Algorithms: Operators and Applications. ISRL, 2019, vol. 160, pp. 139–159. doi: 10.1007/978-3-030-16339-6\_5.
5. Camarena O., Cuevas E., Pérez-Cisneros M. et al. Ls-II: An improved locust search algorithm for solving optimization problems. Math. Problems in Eng., 2018, vol. 2018, art. 4148975. doi: 10.1155/2018/4148975.
6. Alekseev V. Application of artificial locust swarm routing algorithm for VFR flights planning. Math. Modeling, 2021, vol. 4, pp. 124–127.
7. Rodzin S., Kuliyeв E., Zaporozhets D., Rodzina L., Rodzina O. Locust swarm optimization algorithm: A bio-heuristic for global optimization problem. In: LNNS. Proc. CSOC, 2023, vol. 722, pp. 670–678. doi: 10.1007/978-3-031-35311-6\_64.

8. Luque-Chang A., Cuevas E., Fausto F. et al. Social spider optimization algorithm: modifications, applications, and perspectives. *Math. Problems in Eng.*, 2018, vol. 2018, art. 6843923. doi: 10.1155/2018/6843923.
9. Evangeline D., Abirami T. Social spider optimization algorithm: Theory and its applications. *IJITEE*, 2019, vol. 8, no. 10, pp. 327–332. doi: 10.35940/ijitee.I8261.0881019.
10. Rodzin S., Rodzina L. Spider colony optimization algorithm: A bio-heuristic for global optimization problem. In: *LNNS. Proc. CSOC*, 2023, vol. 722, pp. 661–669. doi: 10.1007/978-3-031-35311-6\_63.
11. Kerschke P., Trautmann H. Automated algorithm selection on continuous black-box problems by combining exploratory landscape analysis and machine learning. *Evolutionary Comput.*, 2019, vol. 27, no. 1, pp. 99–127. doi: 10.1162/evco\_a\_00236.
12. Long W., Wu T., Liang X., Xu S. Solving high-dimensional global optimization problems using an improved sine cosine algorithm. *Expert Systems with Applications*, 2019, vol. 123, pp. 108–126. doi: 10.1016/j.eswa.2018.11.032.
13. Dragoi E., Dafinescu V. Review of metaheuristics inspired from the animal kingdom. *Math.*, 2021, vol. 9, no. 18, art. 2335. doi: 10.3390/math9182335.
14. Benitez-Hidalgo A., Nebro A.J., Garcia-Nieto J., Oregi I., Ser J.D. jMetalPy: A Python framework for multi-objective optimization with metaheuristics. *Swarm and Evolutionary Computation*, 2019, vol. 51, art. 100598. doi: 10.1016/j.swevo.2019.100598.
15. Tian Y., Cheng R., Zhang X., Jin Y. A MATLAB platform for evolutionary multi-objective optimization. *IEEE Computational Intelligence Magazine*, 2017, vol. 12, no. 4, pp. 73–87. doi: 10.1109/MCI.2017.2742868.
16. Иванов Д.К., Думина Д.С., Семенов Н.А. Определение весовых коэффициентов для аддитивной фитнес-функции генетического алгоритма // Программные продукты и системы. 2020. Т. 33. № 1. С. 47–53. doi: 10.15827/0236-235X.129.047-053.
17. Saha S., Seal D.B., Ghosh A., Dey K.N. A novel gene ranking method using Wilcoxon rank sum test and genetic algorithm. *IJBRA*, 2016, vol. 12, no. 3, pp. 263–279. doi: 10.1504/IJBRA.2016.078236.

Software &amp; Systems

doi: 10.15827/0236-235X.142.146-154

2024, 37(2), pp. 146–154

### Global optimization based on hybridization of locust swarm and spider colony algorithm

Sergey I. Rodzin <sup>1</sup>✉<sup>1</sup> Southern Federal University, Taganrog, 347922, Russian Federation**For citation**Rodzin, S.I. (2024) 'Global optimization based on hybridization of locust swarm and spider colony algorithm', *Software & Systems*, 37(2), pp. 146–154 (in Russ.). doi: 10.15827/0236-235X.142.146-154**Article info**

Received: 08.01.2024

After revision: 22.02.2024

Accepted: 06.03.2024

**Abstract.** A promising solution to global optimization problems are metaheuristics inspired by nature. They are non-deterministic algorithms that explore solution search space, learn in the search process; they are not tied to a specific task, although they do not guarantee accurate solutions. The purpose of this study is to develop an effective algorithm for solving applied problems of global optimization of multidimensional multi extreme functions in computational phylogenetics problems, in designing electrical circuits, calculations of building engineering safety, calibration of radio propagation models, and others. To achieve this goal, the authors of the paper propose a hybrid algorithm that simulates behavior patterns of a locust swarm and a spider colony. The paper focuses on the issue of reducing the probability of hybrid algorithm premature convergence, maintaining a balance between an algorithm convergence rate and the diversification of a solution search space (intensification/diversification). The paper presents the stages of modified algorithms for a spider colony and a locust swarm that model various patterns of their behavior, which reduces the effect of very good or bad decisions on the search process. The algorithms are hybridized by their sequential combination (preprocessor/postprocessor). The algorithm was tested on seven known multidimensional functions. The results were compared with competing algorithms for a particle swarm, a differential evolution, and a bee colony. The proposed algorithm provides the best results for all considered functions. Verification of the results obtained using the Wilcoxon sum of ranks T-test for independent samples showed that the algorithm results are statistically significant. The developed software application is intended for using in terms of a university course on machine learning and bioinspired optimization, as well as for solving a wide range of scientific and applied problems of search engine optimization.

**Keywords:** algorithm, global optimum, agent, locust swarm, spider colony, behavior pattern, search intensification, search diversification, test function, Wilcoxon test

**Acknowledgements.** The study was carried out at the Southern Federal University and supported by the Russian Science Foundation grant no. 23-21-00089, <https://rscf.ru/project/23-21-00089/>

### References

1. Rodzin, S.I. (2023) 'Current state of bio heuristics: Classification, benchmarking, application areas', *Izv. SFedU. Eng. Sci.*, (2), pp. 280–298 (in Russ.). doi: 10.18522/2311-3103-2023-2-280-298.
2. Rajwar, K., Deep, K., Das, S. (2023) 'An exhaustive review of the metaheuristic algorithms for search and optimization: Taxonomy, applications, and open challenges', *Artificial Intelligence Review*, 56, pp. 13187–13257. doi: 10.1007/s10462-023-10470-y.
3. Molina, D., Poyatos, J., Ser, J.D. et al. (2020) 'Comprehensive taxonomies of nature- and bio-inspired optimization: Inspiration versus algorithmic behavior, critical analysis recommendations', *Cognitive Computation*, 12, pp. 897–939. doi: 10.1007/s12559-020-09730-8.
4. Cuevas, E., Fausto, F., Gonzalez, A. (2019) 'The locust swarm optimization algorithm', in *New Advancements in Swarm Algorithms: Operators and Applications. ISRL*, 160, pp. 139–159. doi: 10.1007/978-3-030-16339-6\_5.
5. Camarena, O., Cuevas, E., Pérez-Cisneros, M. et al. (2018) 'Ls-II: An improved locust search algorithm for solving optimization problems', *Math. Problems in Eng.*, 2018, art. 4148975. doi: 10.1155/2018/4148975.
6. Alekseev, V. (2021) 'Application of artificial locust swarm routing algorithm for VFR flights planning', *Math. Modeling*, 4, pp. 124–127.
7. Rodzin, S., Kuliyeu, E., Zaporozhets, D., Rodzina, L., Rodzina, O. (2023) 'Locust swarm optimization algorithm: A bio-heuristic for global optimization problem', in *LNNS. Proc. CSOC*, 722, pp. 670–678. doi: 10.1007/978-3-031-35311-6\_64.
8. Luque-Chang, A., Cuevas, E., Fausto, F. et al. (2018) 'Social spider optimization algorithm: modifications, applications, and perspectives', *Math. Problems in Eng.*, 2018, art. 6843923. doi: 10.1155/2018/6843923.
9. Evangeline, D., Abirami, T. (2019) 'Social spider optimization algorithm: Theory and its applications', *IJITEE*, 8(10), pp. 327–332. doi: 10.35940/ijitee.I8261.0881019.
10. Rodzin, S., Rodzina, L. (2023) 'Spider colony optimization algorithm: A bio-heuristic for global optimization problem', in *LNNS. Proc. CSOC*, 722, pp. 661–669. doi: 10.1007/978-3-031-35311-6\_63.
11. Kerschke, P., Trautmann, H. (2019) 'Automated algorithm selection on continuous black-box problems by combining exploratory landscape analysis and machine learning', *Evolutionary Comput.*, 27(1), pp. 99–127. doi: 10.1162/evco\_a\_00236.
12. Long, W., Wu, T., Liang, X., Xu, S. (2019) 'Solving high-dimensional global optimization problems using an improved sine cosine algorithm', *Expert Systems with Applications*, 123, pp. 108–126. doi: 10.1016/j.eswa.2018.11.032.
13. Dragoi, E., Dafinescu, V. (2021) 'Review of metaheuristics inspired from the animal kingdom', *Math.*, 9(18), art. 2335. doi: 10.3390/math9182335.
14. Benítez-Hidalgo, A., Nebro, A.J., García-Nieto, J., Oregi, I., Ser, J.D. (2019) 'jMetalPy: A Python framework for multi-objective optimization with metaheuristics', *Swarm and Evolutionary Computation*, 51, art. 100598. doi: 10.1016/j.swevo.2019.100598.
15. Tian, Y., Cheng, R., Zhang, X., Jin, Y. (2017) 'A MATLAB platform for evolutionary multi-objective optimization', *IEEE Computational Intelligence Magazine*, 12(4), pp. 73–87. doi: 10.1109/MCI.2017.2742868.
16. Ivanov, D.K., Dumina, D.S., Semenov, N.A. (2020) 'Determination of weight coefficients for additive fitness function of genetic algorithm', *Software & Systems*, 33(1), pp. 47–53 (in Russ.). doi: 10.15827/0236-235X.129.047-053.
17. Saha, S., Seal, D.B., Ghosh, A., Dey, K.N. (2016) 'A novel gene ranking method using Wilcoxon rank sum test and genetic algorithm', *IJBRA*, 12(3), pp. 263–279. doi: 10.1504/IJBRA.2016.078236.

### Авторы

**Родзин Сергей Иванович**<sup>1</sup>, к.т.н.,  
доцент, профессор, srodzin@sfedu.ru

### Authors

**Sergey I. Rodzin**<sup>1</sup>, Cand. of Sci. (Engineering),  
Associate Professor, Professor, srodzin@sfedu.ru

<sup>1</sup> Южный федеральный университет,  
г. Таганрог, 347922, Россия

<sup>1</sup> Southern Federal University,  
Taganrog, 347922, Russian Federation

## Представление экономической концепции кластера с помощью специализированной онтологии

Д.Л. Напольских <sup>1</sup>✉<sup>1</sup> Поволжский государственный технологический университет,  
г. Йошкар-Ола, 424000, Россия

### Ссылка для цитирования

Напольских Д.Л. Представление экономической концепции кластера с помощью специализированной онтологии // Программные продукты и системы. 2024. Т. 37. № 2. С. 155–163. doi: 10.15827/0236-235X.142.155-163

### Информация о статье

Группа специальностей ВАК: 1.2

Поступила в редакцию: 18.11.2023

После доработки: 11.12.2023

Принята к публикации: 10.01.2024

**Аннотация.** Объектом исследования является специализированная онтология как форма представления экономической концепции кластера для применения в интеллектуальных системах, предметом – иерархия понятий (классов) онтологии домена «Кластеры» и структура отношений между ними. Методологический инструментальный исследования составляют онтологический язык OWL второй версии, редактор онтологий и фреймворк для построения баз знаний Protégé, программные инструменты работы с онтологиями. В статье предлагается структурная схема, отражающая соотношение применяемых в рамках исследования инструментов. Описана последовательность этапов разработки специализированной онтологии для представления экономической концепции: определение перечня основных классов понятий онтологии, формирование таксономической иерархии предметной онтологии, разработка структуры составных понятий, входящих в онтологию, определение отношений между элементами онтологии. Представлен перечень основных классов онтологии домена «Кластеры», разработана таксономическая иерархия кластеров и экономических систем кластерного типа. В ходе исследования проведена систематизация совокупности используемых в рамках онтологии отношений. Помимо стандартных для Protégé универсальных отношений между объектом и классом, а также между подклассом и классом, были выделены 17 типов отношений, необходимых для полного представления концепции кластера. Предлагаемая онтология домена «Кластеры» является основой для интеллектуального анализа различных данных о кластерах в рамках научно-исследовательских задач и при реализации кластерной политики. Представленные в работе результаты являются основой для дальнейших исследований процессов интеграции инновационных кластеров, цифровых платформ и экосистем в контексте проблем управления региональным развитием.

**Ключевые слова:** кластеризация, цифровизация управления регионом, граф знаний, семантические технологии, специализированная онтология, тезаурус, Protégé, Web ontology language

**Благодарности.** Исследование выполнено за счет гранта РНФ № 23-78-10042 «Методология многоуровневой интеграции экономического пространства и синхронизации инновационных процессов как основа устойчивого развития российских регионов (на основе концепции инновационного гиперкластера)», <https://rscf.ru/project/23-78-10042/>

**Введение.** В настоящее время для решения научных и практических задач активно реализуется потенциал интеллектуальных систем, построенных на основе семантических технологий. Процессы цифровизации экономики и государственного управления обуславливают активное внедрение семантических технологий в различные предметные области знаний.

Кластеры и кластерная политика представляют собой недостаточно структурированную область знаний, содержащую противоречивую информацию о кластерах, обусловленную различным контекстом употребления кластерной терминологии [1]. Формирование базы знаний о кластерной политике на основе открытых источников данных о квазикластерах, зачастую не соответствующих научной концепции кластера, может привести к подмене понятий и формальным выводам о процессах кластеризации эконо-

мики [2]. Поэтому возрастает актуальность формализации научной концепции кластера на основе семантических технологий [3, 4].

Семантические сети являются наиболее перспективной формой постановки и решения сложных научно-практических задач социально-экономического развития. Возрастающая значимость больших данных и сквозной характер технологий искусственного интеллекта актуализируют задачи интеграции информации, получаемой в разных форматах и из различных баз знаний, с помощью семантических сетей [5]. Ключевым преимуществом семантических сетей по сравнению с классическими БД является их изначально полуструктурированная модель, позволяющая рассматривать объекты и явления в контексте различных областей знания. Концепция графов знаний [6] получила дальнейшее развитие в виде таких универсальных

баз знаний, как Wikidata и Metaphactory, а также ряда специализированных графов знаний, среди которых можно выделить граф знаний компании Siemens (эксплуатация промышленного оборудования), Statoil (добыча нефти и газа), Pinterest (таргетированная реклама), BioPortal (медицина и биотехнологии) и другие.

Основой (моделью) методологически грамотного построения семантических сетей являются онтологии [7]. Для формализации специализированных знаний о различных предметных областях используются онтологии, являющиеся формой целостного представления системы знаний в рамках конкретного домена (предметной области) с помощью построения иерархии понятий (классов) и структуры отношений между ними [8]. Разработано достаточное количество различных программных средств для создания и редактирования онтологий, среди которых можно выделить фреймворк Protégé [9], а также библиотеку Owlready2 для языка Python [10].

Сегодня существует большое число как универсальных, так и специализированных онтологий, представляющих различные области социально-экономической деятельности [11, 12]. Примерами универсальных онтологий верхнего уровня являются система организации знаний SKOS (Simple Knowledge Organization System), разработанная Консорциумом Всемирной паутины, общая формальная онтология GFO (General Formal Ontology), разработанная Генрихом Эрре и Барбарой Хеллер, базовая формальная онтология верхнего уровня BFO (Basic Formal Ontology), разработанная Барри Смитом.

К специализированным относятся онтологии

- библиографических описаний книг и периодических изданий (BIBO);
- транспортных потоков товаров компаний (SCTG);
- для описания различных организаций (The Organization Ontology);
- товаров (RosettaNet);
- для описания персон (FOAF);
- медицинская (Galen);
- объектов культурного наследия (CIDOC CRM).

В качестве примеров лингвистических онтологий (тезаурусов, представляющих собой иерархическую сеть понятий) можно выделить BabelNet, WordNet и Sensus, а также российские проекты PyТез и YARN.

Таким образом, создание онтологии домена «Кластеры» обусловлено задачей расширения

технологического инструментария субъектов кластерной политики, в частности, механизмов обработки на основе интеллектуальных систем метаданных, необходимых для эффективной реализации государственной кластерной политики. Разработка специализированной предметной онтологии кластера расширяет возможности решения задач кластерной политики за счет анализа больших объемов информации из различных баз знаний [13].

Процесс разработки предметной онтологии домена «Кластеры» состоит из следующих этапов:

- определение перечня основных классов понятий онтологии домена «Кластеры»;
- формирование таксономической иерархии предметной онтологии;
- разработка структуры составных понятий, входящих в онтологию;
- определение отношений между элементами онтологии.

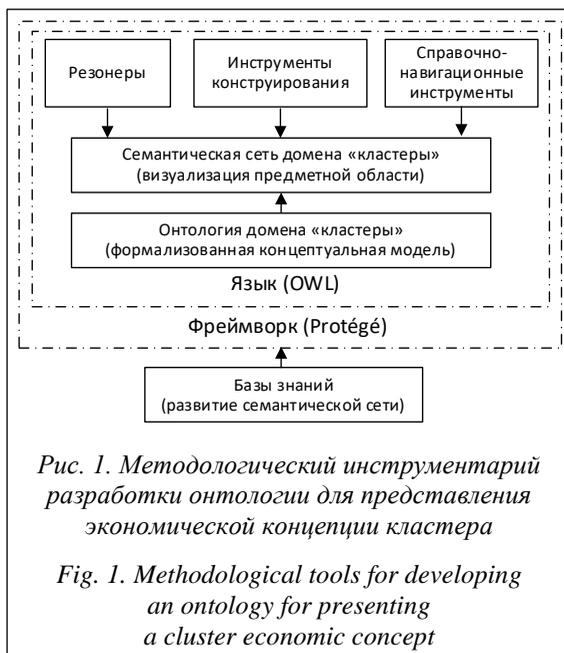
Данная последовательность может применяться для разработки специализированных онтологий с целью представления других экономических концепций и теорий.

### Методология исследования

В рамках исследования использовалась следующая совокупность семантических технологий, объединенных единым методологическим каркасом: онтологический язык OWL второй версии, редактор онтологий и фреймворк для построения баз знаний Protégé [14], программные инструменты работы с онтологиями (резонеры и др.). Инструментарий разработки онтологий для представления экономической концепции кластера показан на рисунке 1.

Сегодня основным стандартизированным языком описания онтологий для семантических сетей, поддерживаемым практически всеми инструментами моделирования онтологий, является OWL. Созданный на основе языка Java редактор онтологий Protégé существенно упрощает процесс создания онтологий, объединяя в рамках единого приложения различные инструменты по формированию иерархии классов и отношений, построению графов и т.д. Редактор онтологий Protégé позволяет использовать OWL, формировать документы на основе языков разметки HTML и XML.

Высокоуровневый язык Python, используемый в качестве языка общего назначения, также может эффективно применяться при ра-



боте с рассматриваемой онтологией. Являясь мультипарадигмальным языком, Python поддерживает решение задач как структурного, так и объектно-ориентированного программирования [15]. Специализированная библиотека Owlready2 существенно расширяет продуктивность использования Python при работе с онтологиями, особенностью Owlready2 являются управление объектами онтологии как обычными объектами Python. Также Owlready2 поддерживает использование SPARQL-запросов и механизмов рассуждений (reasoners) [16].

### Результаты исследования

В рамках онтологии домена «Кластеры» рассматриваются основные экономические понятия, относящиеся к концепции кластера, а также различные виды территориальных экономических систем. Обязательное требование к представлению предметной области в виде онтологии состоит в формировании тезауруса (гlossария), охватывающего основные понятия и термины в рамках исследуемой предметной области. При разработке онтологии домена «Кластеры» с целью конкретизации области знаний был выделен отдельный класс «Неэкономические кластеры», включающий кластеры, изучаемые естественнонаучными дисциплинами (кластеры галактик, кластеры молекул и др.).

Для исключения двусмысленной интерпретации используемых экономических понятий для каждого класса через `rdfs:isDefinedBy` было дано определение, а в разделе `Description` для

каждого класса проведено разграничение непересекающихся классов через предикат `Disjoint With`. Пример ввода определения для класса «Субъект инновационной деятельности» в редакторе `Protégé` представлен на рисунке 2.

Классы понятий являются центральным элементом онтологии домена знаний «Кластеры», большинство рассматриваемых в онтологии классов имеют несколько подклассов, являющихся в большей степени конкретизированными понятиями, чем их надкласс. Верхний уровень таксономической иерархии предметной онтологии домена «Кластеры» представлен на рисунке 3.

Структура таксономических отношений «A Kind Of» (АКО), определяющих уровни понятий (надкласс–класс–подкласс) формируется в виде таксономической иерархии. Например, класс «Кластер» включает в себя подклассы «Агропромышленный кластер», «Инновационный кластер», «Мультикластер», «Промышленный кластер» и «Туристско-рекреационный кластер». Класс «Территориально-отраслевая интегрированная система» включает все кластеры и схожие с ними территориальные экономические системы в виде классов «Протокластер», «Квазикластер», «Латентный кластер», «Территориально-производственный комплекс» и др. Конкретные промышленные кластеры являются как экземплярами надкласса «Территориально-отраслевая интегрированная система» класса «Кластер», так и экземплярами подкласса «Промышленный кластер» (рис. 4).

Для описания сложных составных понятий, входящих в предлагаемую онтологию, используется структурное отношение «Part Of» (является частью), которое также можно интерпретировать как «входит в состав» экономических систем. Например, инновационный кластер является обязательной составной частью инновационного мультикластера, выступая в роли его научно-технологического ядра, а цифровые платформы и цифровые экосистемы входят в состав инновационного гиперкластера, являющегося подтипом инновационного мультикластера.

Разработка и практическое применение предметной онтологии домена «Кластеры» включает задачу оптимизации и описания совокупности используемых в ее рамках отношений. Для специализированных онтологий наряду с общепринятыми типами отношений (таксономических, структурных и др.) вводятся отношения, свойственные рассматриваемому домену (предметной области). В онто-

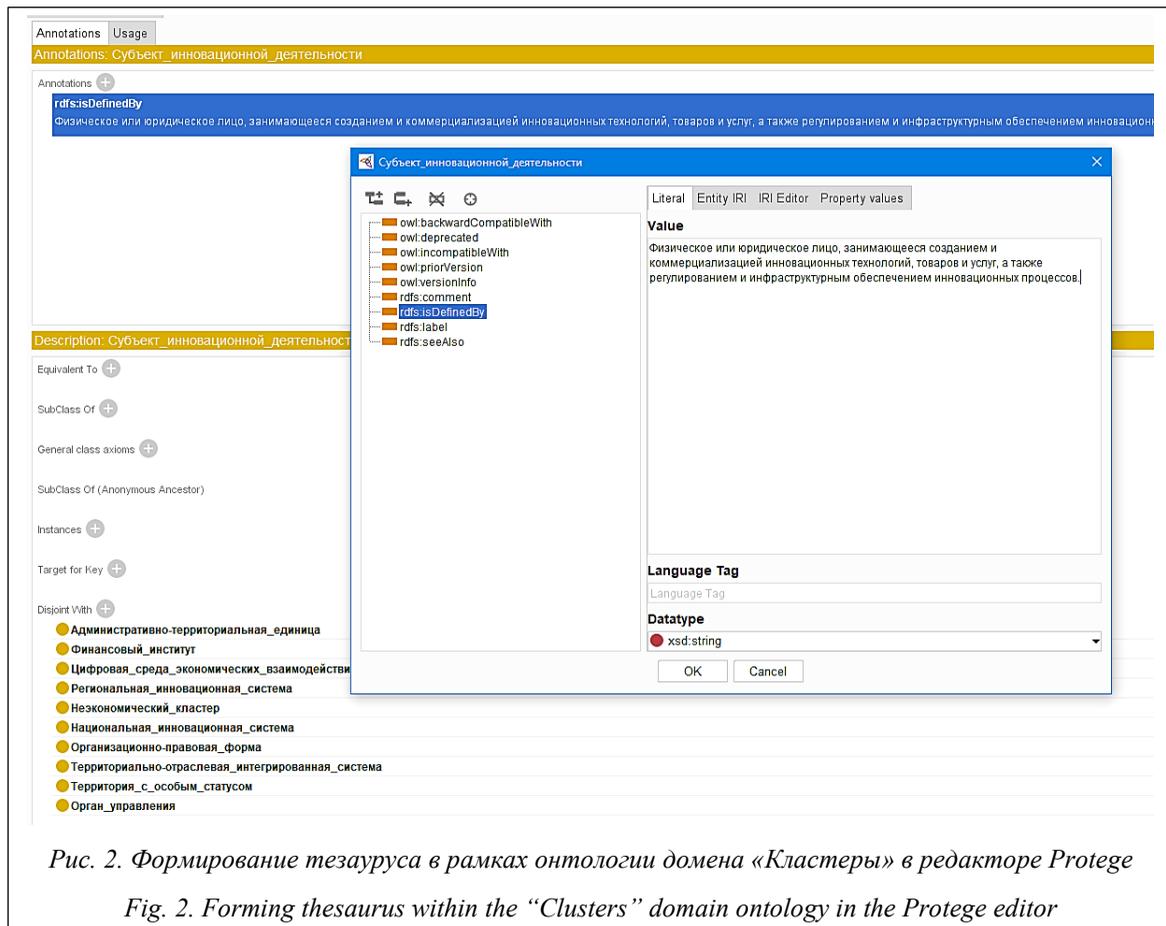


Рис. 2. Формирование тезауруса в рамках онтологии домена «Кластеры» в редакторе Protege

Fig. 2. Forming thesaurus within the “Clusters” domain ontology in the Protege editor

гии домена «Кластеры» используются следующие универсальные отношения: между объектом и классом (IsA), между подклассом и классом (АКО), части и целого (Part Of). К специфическим отношениям относятся «иметь организационно-правовую форму», «иметь основную

специализацию», «иметь сопутствующую специализацию» и другие типы отношений, представленные на рисунке 5.

### Обсуждение результатов

Обзор современных способов создания онтологий и их практического применения в рамках задач моделирования проведен в [17]. В работе сделан подробный сравнительный анализ использования ER-моделей, XML-схем, языков RDF и OWL. При этом не рассмотрены особенности применения при работе с онтологиями языка Python и специализированных библиотек, таких как Owlready2, позволяющей управлять объектами онтологии как обычными объектами Python, использовать SPARQL-запросы и механизмы рассуждений. Проблемы при работе на языке Python с русскоязычными онтологиями с помощью библиотеки Owlready2 были описаны и решены в работе [18]. Отличительной особенностью онтологии домена «Кластеры», являющейся русскоязычной, является учет специфики российской экономической мысли и нормативно-правовой базы кла-

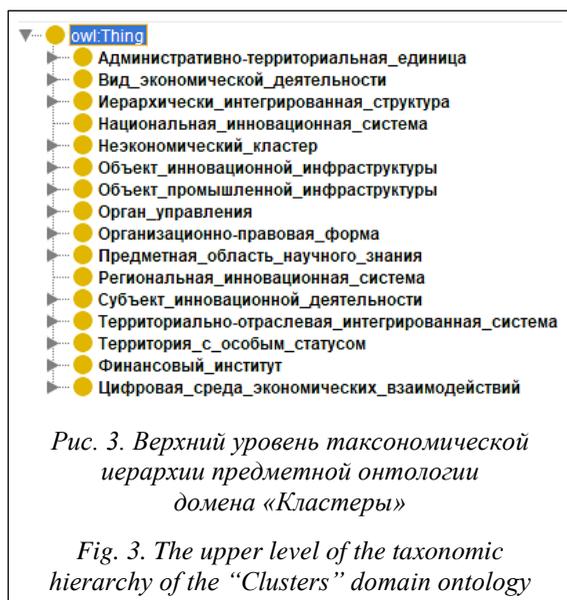


Рис. 3. Верхний уровень таксономической иерархии предметной онтологии домена «Кластеры»

Fig. 3. The upper level of the taxonomic hierarchy of the “Clusters” domain ontology

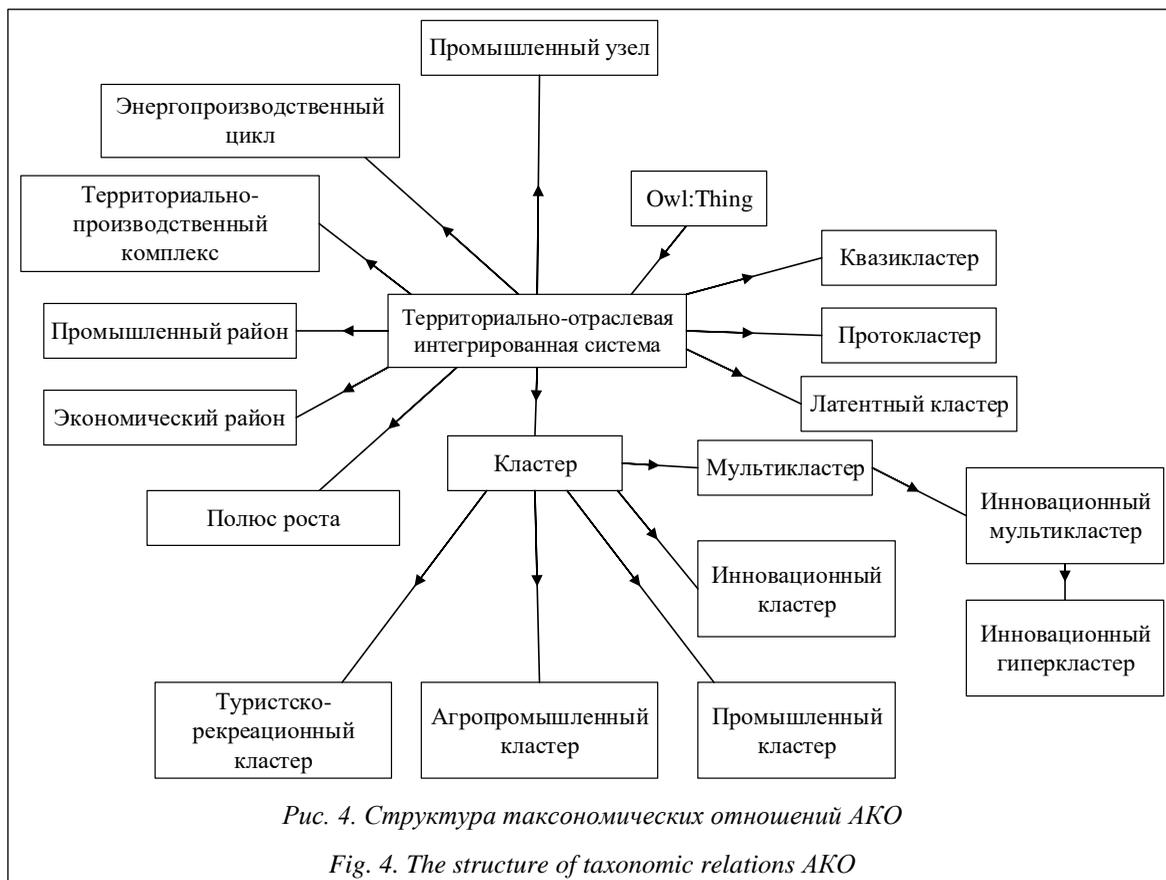


Рис. 4. Структура таксономических отношений АКО

Fig. 4. The structure of taxonomic relations AKO

стерной политики. Онтология включает как адаптированный перевод зарубежных терминов, так и свойственную советско-российской экономической географии терминологию.

Подход к интеграции подпространства предметной области знаний в общее семантическое пространство представлен в работе [19]. В частности, решаются задачи добавления терминов в тезаурус, а также построения эталонного корпуса прикладной предметной области. На основе анализа существующих методик выделены следующие этапы построения онтологий: определение цели онтологии, выделение основных понятий верхнего уровня, выделение связей между ними. Использование в рамках данного исследования указанных выше подходов позволит применять онтологию домена «Кластеры», минимизируя процесс поиска данных о кластеризации экономики без потери общеэкономических результатов и выводов, содержащихся за рамками кластерной политики в таких сферах, как инновационная политика, государственное регулирование экономики и т.д.

В работе [20] выделен ряд ограничений использования семантических сетей для анализа экономических текстов, при этом объект исследования ограничен публикациями экономиче-

ской тематики, не рассматриваются проблемы извлечения экономических данных из других типов источников (БД и т.п.). Сделан вывод о необходимости при исследовании серьезных экономических проблем проводить кросс-текстовый анализ на основе формирования крупной семантической сети, что потенциально может привести к смешению нескольких научных подходов к проблеме. Также исследователям нужно заранее определять фокус анализа, чтобы интерпретировать экономические тексты в рамках единого подхода. Ключевым требованием к онтологии домена «Кластеры», лежащей в основе применения семантических технологий при реализации кластерной политики, является соответствие современным представлениям о пространственном развитии регионов и научной концепции кластера. Особое внимание при разработке онтологии уделено разграничению форм пространственного развития, относящихся к различным теоретическим подходам к проблемам экономического развития территорий, а также соотношению данных форм в контексте именно кластерной политики (фокус исследования). Это является методологическим преимуществом применения предлагаемой онтологии практиками, не

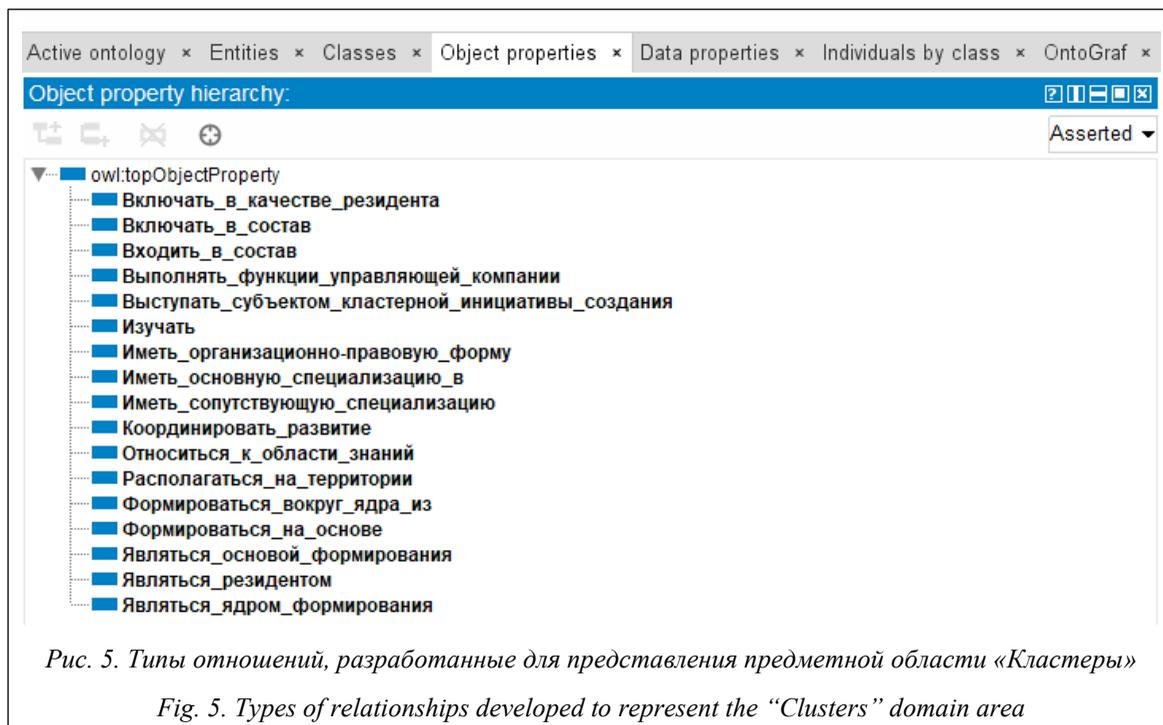


Рис. 5. Типы отношений, разработанные для представления предметной области «Кластеры»

Fig. 5. Types of relationships developed to represent the “Clusters” domain area

погруженными в научный контекст формирования современной концепции кластера.

В исследовании [21] рассмотрены научные проблемы автоматического построения семантических сетей и их применения для получения ответов на вопросы пользователей. В работе рассмотрен ограниченный набор типов вопросов, авторами планируется усовершенствование алгоритма за счет увеличения числа видов отношений между понятиями и добавления поддержки многозначных слов. В онтологии домена «Кластер» изначально учитывается многозначность общенаучного термина «кластер», а также, помимо общеэкономических отношений между понятиями, введены десять типов отношений, имеющих непосредственное отношение к кластерной политике. Введение специфических отношений, таких как «выполнять функции управляющей компании», «включать в качестве резидента», повысит качество построения семантических сетей на основе текстов научных работ и нормативных правовых актов в области кластерной политики.

В работе [22] в качестве основы построения предметной онтологии на основе БД «Татнефть» используется модель данных Ericentre 3.0, представленная в виде ER-диаграмм и набора текстовых файлов на объектно-ориентированном языке Express. При этом для описания самой онтологии используется WOL. Преимуществом данного подхода является разработка онтологии на основе Ericentre 3.0 как отраслевого

стандарта логических моделей БД в нефтегазовой отрасли. Определенная сложность практического применения данного подхода заключается в необходимости конвертации модели Ericentre в язык описания онтологий OWL. На сегодняшний день для российской кластерной политики нет единого отраслевого либо ведомственного стандарта модели данных о процессах кластеризации экономики, на основе которого возможно построение онтологии. Существующие БД о кластерах и объектах инновационной инфраструктуры, формируемые различными организациями, отражают отдельные аспекты кластерной политики. Предлагаемая онтология, изначально разрабатываемая на рекомендованном международным консорциумом W3C языке OWL, вносит определенный вклад в формирование стандарта модели данных в области кластерной политики.

### Заключение

Для представления предметной области «Кластеры» построена модель знаний, при разработке которой обобщены и структурированы различные подходы к определению и классификации кластеров, а также смежных с ними экономических систем. Разработанная онтология домена «Кластеры» является основой для структурирования различных источников данных о кластерах, извлечения необходимых пользователям данных и их контекстуализации.

Дальнейшее научное и практическое применение разработанной предметной онтологии рассматривается в рамках следующих направлений:

- размещение в открытом доступе семантически связанной концептуальной модели кластера для широкого круга лиц, участвующих в реализации кластерной политики;
- выявление неявных знаний о кластерах, актуализация и насыщение данной предметной области на основе семантических технологий;
- извлечение и обработка на основе онтологии, соответствующей современной теоретиче-

ской концепции кластера, неструктурированных данных о кластерах из различных БД, а также нормативных правовых актов и других источников;

- создание алгоритмов и интеллектуальных систем для идентификации и классификации существующих кластеров на основе предлагаемой онтологии.

Полученные научные результаты являются основой для дальнейших исследований процессов интеграции инновационных кластеров, цифровых платформ и экосистем в контексте проблем управления региональным развитием.

### Список литературы

1. Загородников К.А., Просвирина Н.В. Сущность и классификация кластеров в современной инновационной экономике // Экономика: вчера, сегодня, завтра. 2018. Т. 8. № 12А. С. 107–117.
2. Капогузов Е.А., Логинов К.К., Чупин Р.И., Харламова М.С. Прогноз экономической эффективности кластерных проектов в контексте сценариев кластерного развития региона // Terra Economicus. 2019. Т. 17. № 2. С. 40–59.
3. Sedita S., Caloffi A., Lazzeretti L. The invisible college of cluster research: A bibliometric core-periphery analysis of the literature. *Industry and Innovation*, 2018, vol. 22, no. 2, pp. 562–584. doi: 10.1080/13662716.2018.1538872.
4. Morgulis-Yakushev S., Sölvell Ö. Enhancing dynamism in clusters: A model for evaluating cluster organizations' bridge-building activities across cluster gaps. *Competitiveness Review*, 2017, vol. 27, no. 2, pp. 98–112. doi: 10.1108/CR-02-2016-0015.
5. Каленов Н.Е. Об одном подходе к формированию предметных онтологий различных областей науки // Научный сервис в сети интернет: тр. конф. 2020. С. 276–285. doi: 10.20948/abrau-2020-14.
6. Волкова И.А., Шамаева Е.Д. Экскурс в графы знаний // INJOIT. 2023. Т. 11. № 3. С. 75–83.
7. Smith B. The birth of ontology. *J. Knowl. Struct. Syst.*, 2022, vol. 3, no. 1, pp. 57–66.
8. Грибова В.В., Паршкова С.В., Федорищев Л.А. Онтологии для разработки и генерации адаптивных пользовательских интерфейсов редакторов баз знаний // Онтология проектирования. 2022. Т. 12. № 2. С. 200–217. doi: 10.18287/2223-9537-2022-12-2-200-217.
9. Horridge M., Gonçalves R., Nyulas C., Tudorache T., Musen M. WebProtégé: A cloud-based ontology editor. *Proc. WWW'19*, 2019, pp. 686–689. doi: 10.1145/3308560.3317707.
10. Иванов П.И., Мышкина И.Ю., Грудцына Л.Ю. Обзор библиотеки Owlready2 для работы с онтологиями на языке Python // Науч.-технич. вестник Поволжья. 2022. № 12. С. 139–141.
11. Волчек Д.Г. Подход к обучению онтологий на основе анализа метаданных и построения универсальных зависимостей // Экономика: вчера, сегодня, завтра. 2019. Т. 9. № 8А. С. 298–304.
12. Атаева О.М., Серебряков В.А. Онтология цифровой семантической библиотеки LibMeta // Информатика и ее применения. 2018. Т. 12. № 1. С. 2–10. doi: 10.14357/19922264180101.
13. Ketels C., Protsiv S. Cluster presence and economic performance: A new look based on European data. *Regional Studies*, 2021, vol. 55, no. 2, pp. 208–220. doi: 10.1080/00343404.2020.1792435.
14. Дедков Д.А., Бакуменко М.А. Программные продукты для создания онтологий. Система Protege // Тенденции развития Интернет и цифровой экономики: тр. III Всерос. науч.-практич. конф. 2020. С. 205.
15. Савина А.Г., Уханов Д.В., Савин Д.А. Сферы и перспективы применения языка программирования Python // Научные записки ОрелГИЭТ. 2021. Т. 38. № 2. С. 24–28.
16. Jean-Baptiste L. *Ontologies with Python: Programming Owl 2.0 Ontologies with Python and Owlready2*. CA, Apress Berkeley, 2021, 344 p. doi: 10.1007/978-1-4842-6552-9.
17. Антонов А.А., Быков А.Н., Чернышев С.А. Обзор существующих способов формирования онтологии предметной области при моделировании // Междунар. журнал информационных технологий и энергоэффективности. 2021. Т. 6. № 4. С. 12–17.
18. Щукарев И.А. Особенности работы с русскоязычными онтологиями с помощью библиотеки Owlready2 на языке Python // Программные продукты и системы. 2023. Т. 36. № 2. С. 223–227. doi: 10.15827/0236-235X.142.223-227.
19. Атаева О.М., Серебряков В.А., Тучкова Н.П. Интеграция подпространства предметной области в семантическое пространство «математика» // Программные продукты и системы. 2023. Т. 36. № 1. С. 83–96. doi: 10.15827/0236-235X.141.083-096.
20. Пашков С.Г. Семантический сетевой подход: возможности и ограничения (пример образа инфляции в СМИ) // Социологический журнал. 2020. Т. 26. № 2. С. 8–30. doi: 10.19181/socjour.2020.26.2.7262.
21. Потараев В.В., Серебряная Л.В. Автоматическое построение семантической сети для получения ответов на вопросы // Докл. БГУИР. 2020. Т. 18. № 4. С. 44–52. doi: 10.35596/1729-7648-2020-18-4-44-52.

22. Гусенков А.М., Бухараев Н.Р., Биряльцев Е.В. Построение онтологии предметной области на основе логической модели данных // Russian Digital Libraries J. 2020. Т. 23. № 3. С. 390–417. doi: 10.26907/1562-5419-2020-23-3-390-417.

Software &amp; Systems

doi: 10.15827/0236-235X.142.155-163

2024, 37(2), pp. 155–163

## Developing a specialized ontology to represent the economic concept of a cluster

Dmitry L. Napolskikh <sup>1</sup>✉<sup>1</sup> Volga State University of Technology, Yoshkar-Ola, 424000, Russian Federation

### For citation

Napolskikh, D.L. (2024) 'Developing a specialized ontology to represent the economic concept of a cluster', *Software & Systems*, 37(2), pp. 155–163 (in Russ.). doi: 10.15827/0236-235X.142.155-163

### Article info

Received: 18.11.2023

After revision: 11.12.2023

Accepted: 10.01.2024

**Abstract.** The object of the research is a specialized ontology as a form representing the economic concept of a cluster for using in intelligent systems. The subject of the research is the hierarchy of concepts (classes) of the “Clusters” domain ontology and the structure of relations between them. The methodological tools of the research are the second-version OWL ontological language, an ontology editor and the Protégé framework for building knowledge bases, software tools for working with ontologies. The paper proposes a block diagram describing the ratio of the tools used in the study. It also describes a sequence of development stages for a specialized ontology to present an economic concept: a definition of the list of ontology concept main classes; forming a taxonomic hierarchy of a subject ontology; developing the structure of composite concepts included in the ontology; a definition of relations between ontology elements. The paper presents a list of the main ontology classes of the “Clusters” domain, the developed taxonomic hierarchy of clusters and cluster-type economic systems. The study involved the arrangement of the relations used in terms of the ontology. In addition to the Protégé standard universal relations between an object and a class (IsA), as well as between a subclass and a class (AKO), the paper has identified 17 types of relations necessary for a complete representation of the cluster concept. The proposed ontology of the “Clusters” domain is the basis for the intellectual analysis of various data on clusters in terms of research tasks and cluster policy. The results presented in the paper are the basis for further studies of the integration processes for innovation clusters, digital platforms and ecosystems in the context of regional development management problems.

**Keywords:** clustering, region management digitalization, knowledge graph, semantic technologies, specialized ontology, thesaurus, Protégé, Web ontology language

**Acknowledgements.** The research was supported by the grant of the Russian Science Foundation no. 23-78-10042 “Methodology of economic space multilevel integration and innovation process synchronization as a basis for sustainable development of Russian regions (based on the concept of an innovative hypercluster)”, <https://rscf.ru/project/23-78-10042/>

### References

1. Zagorodnikov, K.A., Prosvirina, N.V. (2018) 'Essence and classification of clusters in modern innovation economy', *Economics: Yesterday, Today, Tomorrow*, 8(12A), pp. 107–117 (in Russ.).
2. Kapoguzov, E.A., Loginov, K.K., Chupin, R.I., Kharlamova, M.S. (2019) 'The forecast of cluster economic efficiency in the context of regional cluster development scenarios', *Terra Economicus*, 17(2), pp. 40–59 (in Russ.).
3. Sedita, S., Caloffi, A., Lazzeretti, L. (2018) 'The invisible college of cluster research: A bibliometric core-periphery analysis of the literature', *Industry and Innovatio*, 22(2), pp. 562–584. doi: 10.1080/13662716.2018.1538872.
4. Morgulis-Yakushev, S., Sölvell, Ö. (2017) 'Enhancing dynamism in clusters: A model for evaluating cluster organizations' bridge-building activities across cluster gaps', *Competitiveness Review*, 27(2), pp. 98–112. doi: 10.1108/CR-02-2016-0015.
5. Kalenov, N.E. (2020) 'About one approach to the formation of subject ontologies for science various fields', *Proc. Conf. Sci. Service on the Internet*, pp. 276–285 (in Russ.). doi: 10.20948/abrau-2020-14.
6. Volkova, I.A., Shamaeva, E.D. (2023) 'Excursus in knowledge graphs', *INJOIT*, 11(3), pp. 75–83 (in Russ.).
7. Smith, B. (2022) 'The birth of ontology', *J. Knowl. Struct. Syst.*, 3(1), pp. 57–66.
8. Gribova, V.V., Parshkova, S.V., Fedorishchev, L.A. (2022) 'Ontologies for development and generation adaptive user interfaces of knowledge base editors', *Ontology of Designing*, 12(2), pp. 200–217 (in Russ.). doi: 10.18287/2223-9537-2022-12-2-200-217.
9. Horridge, M., Gonçalves, R., Nyulas, C., Tudorache, T., Musen, M. (2019) 'WebProtégé: A cloud-based ontology editor', *Proc. WWW'19*, pp. 686–689. doi: 10.1145/3308560.3317707.

10. Ivanov, P.I., Myshkina, I.Y., Grudcyna, L.Y. (2022) 'Overview of the Owlready2 library for working with python ontologies', *Sci. and Tech. Volga Region Bull.*, (12), pp. 139–141 (in Russ.).
11. Volchek, D.G. (2019) 'An approach to training ontologies based on the analysis of metadata and the construction of universal dependencies', *Economics: Yesterday, Today and Tomorrow*, 9(8A), pp. 298–304 (in Russ.).
12. Ataeva, O.M., Serebryakov, V.A. (2018) 'Ontology of the digital semantic library LibMeta', *Informatics and Applications*, 12(1), pp. 2–10 (in Russ.). doi: 10.14357/19922264180101.
13. Ketels, C., Protsiv, S. (2021) 'Cluster presence and economic performance: A new look based on European data', *Regional Studies*, 55(2), pp. 208–220. doi: 10.1080/00343404.2020.1792435.
14. Dedkov, D.A., Bakumenko, M.A. (2020) 'Software products for creating ontologies. The Protege system', *Proc. III All-Russ. Conf. Trends in the Development of the Internet and Digital Economy*, pp. 205 (in Russ.).
15. Savina, A.G., Sukhanov, D.V., Savin, D.A. (2021) 'Areas and prospects of application the python programming language', *Sci. J. of OrelSIET*, 38(2), pp. 24–28 (in Russ.).
16. Jean-Baptiste, L. (2021) 'Ontologies with Python: Programming Owl 2.0 Ontologies with Python and Owlready2', CA: Apress Berkeley, 344 p. doi: 10.1007/978-1-4842-6552-9.
17. Antonov, A.A., Bykov, A.N., Chernyshev, S.A. (2021) 'Review of existing methods of formation of domain ontology in modeling', *Int. J. of Information Technology and Energy Efficiency*, 6(4), pp. 12–17 (in Russ.).
18. Shchukarev, I.A. (2023) 'Features of working with Russian-language ontologies using the Owlready2 library in Python', *Software & Systems*, 36(2), pp. 223–227 (in Russ.). doi: 10.15827/0236-235X.142.223-227.
19. Ataeva, O.M., Serebryakov, V.A., Tuchkova, N.P. (2023) 'Integrating the subject area subspace into the "mathematics" semantic space', *Software & Systems*, 36(1), pp. 83–96 (in Russ.). doi: 10.15827/0236-235X.141.083-096.
20. Pashkov, S.G. (2020) 'The semantic network approach: opportunities and restrictions (example of inflation image in the media)', *Sociological J.*, 26(2), pp. 8–30 (in Russ.). doi: 10.19181/socjour.2020.26.2.7262.
21. Potaraev, V.V., Serebryanaya, L.V. (2020) 'Automatic generation of semantic network for question answering', *Doklady BGUIR*, 18(4), pp. 44–52 (in Russ.). doi: 10.35596/1729-7648-2020-18-4-44-52.
22. Gusenkov, A.M., Bukharaev, N.R., Biryaltsev, E.V. (2020) 'Building subject domain ontology on the base of a logical data mod', *Russ. Digital Libraries J.*, 23(3), pp. 390–417 (in Russ.). doi: 10.26907/1562-5419-2020-23-3-390-417.

**Авторы**

**Напольских Дмитрий Леонидович**<sup>1</sup>,  
к.э.н., доцент, NapolskihDL@yandex.ru

**Authors**

**Dmitry L. Napolskikh**<sup>1</sup>, Cand. of Sci. (Economics),  
Associate Professor, NapolskihDL@yandex.ru

<sup>1</sup> Поволжский государственный технологический университет, г. Йошкар-Ола, 424000, Россия

<sup>1</sup> Volga State University of Technology,  
Yoshkar-Ola, 424000, Russian Federation

## Применение глубокого обучения в интерфейсах мозг–компьютер для распознавания движений

Д.В. Павленко<sup>1</sup>✉, Ш.Э. Татарис<sup>1</sup>, В.В. Овчаренко<sup>1</sup>

<sup>1</sup> Крымский федеральный университет имени В.И. Вернадского, г. Симферополь, 295007, Россия

### Ссылка для цитирования

Павленко Д.В., Татарис Ш.Э., Овчаренко В.В. Применение глубокого обучения в интерфейсах мозг–компьютер для распознавания движений // Программные продукты и системы. 2024. Т. 37. № 2. С. 164–169. doi: 10.15827/0236-235X.142.164-169

### Информация о статье

Группа специальностей ВАК: 1.2.1

Поступила в редакцию: 05.11.2023

После доработки: 11.12.2023

Принята к публикации: 10.01.2024

**Аннотация.** Перспективными направлениями применения машинного обучения являются анализ электроэнцефалограммы (ЭЭГ) и разработка нейроинтерфейсов, которые служат для помощи людям с ограниченными возможностями, а также в реабилитационных процедурах. Использование искусственных нейронных сетей позволяет значительно упростить разработку подобных устройств. Нейросети дают возможность автоматически выделять паттерны ЭЭГ, связанные с определенными состояниями, и проводить сложную обработку сигнала в режиме реального времени. Важным аспектом при создании нейроинтерфейсов является разработка программного обеспечения, способного детектировать и классифицировать движения пользователя. В рамках данного исследования решались задачи классификации паттернов сенсомоторных ритмов ЭЭГ, которые связаны с воображаемыми и реальными произвольными движениями верхних конечностей. Была разработана модель по аналогии с архитектурой модели EEGNet. Данная модель предназначена для анализа ЭЭГ в режиме реального времени с отечественного энцефалографа NVX52. Для этой цели были собраны два датасета с записями здоровых людей с использованием неполной международной схемы наложения электродов 10–10, которая включала в себя 32 канала. В ходе исследования удалось достичь точности классификации ряда воображаемых движений до 80 %, распознавания признаков реальных движений до 78 % и превысить показатели распознавания по сравнению с изначальной моделью более чем на 10 %. В перспективе планируется использование данной модели в коррекционных тренингах с применением комплекса, состоящего из неинвазивного интерфейса мозг–компьютер и экзоскелетов кистей рук. Такие комплексы используются в рамках мероприятий, направленных на реабилитацию детей, страдающих детским церебральным параличом.

**Ключевые слова:** интерфейс мозг–компьютер, нейроинтерфейс, ЭЭГ, глубокое обучение, сверточная нейронная сеть, сенсомоторный ритм, детский церебральный паралич

**Благодарности.** Исследование выполнено за счет гранта РФФИ и Республики Крым № 22-15-20035, <https://rscf.ru/project/22-15-20035/>

**Введение.** Взаимодействие человека с электронными устройствами на основе регистрации и анализа электрической активности мозга обеспечивают *интерфейсы мозг–компьютер* (ИМК) [1]. Традиционно ИМК используются в медицинских целях, например, для управления протезами искусственных конечностей или для коррекции различных нарушений нервной системы и улучшения психофизиологического состояния здоровых пользователей [2]. Чаще всего применяются неинвазивные технологии, основанные на регистрации *электроэнцефалограммы* (ЭЭГ). Работа ИМК включает пять основных этапов: сбор данных, обработка сигнала, выделение значимой информации, интерпретация признаков и принятие решения, обратная связь и вывод решения. Хотя эти этапы во многом схожи у многих ИМК, каждая система опирается на разные процессы предоб-

работки сигналов [2], выделения признаков и особые методы классификации [3].

Современные алгоритмы машинного обучения превосходят стандартные методы анализа ЭЭГ при выполнении сложных задач. Например, на основе алгоритмов машинного обучения разрабатываются нейроинтерфейсы для управления различной техникой (инвалидными колясками, автоматизированными больничными койками и другими роботизированными системами) [4]. В работе [5] для классификации эмоций по паттерну ЭЭГ применяется метод, основанный на *сверточных нейронных сетях* (СНС), достигающий 79 % точности распознавания нескольких классов. Подобные системы используются в нейромаркетинге. Также авторы рассматривают алгоритм удаления помех из ЭЭГ при помощи рекуррентных нейронных сетей. Сверточные сети применяются и для других за-

дач, связанных с анализом ЭЭГ. В исследовании [6] для выявления признаков эпилептиформной активности в записях ЭЭГ, предварительно обработанных методом вейвлет-анализа, использована модель архитектуры ResNet18. Модели на базе классических алгоритмов показывают меньшую эффективность по сравнению с СНС.

В настоящее время СНС широко распространены благодаря их успешному применению в сложных задачах классификации и генерации изображений [7], а также для классификации паттернов ЭЭГ, связанных с движениями. Представленные в работах [7, 8] модели СНС имели разные наборы данных и демонстрировали точность от 60 до 94 %.

Одним из примеров использования архитектуры СНС является модель, разработанная в ходе проекта “EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces” [7]. EEGNet применяется для ИМК на основе регистрации и анализа ЭЭГ и некоторых других задач. Основа подхода данной архитектуры заключается в последовательной свертке фрагмента записи сначала во временном диапазоне, а затем в пространственном, то есть по разным осям массива данных [9, 10]. Это позволяет выделить признаки, содержащиеся как в отдельных каналах, так и в их пересечениях, выполняя как временную, так и пространственную фильтрацию. При этом свертка происходит одномерно, хотя для нее применяется слой двухмерной свертки Conv2d [7, 11]. Двухэтапная сверточная последовательность основана на алгоритме общего пространственного шаблона банка фильтров “Filter Bank Common Spatial Pattern” (FBCSP) [12, 13]. Указанная модель и другие подобные модели [14, 15] демонстрирует высокие показатели классификации движений с набором данных BCI Challenge, размещенном на сайте Kaggle (<https://www.kaggle.com/c/inria-bci-challenge/data>). Данные собраны по 56 и более каналам, однако при работе с меньшим числом каналов эффективность моделей резко падает. В настоящее время наиболее распространенным и доступным средством регистрации ЭЭГ являются электроэнцефалографы с 32 каналами. Отмечено, что уменьшение количества каналов ЭЭГ перспективно для создания компактных и удобных в использовании систем современных нейроинтерфейсов [16].

Таким образом, цель представленного проекта заключается в создании на основе EEGNet новой модели для классификации ряда вообра-

жаемых и реальных движений рук с применением электроэнцефалографа с 32 каналами, а также в ее обучении и тестировании на собственном наборе данных. Указанная модель может быть использована при разработке ПО комплекса, включающего ИМК и экзоскелет кистей рук. Такой комплекс управляется в зависимости от параметров ЭЭГ при мысленном представлении движения и используется для коррекции развития детей с *детским церебральным параличом* (ДЦП).

### Метод исследования

Для доработки, обучения и тестирования новой версии модели было собрано несколько датасетов ЭЭГ с фрагментами записей, соответствующих воображаемому и реальному сжатию кистей левой и правой рук.

Поскольку предполагалось, что распознавание реальных движений проще, чем воображаемых, первым этапом исследований стало применение модели глубокого обучения на основе EEGNet для классификации реальных движений. Для этой задачи был собран набор данных из 36 записей ЭЭГ, зарегистрированных у двух испытуемых и содержащих эпохи, соответствующие набору реальных движений рук. ЭЭГ регистрировалась монополярно в лобных (Fp1, Fp2, Fz, F1, F2, F3, F4), центральных (Cz, C1, C2, C3, C4), теменных (Pz, P1, P2, P3, P4) и затылочных (Oz, O1, O2) отведениях, расположенных по неполной международной схеме 10–10 с частотой дискретизации 500 Гц. В качестве референтного использован объединенный ушной электрод, для регистрации ЭЭГ – отечественный энцефалограф NVX 52, программа NeoRec и разработанное на языке Python приложение для отслеживания движений с применением библиотеки MediaPipe. Данное приложение обеспечивает захват видеоизображений с камеры во время регистрации ЭЭГ и позволяет проводить эксперимент более быстро и комфортно для испытуемого, давая возможность совершать движения в произвольное время без предоставления стимулов. Каждая запись длилась произвольное количество времени (в пределах 15 мин.), на протяжении которого испытуемый совершал от 100 до 200 циклов движений (сжатия, разжимания, расслабления кисти левой руки), после чего делали перерыв.

При проведении предобработки проводилась фильтрация при помощи FFT фильтра, выделялись сигналы в частотном диапазоне 1–40 Гц.

Для устранения выбросов удалялись сэмплы с амплитудой сигнала, превышающей 100 мкВ. Проводилась нормализация сэмплов средствами библиотеки scikit-learn.

В задаче распознавания реальных движений модель EEGnet, обученная на данных 32 каналов, продемонстрировала максимальную точность 77 % на валидационной выборке и 67 % на тестовой при бинарной классификации наличия движения и отдыха (рис. 1).

Вторым этапом исследований являлось применение использованной ранее модели для классификации воображаемых движений. С этой целью был собран датасет из 1 382 записей 150 здоровых испытуемых в возрасте от 18 до 30 лет. Запись ЭЭГ осуществлялась с помощью приложения, разработанного на базе программной платформы Open Vibe. Приложение обеспечивало выполнение сценария эксперимента, при котором испытуемым на экране компьютера предъявлялись команды «сжатие левой руки»,

«разжатие левой руки», «сжатие правой руки», «разжатие правой руки», «отдых». В ответ на эти команды надо было кинестетически вообразить соответствующее движение руки или состояние расслабления. С участием каждого испытуемого проводилось по 10 сессий регистрации ЭЭГ, каждая из которых продолжалась 15 минут. В результате был получен датасет, содержащий 120 341 семпл, соответствующий воображаемым движениям или расслаблению. Предобработка данных осуществлялась указанным выше методом.

В задаче классификации с тремя классами состояний – сжатие левой руки, правой руки и расслабление – удалось достичь точности в 68 % на валидационном датасете и 65 % на тестовом (рис. 2).

Для повышения точности классификации и оптимизации модели, а также ее обучения были изменены параметры всех слоев свертки. В первом слое размер фильтров был увеличен

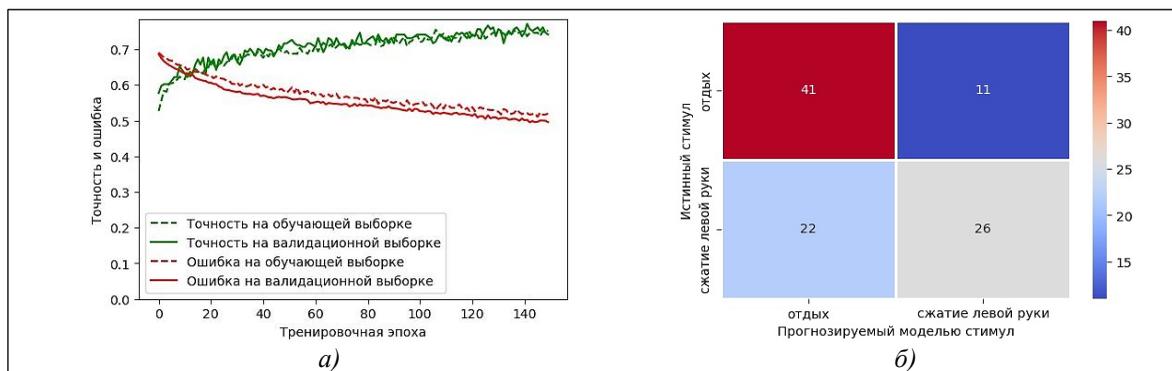


Рис. 1. Обучение модели EEGnet в задаче распознавания реальных движений: а) график обучения модели; б) матрица ошибок на тестовых данных

Fig. 1. Training EEGnet model in the task of recognizing real movements: а) model training schedule; б) test data error matrix

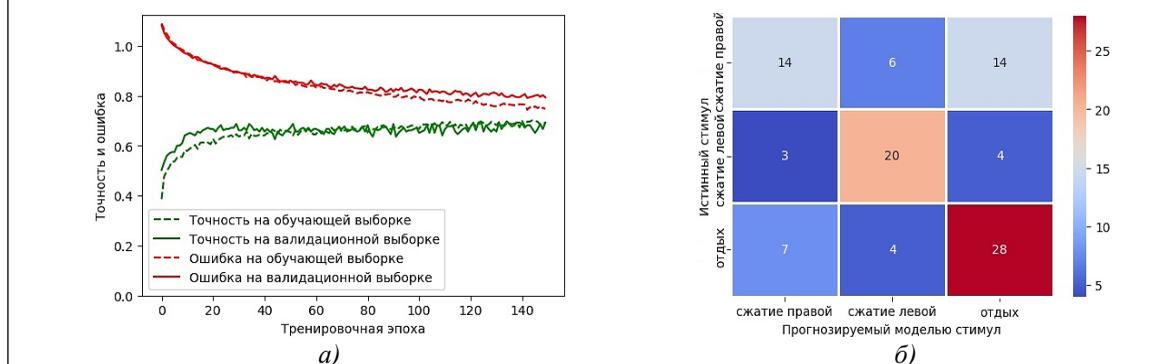


Рис. 2. Обучение модели EEGnet в задаче распознавания воображаемых движений: а) график обучения модели; б) матрица ошибок на тестовых данных

Fig. 2. Training EEGnet model in the task of recognizing imaginary movements: а) model training schedule; б) test data error matrix

до 64, во втором – до 128. Размер ядра в слоях, выполняющих свертку по оси времени, был увеличен до 64, что при частоте дискретизации сигнала 128 Гц соответствует временному отрезку около 0.5 секунды. Значение dropout было уменьшено до 0.1, слои AveragePooling2D заменены на MaxPooling2D.

### Результаты исследования

Новая модель в задаче бинарной классификации реальных движений и расслабления продемонстрировала точность 85 % на валидационном наборе данных и 73 % на тестовом (рис. 3).

В задаче классификации воображаемых движений левой руки, правой руки и расслабления удалось получить точность 80 % на валидационном датасете и 75 % на тестовом (рис. 4).

Таким образом, была проанализирована существующая методика разработки ИМК на основе глубокого обучения с применением

сверточных сетей. Собраны два набора данных, зарегистрированных ЭЭГ с 32 каналами и содержащих паттерны ЭЭГ, соответствующие реальным и воображаемым движениям. Протестирована модель с архитектурой EEGnet, и разработана ее улучшенная версия, показавшая более высокую эффективность в задачах распознавания паттернов ЭЭГ.

### Заключение

В настоящее время в лаборатории «Фаблаб» КФУ имени В.И. Вернадского разрабатывается программа, использующая описанную модель для обработки данных ЭЭГ и классификации движений в режиме реального времени. Это позволит создать комплекс, включающий нейроинтерфейс для проведения коррекционных тренировок на основе биологической обратной связи для детей с ДЦП. Экзоскелеты кистей рук, также входящие в состав комплекса, совершают движения при совпадении активности

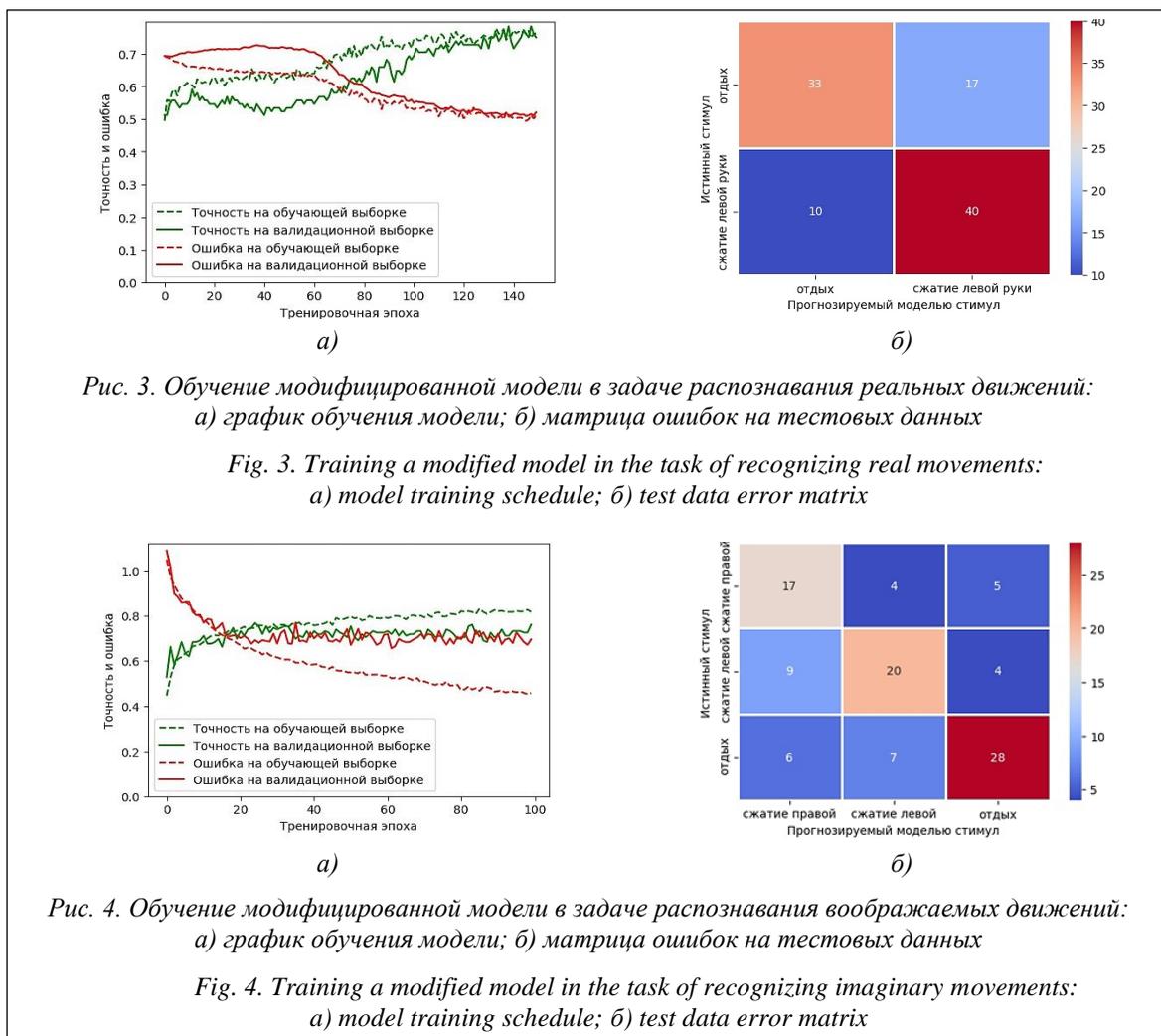


Рис. 3. Обучение модифицированной модели в задаче распознавания реальных движений: а) график обучения модели; б) матрица ошибок на тестовых данных

Fig. 3. Training a modified model in the task of recognizing real movements: а) model training schedule; б) test data error matrix

Рис. 4. Обучение модифицированной модели в задаче распознавания воображаемых движений: а) график обучения модели; б) матрица ошибок на тестовых данных

Fig. 4. Training a modified model in the task of recognizing imaginary movements: а) model training schedule; б) test data error matrix

головного мозга ребенка с параметрами паттернов ЭЭГ здоровых испытуемых. В дальнейшем планируется провести серию клинических

исследований для оценки эффективности данного комплекса с соответствующим набором программ.

#### Список литературы

1. Camargo-Vargas D., Callejas-Cuervo M., Mazzoleni S. Brain-computer interfaces systems for upper and lower limb rehabilitation: A systematic review. *Sensors*, 2021, vol. 21, no. 13, art. 4312. doi: 10.3390/s21134312.
2. Saproo S., Faller J., Shih V., Sajda P., Waytowich N.R. et al. Cortically coupled computing: A new paradigm for synergistic human-machine interaction. *Computer*, 2016, vol. 49, no. 9, pp. 60–68. doi: 10.1109/MC.2016.294.
3. Fu J., Chen S., Jia J. Sensorimotor rhythm-based brain-computer interfaces for motor tasks used in hand upper extremity rehabilitation after stroke: A systematic review. *Brain Sci.*, 2022, vol. 13, no. 1, art. 56. doi: 10.3390/brainsci13010056.
4. Бодин О.Н., Солодимова Г.А., Спиркин А.Н. Нейроинтерфейс для управления роботизированными устройствами // Измерение. Мониторинг. Управление. Контроль. 2019. Т. 30. № 4. С. 70–76. doi: 10.21685/2307-5538-2019-4-8.
5. Шанин И.А., Ступников С.А. Методы анализа данных электроэнцефалографии с применением сверточных и рекуррентных нейронных сетей // Системы и средства информатики. 2021. Т. 31. № 2. С. 36–46. doi: 10.14357/08696527210204.
6. Назариков С.И. Математическая модель для выявления эпилептических приступов на записи ЭЭГ // Изв. вузов. ПИД. 2023. Т. 31. № 5. С. 628–642. doi: 10.18500/0869-6632-003065.
7. Lawhern V.J., Solon A.J., Waytowich N.R., Gordon S.M. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. *JNE*, 2018, vol. 15, no. 5, art. 056013. doi: 10.1088/1741-2552/aae8c.
8. Waytowich N., Lawhern V.J., Garcia J.O., Faller J., Vettel P.S., Vettel J.M. Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. *JNE*, 2018, vol. 15, no. 6, art. 066031. doi: 10.1088/1741-2552/aae5d8.
9. Schirrmester R.T., Springenberg J.T., Fiederer L.D.J., Glasstetter M., Eggenberger K. et al. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 2017, vol. 38, no. 11, pp. 5391–5420. doi: 10.1002/hbm.23730.
10. Gordon S.M., Jaswa M., Solon A.J., Lawhern V.J. Real world BCI: Cross-domain learning and practical applications. *Proc. BCIForReal*, 2017, pp. 25–28. doi: 10.1145/3038439.3038444.
11. Yin Z., Zhang J. Cross-session classification of mental workload levels using EEG and an adaptive deep learning model. *Biomedical Signal Processing and Control*, 2017, vol. 33, pp. 30–47. doi: 10.1016/j.bspc.2016.11.013.
12. Park S.H., Lee D., Lee S.G. Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2018, vol. 26, no. 2, pp. 498–505. doi: 10.1109/TNSRE.2017.2757519.
13. Arpaia P., Esposito A., Natalizio A., Parvis M. How to successfully classify EEG in motor imagery BCI: A metrological analysis of the state of the art. *JNE*, 2022, vol. 19, no. 3. doi: 10.1088/1741-2552/ac74e0.
14. Khademi Z., Ebrahimi F., Kordy H.M. A review of critical challenges in MI-BCI: From conventional to deep learning methods. *J. of Neuroscience Methods*, 2023, vol. 383, art. 109736. doi: 10.1016/j.jneumeth.2022.109736.
15. Tibrewal N., Leeuwis N., Alimardani M. Classification of motor imagery EEG using deep learning increases performance in inefficient BCI users. *PLoS One*, 2022, vol. 17, no. 7, art. e0268880 doi: 10.1371/journal.pone.0268880.
16. Ямаев М.И., Шипицин С.П., Филатова Е.С. Нейронные сети для классификации ЭЭГ: от перцептрона до сверточной LSTM-сети // Междунар. конф. по мягким вычислениям и измерениям: матер. 2018. Т. 1. С. 765–768.

### Applying deep learning in brain-computer interfaces for motion recognition

Dmitry V. Pavlenko <sup>1</sup>✉, Shevhiye E. Tataris <sup>1</sup>, Vladimir V. Ovcharenko <sup>1</sup>

<sup>1</sup> V.I. Vernadsky Crimean Federal University, Simferopol, 295007, Russian Federation

#### For citation

Pavlenko, D.V., Tataris, Sh.E., Ovcharenko, V.V. (2024) 'Applying deep learning in brain-computer interfaces for motion recognition', *Software & Systems*, 37(2), pp. 164–169 (in Russ.). doi: 10.15827/0236-235X.142.164-169

#### Article info

Received: 05.11.2023

After revision: 11.12.2023

Accepted: 10.01.2024

**Abstract.** The promising areas of machine learning application are electroencephalogram (EEG) analysis and development of neural interfaces that serve to help people with disabilities, as well as in rehabilitation procedures. Artificial neural networks make it possible to significantly simplify the development of such devices. Neural networks enable automatic identification of EEG patterns associated with certain states and complex signal processing in real time. One of the important aspects in creating neural interfaces is developing software capable of detecting and classifying user movements. Within the framework of this study, the authors solved the tasks of classifying patterns of EEG sensorimotor rhythms, which are associated with imaginary and real voluntary movements of the upper extremities. The developed model is sim-

ilar to the EEGNet model architecture. This model is designed for real-time EEG analysis using a domestic NVX52 encephalograph. For this purpose, the authors of the paper have collected two datasets with records of healthy people using an incomplete international 10-10 electrode overlay scheme, which included 32 channels. During the study, the authors achieved the accuracy of classifying a number of imaginary movements up to 80 %, the accuracy of recognizing signs of real movements up to 78 % and exceeded recognition rates by more than 10 % compared to the original model. In the future, it is planned to use this model in correctional trainings using a complex consisting of a non-invasive brain-computer interface and hand exoskeletons. Such complexes are used as a part of rehabilitation measures aimed at the rehabilitation of children suffering from cerebral palsy.

**Keywords:** brain-computer interface, neurointerface, EEG, deep learning, convolutional neural network, sensory-motor rhythm

**Acknowledgements.** The study was supported by the Russian Science Foundation and the Republic of Crimea grant no. 22-15-20035, <https://rscf.ru/en/project/22-15-20035/>

### References

1. Camargo-Vargas, D., Callejas-Cuervo, M., Mazzoleni, S. (2021) 'Brain-computer interfaces systems for upper and lower limb rehabilitation: A systematic review', *Sensors*, 21(13), art. 4312. doi: 10.3390/s21134312.
2. Saproo, S., Faller, J., Shih, V., Sajda, P., Waytowich, N.R., et al. (2016) 'Cortically coupled computing: A new paradigm for synergistic human-machine interaction', *Computer*, 49(9), pp. 60–68. doi: 10.1109/MC.2016.294.
3. Fu, J., Chen, S., Jia, J. (2022) 'Sensorimotor rhythm-based brain-computer interfaces for motor tasks used in hand upper extremity rehabilitation after stroke: A systematic review', *Brain Sci.*, 13(1), art. 56. doi: 10.3390/brainsci13010056.
4. Bodin, O.N., Solodimova, G.A., Spirkin, A.N. (2019) 'Neurointerface for controlling robotic devices', *Measurement. Monitoring. Management. Control*, 30(4), pp. 70–76 (in Russ.). doi: 10.21685/2307-5538-2019-4-8.
5. Shanin, I.A., Stupnikov, S.A. (2021) 'Electroencephalography data analysis with convolutional and recurrent neural networks', *Systems and Means of Informatics*, 31(2), pp. 36–46 (in Russ.). doi: 10.14357/08696527210204.
6. Nazarikov, S.I., (2023) 'Mathematical model for identifying epileptic seizures on EEG recordings', *Izv. VUZ. Applied Nonlinear Dynamics*, 31(5), pp. 628–642 (in Russ.). doi: 10.18500/0869-6632-003065.
7. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M. (2018) 'EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces', *JNE*, 15(5), art. 056013. doi: 10.1088/1741-2552/aae8c.
8. Waytowich, N., Lawhern, V.J., Garcia, J.O., Faller, J., Vettel, P.S., Vettel, J.M. (2018) 'Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials', *JNE*, 15(6), art. 066031. doi: 10.1088/1741-2552/aae5d8.
9. Schirrneister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggenberger, K. et al. (2017) 'Deep learning with convolutional neural networks for EEG decoding and visualization', *Human Brain Mapping*, 38(11), pp. 5391–5420. doi: 10.1002/hbm.23730.
10. Gordon, S.M., Jaswa, M., Solon, A.J., Lawhern, V.J. (2017) 'Real world BCI: Cross-domain learning and practical applications', *Proc. BCIforReal*, pp. 25–28. doi: 10.1145/3038439.3038444.
11. Yin, Z., Zhang, J. (2017) 'Cross-session classification of mental workload levels using EEG and an adaptive deep learning model', *Biomedical Signal Processing and Control*, 33, pp. 30–47. doi: 10.1016/j.bspc.2016.11.013.
12. Park, S.H., Lee, D., Lee, S.G. (2018) 'Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(2), pp. 498–505. doi: 10.1109/TNSRE.2017.2757519.
13. Arpaia, P., Esposito, A., Natalizio, A., Parvis, M. (2022) 'How to successfully classify EEG in motor imagery BCI: A metrological analysis of the state of the art', *JNE*, 19(3). doi: 10.1088/1741-2552/ac74e0.
14. Khademi, Z., Ebrahimi, F., Kordy, H.M. (2023) 'A review of critical challenges in MI-BCI: From conventional to deep learning methods', *J. of Neuroscience Methods*, 383, art. 109736. doi: 10.1016/j.jneumeth.2022.109736.
15. Tibrewal, N., Leeuwis, N., Alimardani, M. (2022) 'Classification of motor imagery EEG using deep learning increases performance in inefficient BCI users', *PLoS One*, 17(7), art. e0268880 doi: 10.1371/journal.pone.0268880.
16. Yamaev, M.I., Shipitsin, S.P., Filatova, E.S. (2018) 'Neural networks for EEG classification: From perceptron to convolutional LSTM', *Proc. Int. Conf. SCM*, 1, pp. 765–768 (in Russ.).

### Авторы

**Павленко Дмитрий Владимирович**<sup>1</sup>,  
программист, аспирант, pavlenkoprog@gmail.com  
**Татарис Шевхие Элверовна**<sup>1</sup>, студент,  
лаборант, Tataris.shevkhie1@gmail.com  
**Овчаренко Владимир Викторович**<sup>1</sup>,  
к.м.н., доцент, зав. лабораторией,  
rk\_vladimir@mail.ru

### Authors

**Dmitry V. Pavlenko**<sup>1</sup>, Programmer,  
Postgraduate Student, pavlenkoprog@gmail.com  
**Shevhiye E. Tataris**<sup>1</sup>, Student, Laboratory Assistant,  
Tataris.shevkhie1@gmail.com  
**Vladimir V. Ovcharenko**<sup>1</sup>, Cand. of Sci. (Medics),  
Associate Professor, Head of the Research Center,  
rk\_vladimir@mail.ru

<sup>1</sup> Крымский федеральный университет имени  
В.И. Вернадского, г. Симферополь, 295007, Россия

<sup>1</sup> V.I. Vernadsky Crimean Federal University,  
Simferopol, 295007, Russian Federation

## Сравнительный анализ методов сжатия табличных данных

К.В. Гарев<sup>1,2</sup>✉

<sup>1</sup> ФБГУ «Российский центр научной информации», г. Москва, 119334, Россия  
<sup>2</sup> Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия

### Ссылка для цитирования

Гарев К.В. Сравнительный анализ методов сжатия табличных данных // Программные продукты и системы. 2024. Т. 37. № 2. С. 170–177. doi: 10.15827/0236-235X.142.170-177

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 20.12.2023

После доработки: 24.01.2024

Принята к публикации: 14.02.2024

**Аннотация.** В работе проанализированы методы компрессии табличных данных, используемых в рамках функционирования платформы «Инфраструктура научно-исследовательских данных» (Платформа ИНИД). Оператором платформы является Российский центр научной информации. Основная цель исследования – выявление наиболее оптимальных методик сжатия данных, способных быть интегрированными в структуру Платформы ИНИД для доработки функционала по обмену и управлению научными данными. В ходе работы проведен тщательный анализ пяти наиболее популярных методов сжатия данных, доступных для реализации с использованием программных средств Python: Deflate (gzip), LZMA, Bzip2, Brotli и Snappy. Для каждой из рассмотренных технологий сжатия выявлены их преимущества и недостатки с учетом специфики обработки табличных данных, в том числе коэффициента сжатия, скорости обработки и степени сохранности информации. Результаты данного исследования могут использоваться в практике обмена, хранения, обработки и анализа табличных данных в рамках Платформы ИНИД. Особое внимание уделено анализу преимуществ и недостатков каждого из методов сжатия. Это позволяет сформировать рекомендации по выбору наиболее подходящих технологий, соответствующих строгим требованиям к производительности, эффективности сжатия и надежности сохранности данных. Важным аспектом исследования является фокусировка на возможности оптимизации процессов внутри Платформы ИНИД, что направлено на повышение ее эффективности как инструмента в области работы с научными данными. Кроме того, данная работа способствует углублению понимания потенциала применения современных методов сжатия данных в контексте практик обмена научными данными, открывая перспективу для дальнейших исследований и разработок в этой области.

**Ключевые слова:** методы сжатия, алгоритмы сжатия, Deflate, LZMA, Bzip2, Brotli, Snappy, сравнительный анализ

**Введение.** В последнее время объемы генерируемых и обрабатываемых данных в цифровом мире демонстрируют бурный рост [1]. Это касается как индивидуальных пользователей, так и крупных организаций, сталкивающихся с необходимостью эффективного хранения и передачи информации. В таком контексте алгоритмы сжатия данных приобретают особую актуальность, поскольку позволяют существенно уменьшить размеры файлов и, как следствие, сократить расходы на хранение данных, а также ускорить процесс их передачи через сети.

Каждый из существующих алгоритмов сжатия имеет уникальные характеристики, определяющие их предназначение и области применения. От выбора метода сжатия и реализующего его алгоритма зависят не только конечный размер файла, но и скорость его сжатия-распаковки, что влияет на общую производительность системы. Такие алгоритмы, как LZ77 и LZ78, разработанные в конце 1970-х годов, положили начало целому поколению методов сжатия, каждый из которых по-своему оптимизирует процесс уменьшения размера данных без потерь [2].

В последующие годы были разработаны и внедрены такие алгоритмы, как Deflate, который был представлен как эффективное сочетание сжатия LZ77 с кодированием Хаффмана, Brotli, обеспечивающий более высокие коэффициенты сжатия для веб-данных, а также Snappy и LZO, ориентированные на максимальную скорость сжатия и распаковки. Каждый из этих алгоритмов имеет свое предназначение, будь то максимально возможное сжатие данных или их быстрая обработка без значительных временных затрат.

Данная работа посвящена сравнительному анализу современных методов сжатия данных. Цель исследования – оценить эффективность каждого метода с точки зрения коэффициента сжатия и скорости обработки для последующего использования в рамках функционирования платформы «Инфраструктура научно-исследовательских данных» (Платформа ИНИД) (<https://data.rcsi.science/>).

### Описание алгоритмов сжатия

Цель сжатия – уменьшить размер файлов для экономии пространства на носителях ин-

формации и ускорения передачи данных по сети. Алгоритмы сжатия данных можно классифицировать на две большие категории: сжатие с потерями и без потерь. Каждый алгоритм имеет свои уникальные методы для достижения сжатия.

Сжатие без потерь позволяет точно восстановить исходные данные, в то время как сжатие с потерями приводит к уменьшению размера за счет невозможности точного восстановления оригинала, что приемлемо, например, в аудио, видео и изображениях. Сжатие без потерь использует такие методы, как кодирование Хаффмана, алгоритмы LZ77 и LZ78 и их производные (например, Deflate), а также преобразование Берроуза–Уилера с последующим кодированием длин серий (RLE) и кодированием Хаффмана (как в Bzip2).

Сжатие с потерями применяется в основном к мультимедийным данным, где небольшие потери качества могут быть незаметны человеку или приемлемы с точки зрения конечного использования. Примеры включают JPEG для изображений, MPEG для видео и MP3 для аудио.

Основой теории сжатия данных является понятие энтропии, введенное Клодом Шенноном [3]. Энтропия характеризует среднее количество информации (в битах), содержащееся в каждом символе данных, и определяет теоретический предел сжатия, который может быть достигнут без потерь для заданного набора данных. Алгоритмы сжатия стремятся приблизиться к этому пределу, уменьшая избыточность информации.

Рассмотрим наиболее известные алгоритмы сжатия.

**Deflate** – алгоритм сжатия данных, введенный Филом Кацем в 1993 году и используемый во многих форматах и протоколах, включая ZIP, gzip и PNG. Алгоритм сочетает в себе два основных метода сжатия – LZ77 (Lempel-Ziv 1977) и кодирование Хаффмана. Это позволяет эффективно сжимать данные, уменьшая их размер без потерь [4].

Рассмотрим подробнее основные шаги Deflate.

Первая часть Deflate – это алгоритм сжатия LZ77 (сжатие на базе словаря), в основе которого лежит процедура поиска и замены повторяющихся последовательностей символов в данных. LZ77 использует так называемый словарь – буфер, содержащий уже обработанные данные, для поиска повторяющихся последовательностей.

Прежде всего алгоритм ищет последовательности символов, которые ранее уже встре-

чались в данных. Если такая последовательность найдена, она заменяется ссылкой на ее предыдущее вхождение в словарь. Эта ссылка обычно представляет собой пару значений: смещение назад по тексту до начала повторяющейся последовательности и ее длину. Для данных, которые не могут быть сжаты с помощью ссылок на словарь (то есть уникальные или не имеющие совпадений в словаре), используется их непосредственное представление.

После применения LZ77 к данным Deflate использует кодирование Хаффмана для дальнейшего сжатия. Это метод сжатия данных без потерь, который использует переменную длину кода для представления символов. Символы, встречающиеся чаще, кодируются более короткими кодами, а редкие – более длинными.

Таким образом, сначала на основе частоты встречаемости символов и последовательностей в сжимаемых данных строится дерево Хаффмана. Каждому символу или последовательности присваивается уникальный двоичный код в соответствии с его положением в дереве.

Затем данные кодируются с использованием полученных двоичных кодов, что приводит к их дальнейшему уменьшению в размере.

Итоговый сжатый поток данных в формате Deflate состоит из серии блоков, каждый из которых может быть закодирован независимо. Блоки могут использовать различные стратегии кодирования в зависимости от того, какой подход обеспечивает лучшее сжатие для конкретного набора данных. Deflate также позволяет использовать для кодирования фиксированное или динамическое дерево Хаффмана, что дает дополнительную гибкость в оптимизации процесса сжатия.

**LZMA (Lempel-Ziv-Markov chain Algorithm)** – алгоритм сжатия данных без потерь, который сочетает в себе методы словарного сжатия и адаптивное кодирование арифметическим кодом [5]. Он базируется на алгоритме LZ77 и расширяет его возможности, улучшая эффективность сжатия за счет использования более сложных моделей предсказания.

Процесс сжатия с использованием LZMA начинается с построения словаря входных данных. Затем алгоритм находит повторяющиеся последовательности байтов. Каждая повторяющаяся последовательность заменяется ссылкой на ее предыдущее вхождение в словарь. Этот процесс позволяет сократить размер данных за счет удаления избыточности информации.

После этапа словарного сжатия применяется адаптивное кодирование арифметическим кодом, используемое для дальнейшего умень-

шения размера данных. На этом этапе каждый символ заменяется кодом, который зависит от вероятности его появления в текущем контексте. Алгоритм адаптивно обновляет значения вероятностей появления символов на основе предыдущих символов, что позволяет ему эффективно сжимать данные с учетом их структуры.

Одним из ключевых преимуществ LZMA является высокий коэффициент сжатия, особенно для файлов большого размера или данных с высокой степенью избыточности. Однако этот высокий коэффициент сжатия достигается за счет использования более сложных алгоритмов, что может привести к увеличению времени сжатия и распаковки данных по сравнению с некоторыми другими методами сжатия.

**Bzip2** – алгоритм сжатия данных без потерь, который использует алгоритм Берроуза–Уилера (Burrows–Wheeler Transform, BWT) в сочетании с кодированием длин серий (Run-Length Encoding, RLE) [6] и кодированием Хаффмана. Алгоритм разработан Джулианом Сьюардом, изначально был представлен в 1996 году и предназначен для сжатия текстовых данных с высоким коэффициентом сжатия при приемлемой скорости выполнения [7].

Процесс сжатия в Bzip2 включает несколько ключевых шагов. На первом этапе исходные данные переупорядочиваются таким образом, чтобы создать условия для более эффективного последующего сжатия. BWT группирует похожие символы вместе, делая текст более однородным и улучшая условия для следующих этапов сжатия. После применения BWT исходный текст может содержать серии повторяющихся символов. RLE сокращает эти серии до более компактного представления, записывая символ и количество его последовательных повторений. Затем осуществляется сортировка по контексту. Этот шаг помогает повысить эффективность последующего кодирования Хаффмана, дополнительно упорядочивая символы по их контексту. На последнем этапе применяется кодирование Хаффмана, которое заменяет часто встречающиеся символы или группы символов короткими кодами, а редкие – более длинными. Это дополнительно сжимает данные, основываясь на их частотности.

Алгоритм Bzip2 обрабатывает данные блоками, обычно размером в 900 КБ. Каждый блок сжимается независимо, что позволяет эффективно распараллелить процесс сжатия и распаковки, а также облегчает восстановление данных при частичном повреждении архива.

**Brotli** – алгоритм сжатия данных, разработанный Google и предназначенный в основном

для сжатия веб-контента [8]. Он обеспечивает высокий коэффициент сжатия при сохранении приемлемой скорости обработки данных.

Процесс сжатия с использованием Brotli начинается с построения словаря входных данных, который затем используется для поиска повторяющихся последовательностей. После этого применяются различные методы адаптивного кодирования, такие как кодирование Хаффмана и предсказательное кодирование, для дальнейшего уменьшения размера данных. Алгоритм также может использовать контекстное моделирование для эффективного сжатия текстовых данных.

Одним из ключевых преимуществ Brotli является специальная оптимизация для веб-контента, что делает его особенно эффективным для сжатия HTML, CSS, JavaScript и других типов файлов, используемых на веб-страницах. Он также предлагает высокий коэффициент сжатия при высокой скорости обработки данных, что делает его привлекательным для использования в онлайн-приложениях и сервисах.

**Snappy** – алгоритм сжатия данных, также разработанный Google. Применяется в широком спектре приложений, где требуется высокая скорость сжатия и распаковки данных при относительно низком коэффициенте сжатия [9].

Процесс сжатия с использованием Snappy основан на алгоритме сжатия данных без потерь, обеспечивающем быстрое сжатие и распаковку данных путем замены повторяющихся последовательностей байтов ссылками на их предыдущие вхождения в словаре. Этот метод сжатия приводит к относительно низкому коэффициенту сжатия по сравнению с некоторыми другими алгоритмами, но при этом обеспечивает высокую скорость обработки данных.

Одним из ключевых преимуществ Snappy является высокая скорость сжатия и распаковки данных. Он оптимизирован для обеспечения максимальной производительности при сжатии и распаковке данных в реальном времени, что делает его идеальным для использования в онлайн-сервисах и приложениях, где скорость обработки данных критически важна.

Следует отметить, что Snappy может быть менее эффективным в сжатии данных с высокой степенью избыточности, таких как текстовые данные или данные с повторяющимися паттернами. Это следует учитывать при выборе метода сжатия в зависимости от конкретных характеристик данных и требований приложения.

## Критерии сравнения алгоритмов сжатия

Для определения критериев отбора методов следует учитывать ряд аспектов.

1. *Коэффициент сжатия.* Это ключевой показатель эффективности метода сжатия. Важно выбрать методы сжатия, обеспечивающие наиболее высокий коэффициент, что позволит минимизировать объем хранимых и передаваемых данных, сокращая затраты на хранение и улучшая производительность системы. В данном контексте коэффициент сжатия определяется как отношение объема несжатого файла к объему сжатого.

2. *Сохранность данных.* Учитывая биологическую значимость данных о структурных мотивах белков, важно обеспечить сохранность информации при сжатии. Чтобы избежать искажений или потерь важных деталей структуры белков, предпочтительно использовать такие методы сжатия без потерь, как Deflate, LZMA и Bzip2.

3. *Скорость сжатия.* Важным критерием является скорость сжатия данных. Учитывая, что БД может содержать большие объемы информации, желательно выбирать методы сжатия, обеспечивающие приемлемую скорость обработки данных. Это позволит эффективно работать с данными и обеспечивать быстрый доступ к ним.

4. *Поддержка и совместимость.* Важно выбрать методы сжатия, которые хорошо поддерживаются в современных программных и аппаратных системах, а также имеют широкую совместимость с используемыми платформами и инструментами. Это обеспечит легкость интеграции выбранных методов существующей инфраструктурой и упростит их использование.

Тестовым набором данных, структура которого показана в таблице 1, является один из датасетов, представленных на Платформе ИНИД «База данных структурных мотивов белков: биологические и физико-химические свойства». Набор данных предназначен для проведения структурного анализа белков, ассоциированных с развитием заболеваний, в том числе aberrantных форм белков, которые образованы вследствие аминокислотных замен, модифицирования после синтеза. Набор данных может применяться для решения прикладных медико-биологических задач, таких как разработка новых подходов к диагностике заболеваний, изучение молекулярных основ патогенеза, выявление мишеней белкового происхождения для лекарственных средств и проектирование

миметиков (белков с заданными свойствами). Набор данных содержит 3.96 млн аннотаций структурных мотивов в белковых структурах с указанием внутренних координат (<http://data.rcsi.science/data-catalog/datasets/203/>).

## Сравнительный анализ алгоритмов сжатия

Сравниваемые алгоритмы были реализованы посредством использования python-библиотеки. Вычислительный эксперимент проводился с использованием вычислительных мощностей виртуальной машины со следующими характеристиками: 32 ядра, Intel Xeon, 72 Гб оперативной памяти, сервер под управлением Linux Debian 11.

Результаты сравнительного анализа представлены в таблице 2.

Графическая интерпретация результатов сравнения алгоритмов представлена на рисунке.

Таким образом, на основе проведенного анализа и данных вычислительного эксперимента можно выделить преимущества и недостатки алгоритмов сжатия.

**Deflate.** Преимущества: баланс между скоростью и коэффициентом сжатия. Deflate обеспечивает относительно высокий коэффициент сжатия при сохранении приемлемой скорости сжатия и распаковки.

Широкая поддержка: поддерживается большинством программных и аппаратных систем, что делает его универсальным выбором для сжатия табличных данных.

Недостатки: не оптимизирован для табличных данных. Может быть не таким эффективным для специфических типов табличных данных, как другие алгоритмы, специально разработанные для этой цели.

**LZMA.** Преимущества: высокий коэффициент сжатия. LZMA обеспечивает один из самых высоких коэффициентов сжатия среди алгоритмов без потерь, что может быть критически важным для больших объемов табличных данных.

Эффективность сжатия: особенно эффективен для сжатия больших наборов данных с высокой избыточностью информации.

Недостатки: скорость сжатия и распаковки. LZMA обычно работает медленнее других алгоритмов, что может быть ограничивающим фактором для реальных приложений, требующих быстрого доступа к данным.

**Bzip2.** Преимущества: высокий коэффициент сжатия для определенных типов данных. Bzip2 часто превосходит gzip по коэффициенту

Таблица 1

## Структура набора данных

Table 1

## Dataset structure

Атрибут	Описание	Количество пропусков (NaN) (гш/%)	Формат
experiment_id	Идентификатор трехмерной структуры в Protein Data Bank	0	String
uniprot_id	Идентификатор белка в базе знаний UniProt	1330	String
protein_name	Название белка	101 949	String
chain_id	Идентификатор белковой цепи в экспериментальной структуре	98 257	String
resolution	Разрешение трехмерной структуры белка	2 712 626	Unmeric
experiment_type	Тип эксперимента, в котором получена трехмерная структура белка	0	String
biological_process	Биологический(е) процесс(ы), в которых участвует белок	2 436 633	String
molecular_function	Функция(и), которую выполняет белок в клетке/организме	2 006 228	String
fasta	Аминокислотная последовательность структурного мотива в формате однобуквенного кода	0	String
apfid	Уникальный идентификатор мотива, который содержит указание на экспериментальную структуру, позиции начала и конца мотива	0	String
motif_type	Тип структурного мотива	0	String
motif_len	Длина структурного мотива	0	Integer
motif_start	Позиция начала структурного мотива в цепи белка	1 567	Integer
motif_end	Позиция конца структурного мотива в цепи белка	1 567	Integer
organism	Название организма, которому принадлежит белок (англ.)	98 257	String

сжатия, особенно для файлов большого размера. Это делает его предпочтительным для сжатия объемных табличных данных.

Недостатки: скорость сжатия и распаковки, в сравнении с такими алгоритмами, как Snappy или Deflate, работает медленнее.

Использование ресурсов: Vzip2 может требовать больше оперативной памяти для сжатия и распаковки по сравнению с некоторыми другими алгоритмами, что важно учитывать при работе с ограниченными ресурсами или большими объемами данных.

**Brotli.** Преимущества: оптимизирован для веб-данных. Brotli был специально разработан для сжатия веб-контента, что делает его осо-

бенно подходящим для сжатия табличных данных, передаваемых через веб.

Высокий коэффициент сжатия при высокой скорости: обеспечивает отличный коэффициент сжатия, при этом оставаясь достаточно быстрым для онлайн-приложений.

Недостатки: совместимость и поддержка. Хотя поддержка Brotli растет, она все еще не так универсальна, как у более старых алгоритмов, таких как gzip.

**Snappy.** Преимущества: высокая скорость сжатия и распаковки. Snappy оптимизирован для быстрого сжатия и распаковки, что делает его подходящим алгоритмом для ситуаций, где время обработки критично.

Таблица 2

**Сравнение алгоритмов сжатия**

Table 2

**Comparison of compression algorithms**

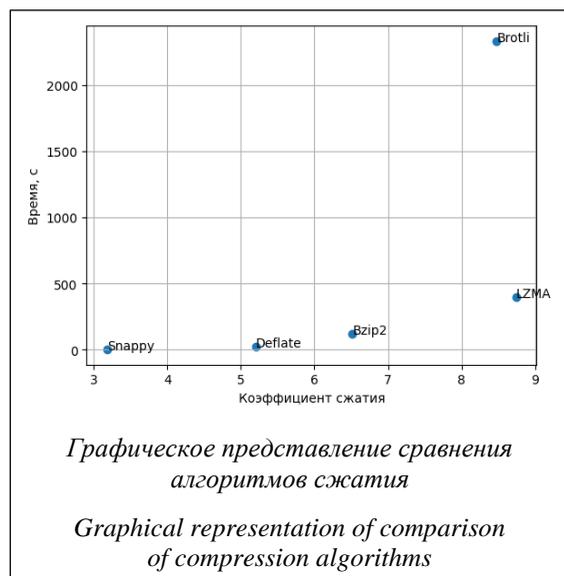
Метод сжатия	Алгоритм сжатия	Коэффициент сжатия	Время сжатия
LZ77 + Huffman	Deflate	5,203	0:00:25.282951
LZ77 + марковские цепи	LZMA	8,742	0:06:38.348505
BWT + RLE + Huffman	Bzip2	6,517	0:01:59.607255
	Brotli	8,476	0:38:52.891844
Поиск совпадений (как в LZ77)	Snappy	3,182	0:00:03.906618

Недостатки: коэффициент сжатия. Может не достигать такого же уровня сжатия, как некоторые другие алгоритмы, особенно для данных с высокой избыточностью.

**Практическое применение**

Использование оптимальных методов сжатия данных, выявленных в рамках исследования, в составе Платформы ИНИД и других информационных систем, оперирующих сверхбольшими объемами данных, позволит значительно оптимизировать процессы обмена, хранения и анализа табличных данных.

Интеграция результатов исследования в Платформу ИНИД обеспечит пользователям возможность эффективно управлять объемами наборами данных и уменьшить нагрузку на инфраструктуру. Благодаря выбору оптимальных методов сжатия, основанному на реальных данных и анализе их характеристик, платформа сможет предоставлять пользователям быстрый доступ к информации и обеспечивать высокую скорость обработки данных.



Использование результатов исследования в Платформе ИНИД также способствует повышению ее конкурентоспособности и привлекательности для научного сообщества. Эффективное управление научными данными, обеспечиваемое оптимальными методами сжатия, будет способствовать развитию научных исследований в различных областях.

**Заключение**

В ходе исследования был проведен сравнительный анализ пяти наиболее распространенных методов сжатия данных – Deflate, LZMA, Bzip2, Brotli и Snappy с использованием табличных данных, представляющих БД структурных мотивов белков. Каждый метод был оценен по коэффициенту сжатия и времени сжатия для определения их эффективности и производительности в контексте данного исследования.

На основе полученных результатов можно сделать следующие выводы.

Deflate продемонстрировал приемлемый коэффициент и низкое время сжатия, что делает его подходящим для быстрой обработки данных с относительно хорошим коэффициентом сжатия.

LZMA показал высокий коэффициент сжатия, однако его время сжатия значительно превышает время других методов, что делает его менее подходящим для приложений, требующих быстрого действия.

Bzip2 продемонстрировал хороший баланс между коэффициентом и временем сжатия, что делает его универсальным методом для различных приложений сжатия данных.

Brotli показал высокий коэффициент сжатия, но его время сжатия значительно превышает время других методов, что делает метод менее подходящим для сценариев, где требуется быстрая обработка данных.

Snappy продемонстрировал низкий коэффициент сжатия, однако обеспечил наименьшее время, что делает его подходящим для задач, где скорость обработки данных приоритетна.

Таким образом, выбор метода сжатия должен основываться на конкретных требованиях и характеристиках приложения, включая коэффициент сжатия, скорость обработки данных и доступные ресурсы.

#### Список литературы

1. Manyika J., Chui M., Brown B., Bughin J., Dobbs R. et al. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey & Company, 2011, 146 p. URL: [https://personal.utdallas.edu/~muratk/courses/cloud11f\\_files/MGI-full-report.pdf](https://personal.utdallas.edu/~muratk/courses/cloud11f_files/MGI-full-report.pdf) (дата обращения: 10.01.2024).
2. Choudhary S.M., Patel A.S., Parmar S.J. Study of LZ77 and LZ78 data compression techniques. IJESIT, 2015, vol. 4, no. 3, pp. 2319–5967.
3. Shannon C.E. A mathematical theory of communication. The Bell System Tech. J., 1948, vol. 27, no. 3, pp. 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x.
4. Oswal S., Singh A., Kumari K. Deflate compression algorithm. IJERGS, 2016, vol. 4, no. 1, pp. 430–436.
5. Adjeroh D., Bell T., Mukherjee A. The Burrows-Wheeler Transform: Data compression, Suffix Arrays, and Pattern Matching. NY, Springer Publ., 2008, 352 p. doi: 10.1007/978-0-387-78909-5.
6. Sarika S., Srilali S. Improved run length encoding scheme for efficient compression data rate. IJERA, 2013, vol. 3, no. 6, pp. 2248–9622.
7. Gupta A., Bansal A., Khanduja V. Modern lossless compression techniques: Review, comparison and analysis. Proc. ICECCT, 2017, pp. 1–8. doi: 10.1109/ICECCT.2017.8117850.
8. Alakuijala J., Farruggia A., Ferragina P. et al. Brotli: A general-purpose data compressor. ACM Transactions on Inform. Sys., 2019, vol. 37, pp. 1–30. doi: 10.1145/3231935.
9. Chen R., Xu L. SnappyR: A new high-speed lossless data compression algorithm. Proc. DCC, 2023, pp. 334–334. doi: 10.1109/DCC55655.2023.00052.

#### Compression methods for tabular data: Comparative analysis

Kirill V. Garev <sup>1,2</sup>✉

<sup>1</sup> Russian Center for Science Information, Moscow, 119334, Russian Federation

<sup>2</sup> Joint Supercomputer Center of RAS, Moscow, 119334, Russian Federation

#### For citation

Garev, K.V. (2024) 'Compression methods for tabular data: Comparative analysis', *Software & Systems*, 37(2), pp. 170–177 (in Russ.). doi: 10.15827/0236-235X.142.170-177

#### Article info

Received: 20.12.2023

After revision: 24.01.2024

Accepted: 14.02.2024

**Abstract.** This paper presents a comparative analysis of table data compression methods used within in terms of the Research Data Infrastructure platform (RDI Platform) operated by the Russian Center for Science Information. The main purpose of the study was to identify the most optimal data compression methods that can be integrated into the RDI Platform structure to enhance the functionality of data exchange and management. The author of the study has carried out a thorough analysis of the five most popular data compression methods available for implementation using the following Python software tools: Deflate (gzip), LZMA, Bzip2, Brotli, and Snappy. The author has analyzed advantages and disadvantages of each considered compression technology, taking into account the specifics of table data processing, including compression ratio, processing speed, and the degree of information preservation. The results of this study can contribute to the practice of exchanging, storing, processing, and analyzing table data in terms of the RDI Platform. The author paid particular attention to analyzing the advantages and disadvantages of each compression method, which allowed forming recommendations for choosing the most suitable technologies that meet strict requirements for performance, compression efficiency, and data preservation reliability. An important aspect of the study is focusing on the possibilities of optimizing processes within the RDI Platform, which increases its efficiency as a tool in the field of working with scientific data. Moreover, this work helps

deepen the understanding of the potential application of modern data compression methods in terms of scientific data exchange practice; it opens prospects for further research and development in this area.

**Keywords:** compression methods, compression algorithms, Deflate, LZMA, Bzip2, Brotli, Snappy, comparative analysis

### References

1. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R. et al. (2011) *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey & Company, 146 p., available at: [https://personal.utdallas.edu/~muratk/courses/cloud11f\\_files/MGI-full-report.pdf](https://personal.utdallas.edu/~muratk/courses/cloud11f_files/MGI-full-report.pdf) (accessed January 10, 2024).
2. Choudhary, S.M., Patel, A.S., Parmar, S.J. (2015) 'Study of LZ77 and LZ78 data compression techniques', *IJESIT*, 4(3), pp. 2319–5967.
3. Shannon, C.E. (1948) 'A mathematical theory of communication', *The Bell System Tech. J.*, 27(3), pp. 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x.
4. Oswal, S., Singh, A., Kumari, K. (2016) 'Deflate compression algorithm', *IJERGS*, 4(1), pp. 430–436.
5. Adjeroh, D., Bell, T., Mukherjee, A. (2008) *The Burrows-Wheeler Transform: Data Compression, Suffix Arrays, and Pattern Matching*. NY: Springer Publ., 352 p. doi: 10.1007/978-0-387-78909-5.
6. Sarika, S., Srilali, S. (2013) 'Improved run length encoding scheme for efficient compression data rate', *IJERA*, 3(6), pp. 2248–9622.
7. Gupta, A., Bansal, A., Khanduja, V. (2017) 'Modern lossless compression techniques: Review, comparison and analysis', *Proc. ICECCT*, pp. 1–8. doi: 10.1109/ICECCT.2017.8117850.
8. Alakujala, J., Farruggia, A., Ferragina, P. et al. (2019) 'Brotli: A general-purpose data compressor', *ACM Transactions on Inform. Sys.*, 37, pp. 1–30. doi: 10.1145/3231935.
9. Chen, R., Xu, L. (2023) 'SnappyR: A new high-speed lossless data compression algorithm', *Proc. DCC*, pp. 334–334. doi: 10.1109/DCC55655.2023.00052.

### Авторы

**Гарев Кирилл Валерьевич**<sup>1,2</sup>,  
начальник управления, научный сотрудник,  
garev.kv@gmail.com, kv@garev.ru

<sup>1</sup> ФБГУ «Российский центр научной информации»,  
г. Москва, 119334, Россия

<sup>2</sup> Межведомственный суперкомпьютерный  
центр РАН, г. Москва, 119334, Россия

### Authors

**Kirill V. Garev**<sup>1,2</sup>, Head of the Department,  
Research Associate,  
garev.kv@gmail.com, kv@garev.ru

<sup>1</sup> Russian Center for Science Information,  
Moscow, 119334, Russian Federation

<sup>2</sup> Joint Supercomputer Center of RAS,  
Moscow, 119334, Russian Federation

## Применение алгоритмов сжатия информации для задачи когнитивного отображения проектного пространства многослойных экранов защиты от радиации

Л.А. Зинченко <sup>1</sup>✉, В.В. Казаков <sup>1</sup>, Б.В. Карышев <sup>1</sup>

<sup>1</sup> МГТУ им. Н.Э. Баумана, г. Москва, 105005, Россия

### Ссылка для цитирования

Зинченко Л.А., Казаков В.В., Карышев Б.В. Применение алгоритмов сжатия информации для задачи когнитивного отображения проектного пространства многослойных экранов защиты от радиации // Программные продукты и системы. 2024. Т. 37. № 2. С. 178–185. doi: 10.15827/0236-235X.142.178-185

### Информация о статье

Группа специальностей ВАК: 2.3.7

Поступила в редакцию: 31.01.2023

После доработки: 14.11.2023

Принята к публикации: 21.11.2023

**Аннотация.** В статье рассмотрены особенности применения алгоритмов снижения размерности в задачах проектирования многослойных радиационно стойких экранов защиты электронной аппаратуры для использования в космическом пространстве. Алгоритмы проецируют исследуемые данные о многослойных защитных экранах из пространства с высокой размерностью в пространство с низкой размерностью. При этом сохраняется семантика данных, что позволяет визуализировать большие наборы информации высокой размерности и многократно упрощает пользовательский визуальный анализ, а также применение некоторых алгоритмов и подходов в автоматизированном режиме. В статье рассмотрено применение следующих алгоритмов снижения размерности: анализ главных компонент (PCA), анализ ядра главных компонент (KernelPCA), стохастическое вложение соседей с t-распределением (t-SNE), равномерное приближение и проекция (UMAP), автоэнкодер (AE), вариационный автоэнкодер (VAE). В случае нейросетевых архитектур сжатия представлены предлагаемые архитектуры сетей, используемые в расчетах и тестировании. По предложенной методике исследована целесообразность комбинации нескольких алгоритмов снижения размерности, применяемых по цепочке. На основании проведенного исследования сделан вывод о результативности применения указанных алгоритмов, а также их комбинации для последующей обработки или визуализации. Кратко рассмотрено программное обеспечение, реализующее один из предложенных подходов в анализе и обработке информации о многослойных радиационно стойких экранах защиты электронной аппаратуры для использования в космическом пространстве. В результате сделан выбор в пользу алгоритма UMAP. Для анализа конфигураций с достаточно большим количеством параметров рекомендуется использовать алгоритм t-SNE с предварительным сжатием алгоритмом UMAP, который упрощает исходный набор данных, улучшая результат работы t-SNE.

**Ключевые слова:** когнитивная визуализация, сжатие данных, алгоритмы снижения размерности, проецирование

**Введение.** Для обеспечения стабильного функционирования электронной аппаратуры в космическом пространстве применяются экраны защиты от радиации. При их проектировании необходимо сравнить большое количество альтернативных проектных решений и выбрать наиболее подходящее [1]. Вычисление характеристик каждого альтернативного проектного решения требует больших временных затрат. При этом проектировщику необходимо провести анализ полученных решений. В существующих программных средствах, таких как SRIM [2] и GEANT4 [3], анализ возможных структур экранов защиты не реализован. Таким образом, возникает необходимость в ПО, упрощающем выбор проектного решения.

Для решения проблемы анализа конфигураций экранов защиты предложена методика с использованием алгоритма t-SNE [4]. Производится декомпозиция параметров, характеризующих структурные составляющие многослой-

ного защитного экрана, на различные уровни – многослойный экран, материалы и отдельные химические элементы. После чего полученные параметры представляются в численном виде. В случае с характеристиками, не являющимися численными, можно использовать кодировку, где каждому возможному значению параметра соответствует свой уникальный численный код, либо так называемый метод one-hot encoding, при котором каждому возможному значению характеристики соответствует элемент множества, принимающий значение 1, если значение характеристики соответствует данному элементу, и 0, если не соответствует. Таким образом, каждой конфигурации соответствует множество численных значений, которые можно рассматривать как точки в пространстве высокой размерности. Далее используется алгоритм t-SNE, проецирующий многомерные данные в двухмерное пространство с сохранением общей структуры данных. После этого строятся

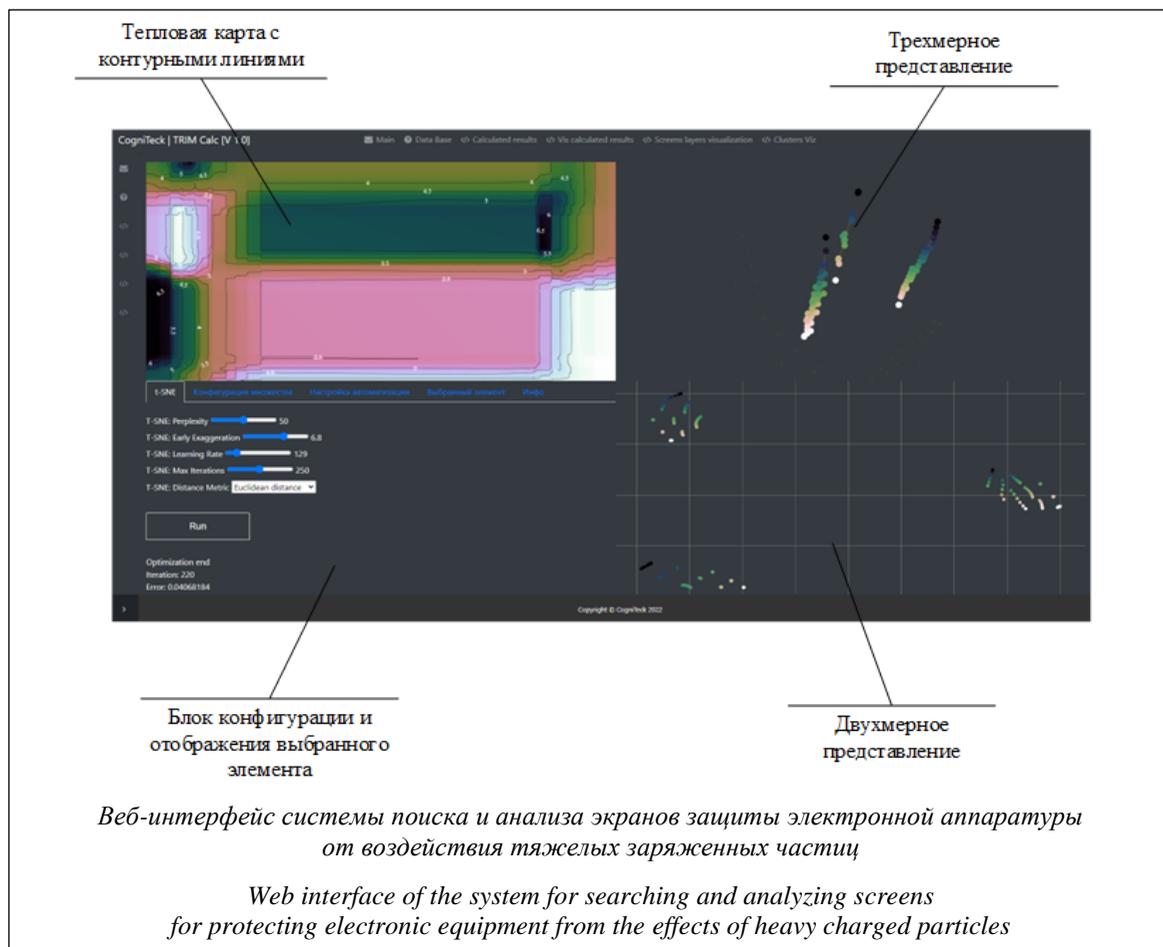
когнитивные визуальные образы на основе полученной проекции точек, отражающих возможные конфигурации, благодаря чему значительно упрощается восприятие данных и разработчик может определить степень сходства конфигураций между собой [5]. Такой подход позволяет приблизительно оценить характеристики группы конфигураций многослойных экранов защиты, проведя расчеты для одной или нескольких из них.

Важным элементом системы является методика сжатия данных, которая напрямую влияет на качество получаемых когнитивных представлений. В случае последующего прогноза некоторых целевых показателей для оптимизации потребления памяти, нагрузки на центральный процессор или графического ядро (в зависимости от алгоритма прогноза и его реализации) сжатие может выступать дополнительным этапом предобработки, повышающим эффективность последующего этапа прогноза. В таком случае выбор наилучшего алгоритма или предлагаемой новой методики, включающей несколько стадий сжатия различными алгоритмами, является крайне важным этапом

при проектировании систем визуализации и прогноза, в связи с чем необходимо провести анализ и оценку алгоритмов сжатия, а также методологии многостадийного сжатия информации для выбора наилучшего решения.

Интерфейс системы анализа конфигураций с когнитивными элементами, изображенный на рисунке, разделен на четыре зоны, три из которых представляют когнитивно-графический образ проекции многомерного множества параметров конфигураций в пространство с низкой размерностью, а один – настройки параметров алгоритма сжатия и графического образа, а также задание исследуемого набора конфигураций экранов защиты. Тепловая карта показывает двухмерное цветовое представление одного из исходных параметров и трехмерное представление, отражающее значение определенного параметра конфигурации, вынося его на ось z. В двухмерном и трехмерном представлениях также цветом выделен один из параметров конфигурации (в данном случае – масса).

Основная задача заключается в том, чтобы оценить различные алгоритмы сжатия много-



мерных данных и их комбинаций и найти оптимальный вариант, позволяющий добиться наименьших потерь информации для последующих задач прогнозирования и визуализации.

В разработанной системе на вход подаются параметры каждого защитного экрана. Параметры химического элемента в составе слоя: процент содержания химических элементов в материале (для составных материалов, например, сплав и другие), порядковый номер химического элемента в периодической таблице, период химического элемента в периодической таблице, группа химического элемента в периодической таблице, подгруппа химического элемента в периодической таблице, атомная масса химического элемента, семейство химического элемента, электроотрицательность химического элемента, температура плавления химического элемента, температура кипения химического элемента, вес материала на 1 м<sup>3</sup>, плотность химического элемента.

Параметры слоя многослойного экрана: вес 1 м<sup>3</sup> материала (слоя многослойного экрана), общий вес слоя, количество химических элементов, из которых состоит материал (слоя многослойного экрана), толщина слоя многослойного экрана.

Параметры многослойного экрана в целом: количество слоев в данном многослойном экране, вес многослойного экрана площадью 1 м<sup>2</sup> (толщина экрана определяется исходя из параметров его слоев), общая толщина многослойного экрана (в мм).

Часть из приведенных параметров используется напрямую в структуре данных для непосредственного сжатия, а у некоторых (таких, как порядковый номер в периодической таблице) необходимо корректно сопоставить данные о химическом элементе и его объекте в рамках разработанной системы.

Предложенный подход работы с многомерными данными с применением алгоритма t-SNE был реализован в программном комплексе CogShield [6]. Для дальнейшего повышения качества визуального анализа многомерных данных необходимо проанализировать иные алгоритмы сжатия и их комбинации и предложить подход в оценке алгоритмов, не зависящий от задачи приложения.

### Обзор алгоритмов снижения размерности

Существует большое количество алгоритмов снижения размерности (см. [6]), которые можно комбинировать между собой. Выбор ал-

горитма или их комбинации напрямую влияет на результативность описанной выше методики проектирования экрана защиты от радиации. При этом важно выбрать для использования такой алгоритм, который максимально сохраняет полезную информацию в данных и исключает дублирование информации или малозначимые компоненты.

**Автоэнкодер (Autoencoder – AE)** представляет две связанные между собой нейронные сети: кодер, сжимающий данные до заданной размерности, и декодер, разжимающий их обратно [7]. Во время обучения их объединяют в одну нейросеть, задача которой сжать вектор, а потом восстановить его как можно точнее. Чаще всего обучение происходит методом градиентного спуска. Существенным недостатком автоэнкодера является его низкая вычислительная эффективность.

**Вариационный автоэнкодер (Variational Autoencoder – VAE)** позволяет сформировать сжатое множество в соответствии с заданным законом распределения (обычно используется нормальное распределение) [8]. Это приводит к тому, что вариационный декодер преобразует точки в окрестности, полученные энкодером, в изображения, похожие на исходные. Благодаря этому свойству вариационный автоэнкодер используется для генерации данных, схожих с элементами некоего множества. При исследовании выбрана одинаковая структура вариационного автоэнкодера и автоэнкодера. В качестве функции активации для нейронных сетей использована ReLU [9], так как она позволяет автоэнкодерам выявлять нелинейные закономерности в данных. В эксперименте энкодер и декодер включали по пять слоев, причем количество нейронов зависело от размера входного вектора. В таблице 1 представлено количество нейронов в каждом слое кодера. Декодер имел симметричную структуру.

Таблица 1

Структура кодера

Table 1

Coder structure

Слой				
1	2	3	4	5
Входной вектор размера $n$	$2n$	$n$	$n$	Вектор скрытого состояния (в данной работе размер равен 2)

**Метод анализа главных компонент (PCA)** состоит в выборе главных компонент – новой системы координат, лучше описываю-

щей данные [10]. В случае с нормальным распределением данные напоминают многомерный эллипс, главные компоненты совпадают с осями этого эллипса. Затем вычисляются дисперсии значений по компонентам. Далее компонента, вносящая минимальный вклад в информативность, удаляется.

**Метод KernelPCA** представляет собой улучшенную версию PCA, способную выявлять нелинейные зависимости в данных [11].

**Алгоритм t-SNE** [12] включает следующие шаги. Для всех элементов множества вводится и вычисляется функция, преобразующая многомерное евклидово расстояние между точками в условные вероятности, отражающие сходство точек между собой. Элементы исходного множества случайным образом проецируются в пространство низкой размерности, и для каждого элемента нового множества вычисляется аналогичный показатель. Посредством изменения положения элементов множества-проекции увеличивается сходство структуры исходного множества и его проекции. Сходство оценивается по сумме дивергенций Кульбака-Лейблера [13, 14] для всех точек. У алгоритма t-SNE есть очень важный параметр – перплексия. При ее высоком значении сохраняется глобальная структура данных, а при низком – локальная.

**Алгоритм UMAP** [15] по принципу работы похож на алгоритм t-SNE. Его преимуществами являются улучшенное быстродействие, сохранение глобальной структуры множества при проецировании: расстояние между кластерами после сжатия t-SNE ничего не говорит о схожести кластеров между собой, в то время как после алгоритма UMAP близость кластеров свидетельствует об их схожести. Алгоритм UMAP в отличие от алгоритма t-SNE можно эффективно применять не только для отображения данных, но и для общего снижения размерности с целью избавления от незначимых компонент и мультиколлинеарности в данных.

### Сравнение алгоритмов и их комбинаций

Для количественного определения качества работы алгоритма снижения размерности в работе использованы алгоритмы классификации, для оценки результативности которых была использована метрика Ассигасу, вычисляемая как отношение количества правильно распознанных элементов к их общему количеству. Сначала алгоритм классификации обучался на исходном наборе данных, а после – на сжатом.

Разность между результатом после сжатия и до сжатия принята как показатель качества работы алгоритма сжатия. Чем он выше, тем меньше информации было потеряно при сжатии. Применялись следующие алгоритмы: градиентный бустинг, k-средних и k-ближайших соседей. Если точность прогноза алгоритма после сжатия падала ниже вероятности случайного угадывания, то в качестве результата принималось число –1. При иных значениях для неудачного сжатия результаты менялись.

Для исследования были выбраны следующие наборы данных:

- случайная выборка из 1 000 элементов из набора MNIST (<http://yann.lecun.com/exdb/mnist/>), представляющего собой набор черных белых рукописных цифр 28×28 пикселей, подаваемых на вход в виде числовых массивов; целевое значение для прогноза – значение цифры;
- Water Quality (<https://www.kaggle.com/datasets/adityakadiwal/water-potability>) – набор химических и физических характеристик воды; целевое значение для прогноза – пригодность воды для питья;
- Mobile Price Classification (<https://www.kaggle.com/datasets/iabhishekoofficial/mobile-price-classification>) – набор характеристик телефонов; целевое значение для прогноза – стоимостная группа;
- Titanic (<https://www.kaggle.com/c/titanic>) – информация о пассажирах; целевое значение для прогноза – выжил ли человек после крушения;
- Iris (<https://www.kaggle.com/datasets/uciml/iris>) – набор параметров цветков ириса; целевое значение для прогноза – подвид;
- Student's scores (<https://www.kaggle.com/datasets/samarsaeedkhan/scores>) – набор оценок студентов по различным предметам; целевое значение для прогноза – удовлетворенность оценками.

В ходе эксперимента алгоритмы UMAP и t-SNE использовались с различными значениями количества соседей и переплексии соответственно, после чего выбирался наилучший результат. Параметры остальных алгоритмов в ходе эксперимента не менялись.

В процессе работы также была исследована результативность предварительного сжатия с использованием алгоритмов PCA, KernelPCA и UMAP. Остальные алгоритмы из перечисленных выше для многоэтапного сжатия не подходят, так как алгоритм t-SNE не способен сжи-

мать данные в размерности больше 3, а автоэнкодеры обладают слишком низкой вычислительной эффективностью. Три последних набора данных обладают относительно низкой исходной размерностью (3–6 измерений), поэтому исследовать на них предварительное сжатие не имело смысла, поскольку все отображение будет выполнено первым примененным алгоритмом.

Для эксперимента использовались следующие реализации алгоритмов:

- t-SNE – пакет sklearn (<https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html>);
- UMAP – реализация авторов алгоритма (<https://umap-learn.readthedocs.io/en/latest/index.html>);
- PCA – пакет sklearn (<https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>);
- KernelPCA – пакет sklearn (<https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.KernelPCA.html>);
- AE и VAE – реализованы с использованием keras (<https://keras.io/>).

Для выполнения эксперимента был использован шестиядерный процессор AMD Ryzen 5 5500U с частотой 2,1 ГГц и 16 Гб оперативной памяти. Результаты эксперимента представлены в таблицах 2 и 3.

Таблица 2

**Результаты работы алгоритмов без предварительного сжатия**

Table 2

**Results of algorithms without precompression**

Метод	Количество корректно сжатых наборов данных (из 6)	Изменение среднего значения метрики Accuracy
t-SNE	6	0,0497
UMAP	6	0,0454
KernelPCA	5	-0,0942
PCA	3	-0,1789
AE	3	-0,2030
VAE	1	-0,3886

Анализ результатов предлагаемого подхода оценки алгоритмов сжатия в таблицах 2 и 3 позволяет сделать вывод, что при проецировании лучше всего сохраняют целевое значение алгоритмы UMAP и t-SNE. Предварительное сжатие позволяет немного улучшить результат. В работе [16] описано многостадийное сжатие,

но нет описания метода поэтапного сжатия UMAP\_и\_t-SNE и других приведенных в данной работе. Наилучший результат дает предварительное сжатие алгоритмом UMAP с последующим сжатием t-SNE.

Таблица 3

**Результаты работы алгоритмов с предварительным сжатием**

Table 3

**Results of algorithms with precompression**

Метод	Количество корректно сжатых наборов данных (из 3)	Изменение среднего Accuracy
UMAP_и_t-SNE	3	0,0044
UMAP_и_PCA	3	-0,0054
KernelPCA_и_UMAP	3	-0,0080
PCA_и_t-SNE	3	-0,0089
UMAP	3	-0,0122
t-SNE	3	-0,0149
KernelPCA_и_t-SNE	3	-0,0337
KernelPCA_и_PCA	3	-0,0886
KernelPCA_и_AE	3	-0,1090
UMAP_и_KernelPCA	2	-0,1156
PCA_и_UMAP	2	-0,1212
UMAP_и_AE	2	-0,1633
AE	2	-0,1904
KernelPCA	2	-0,2098
PCA_и_AE	2	-0,2099
KernelPCA_и_VAE	2	-0,2309
PCA_и_KernelPCA	2	-0,2392
UMAP_и_VAE	2	-0,2733
PCA_и_VAE	2	-0,2738
PCA	1	-0,3149
VAE	0	-0,4004

Результаты показывают, что в сравнении с современными аналогами (например, KernelPCA) и подходами в сжатии данных алгоритмы UMAP и t-SNE, а также предлагаемая методика двухэтапного сжатия (метод UMAP\_и\_t-SNE из таблицы 3) позволяют получить в среднем более высокие результаты прогноза целевых значений в исследуемых тестовых датасетах, в том числе и по сравнению с методом прогноза без сжатия, что является важным наблюдением, которое может позволить не только оптимизировать использование памяти и загрузки процессора при самом прогнозе целевых значений, но и поднять качество прогноза необходимых показателей исследуемых

объектов. Интуиция дальнейшего применения полученных результатов в задачах когнитивной визуализации многомерных данных заключается в том, что алгоритм сжатия, позволяющий получить более точный прогноз целевой метрики методами машинного обучения при построении когнитивного образа для визуальной аналитики экспертом, позволит выявить больше закономерностей и инсайтов в процессе взаимодействия эксперта и системы анализа.

Рассматривая полученные результаты применительно к разработанной системе анализа многослойных радиационно стойких защитных экранов, можно сделать вывод о том, что необходимо добавить возможность сжатия методом UMAP и t-SNE, показавшим наилучшие результаты, и сделать его сжатием по умолчанию, а также добавить сжатие методом UMAP с возможностью переключения между алгоритмами, чтобы дать возможность эксперту системы посмотреть на данные в разных сжатиях и выбрать наиболее подходящее с точки зрения эксперта представление.

## Заключение

В статье предложена методика сравнения алгоритмов сжатия для последующей визуализации многомерных данных. По предложенной методике было выполнено сравнение алгоритмов t-SNE, UMAP, PCA, KernelPCA, AE, VAE. Также исследована эффективность предварительного сжатия данных и двухэтапного применения алгоритмов снижения размерности.

При построении когнитивных образов для последующего визуального анализа пространства решений рекомендуется использовать алгоритм UMAP, поскольку он сохраняет целевое значение так же хорошо, как и t-SNE, то есть позволяет проводить анализ с большей точностью, чем другие алгоритмы, но требует меньших вычислительных затрат. Для анализа конфигураций с достаточно большим количеством параметров рекомендуется использовать алгоритм t-SNE с предварительным сжатием алгоритмом UMAP, который упрощает исходный набор данных, тем самым улучшая результат работы t-SNE.

## Список литературы

1. Терехов В.В., Михайличенко С.С. Методика определения толщины экрана для защиты микроэлектромеханических систем от воздействия высокоэнергичных ионов // Наукоемкие технологии и интеллектуальные системы: матер. Междунар. науч.-технич. конф. 2016. С. 356–358.
2. Mitsi E., Koutsomitis K., Apostolopoulos G. On the use of SRIM for calculating arc-dpa exposure. ArXiv, 2023, art. 2307.12867. URL: <https://arxiv.org/abs/2307.12867> (дата обращения: 12.11.2023).
3. Sytov A., Bandiera L., Cho K. et al. Geant4 simulation model of electromagnetic processes in oriented crystals for accelerator physics. J. Korean Phys. Soc., 2023, vol. 83, pp. 132–139 (дата обращения: 12.11.2023). doi: 10.1007/S40042-023-00834-6.
4. Шахнов В.А., Казаков В.В., Глушко А.А. и др. О возможности применения алгоритмов эволюционных вычислений при проектировании экранов радиационной защиты // Радиолокация и связь – перспективные технологии: XVII науч.-технич. конф. 2019. С. 87–90.
5. Казаков В.В. Генерация когнитивного представления информации о многослойных радиационно-стойких защитных экранах // ИС&ИТ: сб. тр. конф. 2022. Т. 1. С. 187–194.
6. Waggoner Ph.D. Modern Dimension Reduction (Elements in Quantitative and Computational Methods for the Social Sciences). Cambridge University Press, 2021, 98 p.
7. Meneghetti L., Demo N., Rozza G. A dimensionality reduction approach for convolutional neural networks. Appl. Intell., 2023, vol. 53, pp. 22818–22833. doi: 10.1007/s10489-023-04730-1.
8. Гергет О.М., Лаптев В.В. Разработка моделей искусственных нейронных сетей для вариационного автоэнкодера // МСИТ: матер. XVIII Междунар. науч.-технич. конф. 2021. С. 51–52.
9. Соснин А.С., Сулова И.А. Функции активации нейросети: сигмоида, линейная, ступенчатая, RELU, TANH // Наука. Информатизация. Технологии. Образование: матер. XII Междунар. науч.-практ. конф. 2019. С. 237–246.
10. Bilokon P., Finkelstein D. Iterated and exponentially weighted moving principal component analysis. SSRN, 2021. URL: <https://ssrn.com/abstract=3913940> (дата обращения: 12.11.2023). doi: 10.2139/ssrn.3913940.
11. Avella-Medina M., Davis R.A., Samorodnitsky G. Kernel PCA for multivariate extremes. ArXiv, 2022, art. 1207.3538. URL: <https://arxiv.org/pdf/2211.13172.pdf> (дата обращения: 12.11.2023).
12. Cai T.T., Ma R. Theoretical foundations of t-SNE for visualizing high-dimensional clustered data. JMLR, 2022, vol. 23, pp. 1–54.
13. Bonnici V. Kullback-Leibler divergence between quantum distributions, and its upper-bound. ArXiv, 2020, art. 2008.05932. URL: <https://arxiv.org/abs/2008.05932> (дата обращения: 12.11.2023).
14. Pain J.-C. Kullback-Leibler divergence for the Fréchet extreme-value distribution. ArXiv, 2023, art. 2303.13153. URL: <https://arxiv.org/pdf/2303.13153.pdf> (дата обращения: 12.11.2023).

15. McInnes L., Healy J., Saul N., Großberger L. UMAP: Uniform manifold approximation and projection. *J. of Open Source Software*, 2018, vol. 3, no. 29, art. 861. doi: 10.21105/joss.00861.

16. Dimple, Misra K., Arun K.G. Evidence for two distinct populations of Kilonova-associated gamma-ray bursts. *ApJL*, 2023, vol. 949, no. 2, art. L22. doi: 10.3847/2041-8213/acd4c4.

Software &amp; Systems

doi: 10.15827/0236-235X.142.178-185

2024, 37(2), pp. 178–185

### The problem of cognitive representation of the project space of multilayer radiation shields: Information compression algorithms

Ludmila A. Zinchenko <sup>1</sup>✉, Vadim V. Kazakov <sup>1</sup>, Boris V. Karyshev <sup>1</sup>

<sup>1</sup> Bauman Moscow State Technical University, Moscow, 105005, Russian Federation

#### For citation

Zinchenko, L.A., Kazakov, V.V., Karyshev, B.V. (2024) 'The problem of cognitive representation of the project space of multilayer radiation shields: Information compression algorithms', *Software & Systems*, 37(2), pp. 178–185 (in Russ.). doi: 10.15827/0236-235X.142.178-185

#### Article info

Received: 31.01.2023

After revision: 14.11.2023

Accepted: 21.11.2023

**Abstract.** The paper focuses on dimensional reduction algorithm application in the problems of designing multilayer radiation shields for protecting electronic equipment in outer space. The considered algorithms project the studied data on multilayer shields from high-dimensional space into low-dimensional space with preservation of data semantics, which allows visualizing large sets of high-dimensional information and simplifies user visual analysis, as well as application of some algorithms and approaches in an automated mode. The paper analyses the application of the following dimensionality reduction algorithms: principal component analysis (PCA), kernel principal component analysis (KernelPCA), stochastic neighbour embedding with t-distribution (t-SNE), uniform approximation and projection (UMAP), autoencoder (AE), variational autoencoder (VAE). In terms of neural network compression architectures, the paper presents network architectures used in computation and testing. Moreover, according to the proposed methodology, the authors of the paper investigate feasibility of combining several dimensionality reduction algorithms applied along the chain. Based on the conducted research the authors make a conclusion about the efficiency of the mentioned algorithms, as well as their combination for further processing or visualization. There is a brief description of the software that implements one of the proposed approaches in analyzing and processing information about multilayer radiation shields for electronic equipment used in outer space. On the basis of the conducted research it is recommended to use the UMAP algorithm. To analyze configurations with a sufficiently large number of parameters, it is recommended to use the t-SNE algorithm with precompression by the UMAP algorithm, which simplifies the initial data set, thus improving the result of t-SNE.

**Keywords:** cognitive visualization, data compression, dimensionality reduction, projection

#### References

1. Terekhov, V.V., Mikhaylichenko, S.S (2016) 'Method for modeling the thickness of shield for protection microelectromechanical systems from the impact of high-energy ions', *Proc. Int. Sci-Tech. Conf. Tech. & Systems*, pp. 356–358 (in Russ.).
2. Mitsi, E., Koutsomitis, K., Apostolopoulos, G. (2023) 'On the use of SRIM for calculating arc-dpa exposure', *ArXiv*, art. 2307.12867, available at: <https://arxiv.org/abs/2307.12867> (accessed November 12, 2023).
3. Sytov, A., Bandiera, L., Cho, K. et al. (2023) 'Geant4 simulation model of electromagnetic processes in oriented crystals for accelerator physics', *J. Korean Phys. Soc.*, 83, pp. 132–139. doi: 10.1007/S40042-023-00834-6. (accessed November 12, 2023).
4. Shakhnov, V.A., Kazakov, V.V., Glushko, A.A. et al. (2019) 'About possible application of evolutionary computing algorithms for designing radiation protection screens', *Proc. XVII Sci-Tech. Conf. Radar and Communications – Promising Tech.*, pp. 87–90 (in Russ.).
5. Kazakov, V.V. (2022) 'Generation of cognitive representation of information about multilayer radiation-resistant protective screens', *Proc. IS&IT*, 1, pp. 187–194 (in Russ.).
6. Waggoner, Ph.D. (2021) *Modern Dimension Reduction (Elements in Quantitative and Computational Methods for the Social Sciences)*. Cambridge University Press, 98 p.

7. Meneghetti, L., Demo, N., Rozza, G. (2023) 'A dimensionality reduction approach for convolutional neural networks', *Appl. Intell.*, 53, pp. 22818–22833. doi: 10.1007/s10489-023-04730-1.
8. Gerget, O.M., Laptev, V.V. (2021) 'Development of models of artificial neural networks of variational autoencoder', *Proc. Int. Sci-Tech. Conf. YMIT*, pp. 51–52 (in Russ.).
9. Sosnin, A.S., Suslova, I.A. (2019) 'Functions of neural net activation: SIGMOID, LINEAR, STEP, RELU, TANH', *Proc. Sci-Tech. Conf. The Sci. Informatization. Tech. Education*, pp. 237–246 (in Russ.).
10. Bilokon, P., Finkelstein, D. (2021) 'Iterated and exponentially weighted moving principal component analysis', *SSRN*, available at: <https://ssrn.com/abstract=3913940> (accessed November 12, 2023). doi: 10.2139/ssrn.3913940.
11. Avella-Medina, M., Davis, R.A., Samorodnitsky, G. (2022) 'Kernel PCA for multivariate extremes', *ArXiv*, art. 1207.3538, available at: <https://arxiv.org/pdf/2211.13172.pdf> (accessed November 12, 2023).
12. Cai, T.T., Ma, R. (2022) 'Theoretical foundations of T-SNE for visualizing high-dimensional clustered data', *JMLR*, 23, pp. 1–54.
13. Bonnici, V. (2020) 'Kullback-Leibler divergence between quantum distributions, and its upper-bound', *ArXiv*, art. 2008.05932, available at: <https://arxiv.org/abs/2008.05932> (accessed November 12, 2023).
14. Pain, J.-C. (2023) 'Kullback-Leibler divergence for the Fréchet extreme-value distribution', *ArXiv*, art. 2303.13153, available at: <https://arxiv.org/pdf/2303.13153.pdf> (accessed November 12, 2023).
15. McInnes, L., Healy, J., Saul, N., Großberger, L. (2018) 'UMAP: Uniform manifold approximation and projection', *J. of Open Source Software*, 3(29), art. 861. doi: 10.21105/joss.00861.
16. Dimple, Misra, K., Arun, K.G. (2023) 'Evidence for two distinct populations of Kilonova-associated gamma-ray bursts', *ApJL*, 949(2), art. L22. doi: 10.3847/2041-8213/acd4c4.

**Авторы**

**Зинченко Людмила Анатольевна**<sup>1</sup>, д.т.н.,  
профессор, [lyudmillaa@mail.ru](mailto:lyudmillaa@mail.ru)  
**Казakov Вадим Вячеславович**<sup>1</sup>,  
аспирант, [kazakov-vadim-k@yandex.ru](mailto:kazakov-vadim-k@yandex.ru)  
**Карышев Борис Владимирович**<sup>1</sup>,  
студент, [boris.karyshev@gmail.com](mailto:boris.karyshev@gmail.com)

<sup>1</sup> МГТУ им. Н.Э. Баумана,  
г. Москва, 105005, Россия

**Authors**

**Ludmila A. Zinchenko**<sup>1</sup>, Dr.Sc. (Engineering),  
Professor, [lyudmillaa@mail.ru](mailto:lyudmillaa@mail.ru)  
**Vadim V. Kazakov**<sup>1</sup>, Postgraduate Student,  
[kazakov-vadim-k@yandex.ru](mailto:kazakov-vadim-k@yandex.ru)  
**Boris V. Karyshev**<sup>1</sup>, Student,  
[boris.karyshev@gmail.com](mailto:boris.karyshev@gmail.com)

<sup>1</sup> Bauman Moscow State Technical University,  
Moscow, 105005, Russian Federation

## Сегментация файлов неисполняемых форматов для выявления угроз нарушения информационной безопасности, реализуемых в форме эксплоитов

А.Н. Архипов<sup>1</sup>✉, С.Е. Кондаков<sup>1</sup>

<sup>1</sup> Московский государственный технический университет им. Н.Э. Баумана, г. Москва, 105005, Россия

### Ссылка для цитирования

Архипов А.Н., Кондаков С.Е. Сегментация файлов неисполняемых форматов для выявления угроз нарушения информационной безопасности, реализуемых в форме эксплоитов // Программные продукты и системы. 2024. Т. 37. № 2. С. 186–192. doi: 10.15827/0236-235X.142.186-192

### Информация о статье

Группа специальностей ВАК: 2.3.6

Поступила в редакцию: 11.12.2023

После доработки: 16.02.2024

Принята к публикации: 21.02.2024

**Аннотация.** В статье описана прикладная задача сегментации файлов неисполняемых форматов с целью выявления угроз нарушения информационной безопасности, реализуемых в форме эксплоитов (вредоносного кода). Актуальность исследования обусловлена необходимостью решения научной задачи выявления эксплоитов, в том числе тех, для которых применялись технологии обфускации, и повышения эффективности подсистемы антивирусной защиты информации за счет повышения чувствительности и специфичности выявления эксплоитов. Цель авторов исследования – создание алгоритма сегментации файлов неисполняемых форматов, позволяющего представить их в виде блоков (фрагментов), обеспечивающих максимальную вероятность вхождения в их состав элементов эксплоита. Сегментация файлов неисполняемых форматов применяется для обеспечения возможности последующего углубленного анализа не всего файла целиком, а его фрагментов на наличие вредоносного кода. Предметом исследования является множество методов, методик, моделей, алгоритмов сегментации файлов неисполняемых форматов с целью выявления угроз нарушения информационной безопасности, реализуемых в форме эксплоитов (вредоносного кода). В работе применяются научные методы анализа, измерения и сравнения. Авторами данной статьи разработан алгоритм сегментации, который позволяет представить файл неисполняемого формата в виде блоков оптимальных размеров, необходимых для выявления в их составе элементов эксплоита. Алгоритм базируется на математической модели эксплоита, внедренного в файл неисполняемого формата, разработанной авторами и математически описывающей структуру, составные элементы и показатели, характеризующие его. Предлагаемый алгоритм может использоваться для создания новых методов, методик, моделей, алгоритмов и средств, направленных на повышение эффективности защиты информации от воздействия вредоносного кода, распространяемого в форме эксплоитов, в том числе созданных с применением технологий обфускации (запутывания) программного кода.

**Ключевые слова:** компьютерная атака, система защиты информации, выявление эксплоитов, алгоритм сегментации, технологии обфускации, антивирусная защита информации, компьютерный вирус, вредоносный код

**Введение.** Выявление угроз нарушения информационной безопасности, реализуемых посредством эксплоитов, является одним из приоритетных направлений, определенных в рамках функционирования и развития государственной системы обнаружения, предупреждения и ликвидации последствий компьютерных атак, направленных на информационные ресурсы Российской Федерации [1].

Решение указанной задачи достигается за счет совершенствования и разработки новых методов, алгоритмов и средств анализа файлов неисполняемых форматов, потенциально пригодных для внедрения в их состав эксплоитов.

В ряде литературных источников, посвященных данной тематике, одним из основных этапов существующей методологии выявления эксплоитов в файлах неисполняемых форматов выделяется сегментация, которая осуществля-

ется для возможности последующего углубленного анализа не всего файла целиком, а его фрагментов на наличие вредоносного кода. Анализируемый файл представляется в виде конечного множества блоков (фрагментов) определенной длины.

В научных работах других исследователей условно можно выделить два основных подхода к решению данной задачи. Первый предполагает сегментацию файла на его структурные элементы в соответствии с форматом [2–4], второй, значительно чаще применяемый для решения задачи обнаружения вредоносного кода в файлах неисполняемых форматов как наиболее эффективный, – применение метода скользящего окна [5–7].

К основному недостатку обоих подходов можно отнести неоптимальность размеров получаемых сегментов (блоков), которые могут

значительно отличаться от искомым в их составе элементов эксплоита, что значительно снижает эффективность обнаружения.

Если при первом подходе размеры сегментов определяются структурой файла в соответствии со спецификацией на формат, то параметры скользящего окна подбираются либо экспертным методом, либо в рамках обучения модели бинарной классификации с использованием технологий машинного обучения и искусственного интеллекта и не обосновываются математически, например, на основе математического моделирования эксплоита и его составных частей [8, 9].

Также не учитывается специфика конкретного формата файла – его структурные особенности и предусмотренные методы преобразования данных (смена кодировки, сжатие и др.).

Кроме того, не рассматривается возможность злоумышленника специально преобразовать элементы эксплоита к виду, сохраняющему заложенную функциональность, но затрудняющему анализ, – обфускации.

Названные аспекты могут значительно снизить эффективность выявления угроз нарушения информационной безопасности, реализуемых посредством эксплоитов, в файлах исполняемых форматов [10].

Указанные недостатки обуславливают актуальность разработки новых и/или совершенствование существующих методов и алгоритмов сегментации.

В частности, в научной и практической литературе не представлен алгоритм сегментации, учитывающий строение формата исполняемого файла, применение технологий обфускации и одновременно позволяющий осуществлять деление файла на сегменты математически обоснованной длины, обеспечивающей максимальную вероятность вхождения в их состав элементов эксплоита и минимизацию вероятности вхождения иных данных. Разработка этого алгоритма формирует результат, обладающий научной новизной.

### Постановка задачи

Дано:  $X$  – множество файлов исполняемых форматов с внедренными обфусцированными эксплоитами (вредоносные файлы) и без таковых (чистые файлы).

Вербальная постановка научной задачи: разработать алгоритм деления файлов исполняемого формата на сегменты длины, позволяющей обеспечить максимизацию показателей эффективности сегментации.

Формальная постановка научной задачи: разработать такой алгоритм  $A_c$ , что

$$X \xrightarrow{A_c} \{b_1^{z_j}, b_2^{z_j}, \dots, b_k^{z_j}\} : \max(w_q),$$

$$q = 1, \dots, n, \quad j = 1, \dots, m,$$

где  $\{b_1^{z_j}, b_2^{z_j}, \dots, b_k^{z_j}\}$  – множество сегментов, получаемых по результатам сегментации;  $z_j$  – длина сегментов;  $w_q$  – показатели эффективности сегментации;  $n$  – количество показателей эффективности сегментации;  $m$  – количество значений длины сегментов.

### Выбор показателей и критериев их оценки

Эксплоит структурно состоит из двух частей (модуля эксплуатации уязвимости и полезной нагрузки), которые обладают разными численными характеристиками [8, 9], отражающими их длины. Поэтому целесообразно разделить файл на сегменты размера  $z_1$  и  $z_2$ , где сегменты длины  $z_1$  будут применяться для потенциального поиска содержимого модуля эксплуатации уязвимости, а  $z_2$  – полезной нагрузки эксплоита.

Одновременно, учитывая, что по результатам сегментации необходимо получить сегменты длины, обеспечивающей максимальную вероятность вхождения в их состав элементов эксплоита и минимизацию вероятности вхождения иных данных, в качестве показателей эффективности сегментации определим следующие:

$w_1$  – вероятность события, при котором в сегмент  $b_i$  длины  $z_1$  по результатам сегментации войдет только содержимое модуля эксплуатации эксплоита;

$w_2$  – вероятность события, при котором в сегмент  $b_i$  длины  $z_2$  по результатам сегментации войдет только содержимое полезной нагрузки эксплоита.

Указанные показатели могут быть найдены по формулам

$$w_1 = \frac{L_M}{z_1}, \quad w_2 = \frac{L_S}{z_2}, \quad (1)$$

где  $z_1, z_2$  – размеры анализируемых сегментов в байтах;  $L_S$  – размер модуля полезной нагрузки в байтах;  $L_M$  – размер модуля эксплуатации в байтах.

При идеальных значениях длин сегментов, получаемых при сегментации, показатели  $w_1$  и  $w_2$  будут равны 1, но в реальных условиях достижение указанных значений показателей маловероятно. Поэтому в качестве критерия

оценки эффективности показателей сегментации определим требования, заданные в постановке задачи:

$$\max(w_1) \rightarrow 1, \max(w_2) \rightarrow 1. \quad (2)$$

### Разработка алгоритма сегментации

Учитывая особенности построения эксплоита, отраженные в его математической модели, а именно нахождение элементов эксплоита в его структурообразующих блоках, на первом этапе необходимо разделить исследуемый файл на формирующие структурные элементы, используемые для его построения в соответствии со спецификацией на формат.

Таким образом, на первом этапе  $A_c^1$  сегментации исследуемый файл представляется в виде конечного множества его структурных элементов в соответствии со спецификацией на конкретный формат:

$$X \xrightarrow{A_c^1} \{s_1, s_2, \dots, s_p\}, \quad (3)$$

где  $s_p$  – структурные элементы сегментируемого файла;  $p$  – количество таких элементов.

На втором этапе предлагаемого алгоритма структурные элементы исследуемого файла  $s_p$  разделяем методом скользящего окна на сегменты длиной  $z_1$  и  $z_2$ , где сегменты длины  $z_1$  будут для потенциального поиска содержимого модуля эксплуатации уязвимости, а  $z_2$  – полезной нагрузки эксплоита.

С учетом изложенного на втором этапе  $A_c^2$  каждый из элементов  $s_p$ , полученных на предыдущем этапе (3), будет сегментирован методом скользящего окна в соответствии с выражением

$$s_p \xrightarrow{A_c^2} \begin{cases} \sum_{j=1}^a b_j^{z_1}, \\ \sum_{i=1}^e b_j^{z_2}, \end{cases} \quad (4)$$

где  $a$  и  $e$  – число сегментов длины  $z_1$  и  $z_2$  соответственно.

При этом их значения могут быть получены по следующим формулам:

$$a = (L_F - z_1) + 1, e = (L_F - z_2) + 1, \quad (5)$$

где  $L_F$  – размер анализируемого файла в байтах.

Пусть числовые значения размера модуля полезной нагрузки и модуля эксплуатации в байтах для исследуемого множества файлов неисполняемых форматов с внедренными эксплоитами (вредоносные файлы) определяются неизвестной числовой функцией, определенной на множестве элементарных исходов. Тогда предсказать заранее, какое из своих значе-

ний она примет, как правило, невозможно, можно лишь указать вероятность, с которой будет принято то или иное значение, или вероятность того, что ее значения будут находиться в каком-либо числовом промежутке.

Тогда числовые значения размера модуля полезной нагрузки  $L_S$  и модуля эксплуатации  $L_M$  можно рассматривать как случайные величины, так как они удовлетворяют ее классическому определению [11]: заданы вероятностное пространство  $(\Omega, F, P)$ , числовая функция  $\xi(\omega)$  определена для всех  $\omega \in \Omega$ , если для любого числа  $c$  выполняется условие  $\{\omega \in \Omega: \xi(\omega) \leq c\} \in F$ .

С учетом выражений (1) и (2) значения длины сегментов в байтах должны удовлетворять условию

$$z_1 \approx L_S, z_2 \approx L_M. \quad (6)$$

Случайная величина будет полностью описана с вероятностной точки зрения, если задать ее распределение, то есть точно указать, какой вероятностью обладает каждое из возможных событий. Этим будет установлен закон распределения случайной величины [11].

Поскольку реальный закон распределения неизвестен, получим его приближенные значения эмпирически.

Для этого сформируем коллекцию файлов неисполняемых форматов, содержащих обфусцированные эксплоиты, из образцов, размещенных в открытом доступе (<https://bazaar.abuse.ch/>) (вредоносные файлы), и без таковых (чистые файлы), сгенерированные в автоматическом режиме с использованием штатных шаблонов приложений Microsoft Office.

В качестве обфускаторов использованы свободно доступные на хостинге IT-проектов Github реализации, подходящие для автоматизированного запутывания большого количества исходных кодов.

Для подтверждения наличия/отсутствия вредоносного кода в сформированной коллекции применялись средства (технологии) антивирусной защиты информации, размещенные в открытом доступе (<https://virustotal.com/>).

На базе полученной коллекции эксплоитов (2 500 шт.) проведем анализ значений размеров модуля полезной нагрузки  $L_S$  и модуля эксплуатации  $L_M$  и запишем их в виде рядов распределений.

Построенные гистограммы плотности вероятности случайных величин (<http://www.swsys.ru/uploaded/image/2024-2/9.jpg>), отображающих размеры упомянутых элементов эксплоита, позволяют выдвинуть предположение о нормальности распределений данных случайных величин.

Для подтверждения данного предположения воспользуемся специализированными статистическими методами Колмогорова–Смирнова и Шапиро–Уилка [12, 13].

По результатам выполнения указанных статистических тестов получены следующие значения контрольных критериев:  $D_1 = 0.01662$ ,  $D_2 = 0.01601$ ,  $W_1 = 0.9992$ ,  $W_2 = 0.9987$ .

Отличия полученной эмпирически и эталонной (нормальной) функций распределений при их проверке статистическими методами Колмогорова–Смирнова и Шапиро–Уилка можно увидеть на построенных графиках (<http://www.swsys.ru/uploaded/image/2024-2/10.jpg>).

На основании полученных результатов можно сделать вывод о том, что случайные величины значений размера модуля полезной нагрузки  $L_S$  и модуля эксплуатации  $L_M$  являются нормально распределенными с уровнем значимости, равным 0.05.

Поэтому с учетом критериев (2) значения для параметров  $z_1$  и  $z_2$  будут следующие:

$$z_1 \approx L_S \approx E_S, \quad z_2 \approx L_M \approx E_M, \quad (7)$$

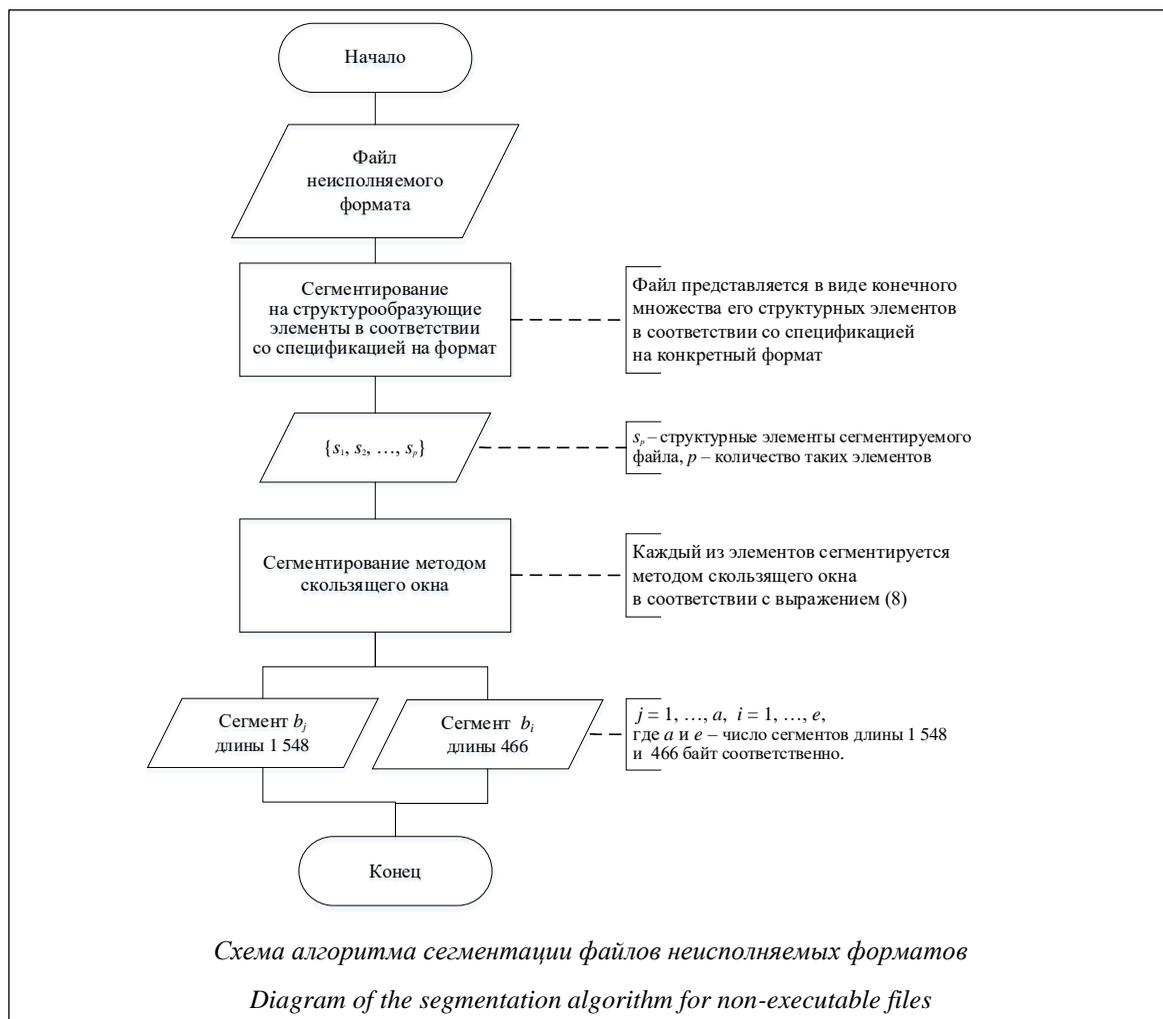
где  $E_S$  и  $E_M$  – эмпирические математические ожидания значений размера модуля полезной нагрузки  $L_S$  и модуля эксплуатации  $L_M$  соответственно.

Таким образом, сегментацию структурных элементов исследуемого файла  $s_p$  будем осуществлять на блоки размера  $z_1 = 1\,548$  байт и  $z_2 = 466$ .

Обобщая полученные результаты, запишем алгоритм сегментации файлов неисполняемых форматов с указанием конкретных параметров сегментирования с учетом выражений (3)–(5) и (7):

$$X \xrightarrow{A_c^1} \{s_1, s_2, \dots, s_p\} \xrightarrow{A_c^2} \begin{cases} (L_F^{-1548} + 1) \sum_{j=1}^{L_F^{-1548} + 1} b_j^{1548}, \\ (L_F^{-466} + 1) \sum_{i=1}^{L_F^{-466} + 1} b_i^{466}. \end{cases} \quad (8)$$

На рисунке представлена схема алгоритма сегментации файлов неисполняемых форматов в задаче выявления эксплоитов.



### Результаты исследования эффективности применения предлагаемого алгоритма

С формальной точки зрения выявление угроз нарушения информационной безопасности, реализуемых посредством эксплоитов, сводится к решению научной задачи бинарной классификации файлов неисполняемых форматов на безопасные и вредоносные.

С учетом изложенного в качестве метода оценки эффективности бинарной классификации выберем ROC-анализ [14–16].

В качестве показателя оценки эффективности бинарной классификации определим

$$q = \frac{q_1 + q_2}{2}, \tag{9}$$

где  $q_1$  – чувствительность,  $q_2$  – специфичность, которые вычисляются по формулам

$$q_1 = \frac{TP}{TP + FN}, q_2 = \frac{TN}{TN + FP},$$

где  $TP$  – число верно классифицированных вредоносных объектов;  $FN$  – число объектов, классифицированных как отрицательные (ошибка I рода);  $TN$  – число верно классифицированных безопасных объектов;  $FP$  – число объектов, классифицированных как положительные (ошибка II рода).

В качестве критерия оценки эффективности используем условие, предусмотренное в методе:

$$\max(q) \rightarrow 1.$$

Описанный алгоритм сегментации реализован в программе для ЭВМ AntigenExploits (Свидетельство о госрегистрации № 2023687464), разработанной авторами в качестве эмпирической реализации методики выявления угроз нарушения информационной безопасности, реализуемых посредством эксплоитов.

Исследования эффективности применения предлагаемого алгоритма проведены с использованием указанной программы.

Эксперименты проводились в следующем порядке.

1. Подготовка тестовой выборки из 2 000 файлов неисполняемых форматов, не используемых при разработке алгоритма сегментации, содержащей файлы с внедренными эксплоитами (вредоносные файлы – 1 000 образцов) и без таковых (безопасные файлы – 1 000 образцов).

2. Анализ тестовой выборки с помощью существующих отечественных средств антивирусной защиты информации, включенных в ре-

естр российского программного обеспечения, а также в программы AntigenExploits.

3. Анализ результатов экспериментального исследования (см. таблицу).

### Результаты экспериментального исследования

#### Experimental study results

Средство антивирусной защиты информации	Значение критерия $q$ (9)
Программа AntigenExploits	0.97
Kaspersky Standard	0.9
Dr.Web Антивирус	0.70
NANO Антивирус Pro	0.67
Антивирус VR Protect для Linux	0.69

Результаты оценки показали, что алгоритм сегментации, предлагаемый авторами, в конкретном исследовании позволил повысить эффективность выявления угроз нарушения информационной безопасности, реализуемых посредством обфусцированных эксплоитов, по сравнению с существующими средствами антивирусной защиты информации более чем на 7 %.

### Заключение

В статье представлен алгоритм сегментации файлов неисполняемых форматов, применимый в задаче выявления эксплоитов, определяющий порядок, применяемые методы и численные параметры сегментации, обоснованные математически.

Алгоритм является универсальным и позволяет разбивать любые файлы неисполняемых форматов на сегменты наилучшей длины, обеспечивающей максимальную вероятность вхождения в их состав элементов эксплоита и минимизацию вероятности вхождения иных данных.

При этом сегментирование файла неисполняемого формата предложенным алгоритмом применимо для выявления эксплоитов, подвергшихся процедуре обфускации.

Указанные аспекты отличают представленный алгоритм сегментации от существующих аналогов.

В качестве направления для дальнейших исследований в данной области целесообразно рассмотреть возможность оптимизации шага (сдвига по фазе) предложенного алгоритма сегментации.

## Список литературы

1. Ланецкая А.Ю., Александрова Е.Н. Современные угрозы информационной безопасности // Междунар. журнал гуманитарных и естественных наук. 2022. № 7-2. С. 192–195.
2. Zhou X., Pang J. Expdf: Exploits detection system based on machine-learning. *Int. J. of Computational Intelligence*, 2019, vol. 12, no. 2, pp. 1019–1028. doi: 10.2991/ijcis.d.190905.001.
3. Falah A., Pan L., Abdelrazek M., Doss R. Identifying drawbacks in malicious PDF detectors. In: *CCIS. Proc. FNSS*, 2019, vol. 878, pp. 128–139. doi: 10.1007/978-3-319-94421-0\_10.
4. Kumar R., Geetha S. Malware classification using XGboost-gradient boosted decision tree. *ASTES J.*, 2020, vol. 5, no. 5, pp. 536–549. doi: 10.25046/aj050566.
5. Yousefi-Azar M., Varadharajan V., Hamey L., Chen S. Mutual information and feature importance gradient boosting: Automatic byte n-gram feature reranking for Android malware detection. *Software: Practice and Experience*, 2021, vol. 51, no. 7, pp. 1518–1539. doi: 10.1002/spe.2971.
6. Jeong Y.S., Woo J., Kang A.R. Malware detection on byte streams of hangul word processor files. *Appl. Sci.*, 2019, vol. 9, no. 23, art. 5178. doi: 10.3390/app9235178.
7. Kang A.R., Jeong Y.S., Kim S.L., Woo J. Malicious PDF detection model against adversarial attack built from benign PDF containing JavaScript. *Appl. Sci.*, 2019, vol. 9, no. 22, art. 4764. doi: 10.3390/app9224764.
8. Кондаков С.Е., Архипов А.Н. Математическая модель эксплоита, внедренного в файл неисполняемого формата // Изв. ИИФ. 2023. Т. 69. № 3. С. 93–96.
9. Архипов А.Н., Кондаков С.Е. Разработка математической модели эксплоита, внедренного в файл неисполняемого формата, с учетом полезной нагрузки // Приборы. 2023. № 11. С. 39–46.
10. Архипов А.Н., Пиков В.А., Кабаков В.В. Порядок и результаты экспериментальных исследований влияния обфускации на качество выявления угроз информационной безопасности, реализуемых посредством эксплоитов, в файлах неисполняемых форматов // Вопросы защиты информации. 2023. № 2. С. 32–37.
11. Феллер В. Введение в теорию вероятностей и ее приложения; [пер. с англ.]. М.: Мир, 1984. Т. 1. 528 с.
12. Леман Э. Проверка статистических гипотез; [пер. с англ.]. М.: Наука, 1979. 408 с.
13. Shapiro S.S., Wilk M.B. An analysis of variance test for normality (complete samples). *Biometrika*, 1965, vol. 52, no. 3-4, pp. 591–611. doi: 10.1093/BIOMET/52.3-4.591.
14. Брюс П., Брюс Э. Практическая статистика для специалистов Data Science; [пер. с англ.]. СПб: БХВ-Петербург, 2018. 304 с.
15. Старовойтов В.В., Голуб Ю.И. Сравнительный анализ оценок качества бинарной классификации // Информатика. 2020. Т. 17. № 1. С. 87–101. doi: 10.37661/1816-0301-2020-17-1-87-101.
16. Vujovic Ž.Đ. Classification model evaluation metrics. *IJACSA*, 2021, vol. 12, no. 6, pp. 599–606. doi: 10.14569/IJACSA.2021.0120670.

**Algorithm for segmentation of non-executable files in terms of exploit detection**Alexander N. Arkhipov <sup>1</sup>✉, Sergey E. Kondakov <sup>1</sup><sup>1</sup> Bauman Moscow State Technical University, Moscow, 105005, Russian Federation**For citation**Arkhipov, A.N., Kondakov, S.E. (2024) 'Algorithm for segmentation of non-executable file formats in terms of exploit detection', *Software & Systems*, 37(2), pp. 186–192 (in Russ.). doi: 10.15827/0236-235X.142.186-192**Article info**

Received: 11.12.2023

After revision: 16.02.2024

Accepted: 21.02.2024

**Abstract.** The paper solves the applied task of segmenting non-executable files in order to identify information security threats implemented in the form of exploits (malicious code). The relevance of the study is due to the need to solve the scientific problem of detecting exploits, including those undergoing obfuscation technologies, as well as and to increase the effectiveness of the anti-virus information protection subsystem by increasing the sensitivity and specificity of detecting exploits. The aim of the work is to develop an algorithm for segmenting non-executable files, which allows presenting them in the form of blocks (fragments) that ensure the maximum probability entering exploit elements into their composition. Segmentation of non-executable files enables subsequent in-depth analysis of not the entire file, but its fragments for the malicious code. The subject of the study is a variety of methods, techniques, models, algorithms for segmenting non-executable files in order to identify information security threats implemented in the form of exploits (malicious code). The research uses scientific methods of analysis, measurement and comparison. The authors of the paper have developed a

segmentation algorithm that allows presenting a non-executable file in the form of blocks (fragments) of optimal sizes necessary to identify exploit elements in their composition. The proposed algorithm is based on an exploit mathematical model embedded in a non-executable file. The model was developed by the authors and mathematically describes the structure, constituent elements and indicators that characterize the algorithm. The proposed algorithm can be used to create new methods, techniques, models, algorithms and tools aimed at improving the effectiveness of protecting information from the effects of malicious code distributed in the form of exploits, including those created using program code obfuscation (obfuscation) technologies.

**Keywords:** computer attack, information protection system, exploit detection, segmentation algorithm, obfuscation technologies, anti-virus information protection, computer virus, malicious code

## References

1. Lanetskaya, A.Yu., Aleksandrova, E.N. (2022) 'Modern threats to information security', *Int. J. of Humanities and Natural Sci.*, (7-2), pp. 192–195 (in Russ.).
2. Zhou, X., Pang, J. (2019) 'Expdf: Exploits detection system based on machine-learning', *Int. J. of Computational Intelligence*, 12(2), pp. 1019–1028. doi: 10.2991/ijcis.d.190905.001.
3. Falah, A., Pan, L., Abdelrazek, M., Doss, R. (2019) 'Identifying drawbacks in malicious PDF detectors', in *CCIS. Proc. FNSS*, 878, pp. 128–139. doi: 10.1007/978-3-319-94421-0\_10.
4. Kumar, R., Geetha, S. (2020) 'Malware classification using XGboost-gradient boosted decision tree', *ASTES J.*, 5(5), pp. 536–549. doi: 10.25046/aj050566.
5. Yousefi-Azar, M., Varadharajan, V., Hamey, L., Chen, S. (2021) 'Mutual information and feature importance gradient boosting: Automatic byte n-gram feature reranking for Android malware detection', *Software: Practice and Experience*, 51(7), pp. 1518–1539. doi: 10.1002/spe.2971.
6. Jeong, Y.S., Woo, J., Kang, A.R. (2019) 'Malware detection on byte streams of hangul word processor files', *Appl. Sci.*, 9(23), art. 5178. doi: 10.3390/app9235178.
7. Kang, A.R., Jeong, Y.S., Kim, S.L., Woo, J. (2019) 'Malicious PDF detection model against adversarial attack built from benign PDF containing JavaScript', *Appl. Sci.*, 9(22), art. 4764. doi: 10.3390/app9224764.
8. Kondakov, S.E., Arkhipov, A.N. (2023) 'Mathematical model of an exploit embedded in a nonexecutable file', *Proc. of IEP*, 69(3), pp. 93–96 (in Russ.).
9. Arkhipov, A.N., Kondakov, S.E. (2023) 'Development of a mathematical model of an exploit embedded in a non-executable file format, taking into account the payload', *Instruments*, (11), pp. 39–46 (in Russ.).
10. Arkhipov, A.N., Pikov, V.A., Kabakov, V.V. (2023) 'The order and results of experimental studies of the influence of obfuscation on the quality of detecting information security threats implemented through exploits in files of non-executable formats', *Information Security Questions*, (2), pp. 32–37 (in Russ.).
11. Feller, W. (1957) *An Introduction to Probability Theory and its Applications*, NY: John Wiley & Sons, 526 p. (Russ. ed.: (1984) Moscow, 528 p.).
12. Lehmann, E.L. (1959) *Testing Statistical Hypotheses*, NY: John Wiley & Sons, 388 p. (Russ. ed.: (1979) Moscow, 408 p.).
13. Shapiro, S.S., Wilk, M.B. (1965) 'An Analysis of Variance Test for Normality (Complete Samples)', *Biometrika*, 52(3/4), pp. 591–611. doi: 10.1093/BIOMET/52.3-4.591.
14. Bruce, P., Bruce, E. (2017) *Practical Statistics for Data Scientists*, CA: O'Reilly Media Publ., 317 p. (Russ. ed.: (2018) St. Petersburg, 304 p.).
15. Starovoytov, V.V., Golub, Yu.I. (2020) 'Comparative study of quality estimation of binary classification', *Informatcs*, 17(1), pp. 87–101 (in Russ.). doi: 10.37661/1816-0301-2020-17-1-87-101.
16. Vujovic, Ž.Đ. (2021) 'Classification model evaluation metrics', *IJACSA*, 12(6), pp. 599–606. doi: 10.14569/IJACSA.2021.0120670.

## Авторы

**Архипов Александр Николаевич**<sup>1</sup>,  
ассистент кафедры, diskpart111@mail.ru  
**Кондаков Сергей Евгеньевич**<sup>1</sup>,  
к.т.н., доцент, sergeikondakov@list.ru

## Authors

**Alexander N. Arkhipov**<sup>1</sup>,  
Teaching Assistant, diskpart111@mail.ru  
**Sergey E. Kondakov**<sup>1</sup>, Cand. of Sci. (Engineering),  
Associate Professor, sergeikondakov@list.ru

<sup>1</sup> Московский государственный  
технический университет им. Н.Э. Баумана,  
г. Москва, 105005, Россия

<sup>1</sup> Bauman Moscow State  
Technical University,  
Moscow, 105005, Russian Federation

## Реализация программного комплекса психодиагностики с обратной связью на основе компьютерного зрения

А.В. Ивашенко<sup>1</sup>✉, М.В. Александрова<sup>2</sup>, Д.С. Жейков<sup>1</sup>,  
Е.В. Мазанкина<sup>1</sup>, Е.В. Захарова<sup>1</sup>, А.В. Колсанов<sup>1</sup>

<sup>1</sup> Самарский государственный медицинский университет, г. Самара, 443099, Россия

<sup>2</sup> Компания «Открытый код», г. Самара, 443001, Россия

### Ссылка для цитирования

Ивашенко А.В., Александрова М.В., Жейков Д.С., Мазанкина Е.В., Захарова Е.В., Колсанов А.В. Реализация программного комплекса психодиагностики с обратной связью на основе компьютерного зрения // Программные продукты и системы. 2024. Т. 37. № 2. С. 193–200. doi: 10.15827/0236-235X.142.193-200

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 21.09.2023

После доработки: 09.01.2024

Принята к публикации: 18.03.2024

**Аннотация.** В статье представлен программный комплекс психодиагностики, позволяющий учитывать вовлеченность респондентов при прохождении тестов. Предметом исследования является управление глубиной погружения пользователей интерактивных пользовательских интерфейсов. Предлагается формальная модель иммерсивной среды, предназначенная для реализации системы управления погружением пользователя в процессы человеко-компьютерного взаимодействия путем мониторинга и контроля выполняемых действий в ответ на возникающие события. В качестве таких событий могут рассматриваться воздействия на пользователя в реальном мире, а также аудиовизуальные стимулы, формируемые в иммерсивной среде автоматически. Ответное поведение должно включать цепочки действий, автоматически фиксируемые средствами компьютерного зрения и окулографии. Разработанный программный комплекс психологической диагностики и медицинской реабилитации позволяет в отличие от аналогов реализовать обратную связь по мониторингу и контролю вовлеченности пациентов. Для этого в состав системы входит подсистема компьютерного зрения, включающая видеокамеру и программное обеспечение видеоконтроля двигательной активности пациента. Программное обеспечение отслеживает общий объем мимических движений, движений головы, а также идентифицирует текущее эмоциональное состояние с помощью искусственной нейронной сети. Для обеспечения высокой вовлеченности пользователя предусмотрена персонализация пользовательского интерфейса и тестов. Практическая значимость работы состоит в реализации мониторинга вовлеченности пациентов при прохождении психологического тестирования и использования этой информации для персонализации медицинских услуг. Анализ вовлеченности пользователей программного комплекса психологической диагностики позволяет дополнить результаты психологического тестирования и адаптировать последовательность и содержание тестов, сохраняя интерес пользователя и снижая влияние внешних раздражителей.

**Ключевые слова:** человеко-компьютерное взаимодействие, автоматизация психологического тестирования, пользовательские интерфейсы, иммерсивная реальность, акцентная визуализация

**Введение.** Автоматизация психологической диагностики является важной задачей цифровой трансформации здравоохранения. Реализация современных интерактивных пользовательских интерфейсов в составе ПО психологического тестирования позволяет сократить время, необходимое для полноценного психодиагностического обследования, и повысить достоверность установления и измерения индивидуально-психологических отличий пациентов. При этом на практике необходимо реализовать персонализированный подход, подбирая последовательность и содержание тестов, а также способ их визуализации в пользовательском интерфейсе с учетом индивидуальных особенностей восприятия.

Современные технологии программной реализации человеко-компьютерного взаимодействия предоставляют широкие возможности по персонализации пользовательских интерфейсов. Однако единого критерия, позволяющего

интегрально оценивать вовлеченность пользователя и строить обратную связь в системе управления пользовательским интерфейсом, до настоящего времени нет. Для решения этой проблемы в данной статье предлагается модель иммерсивной среды с управляемой глубиной погружения, на основе которой создана система мониторинга вовлеченности пользователя и управления ею.

Анализ вовлеченности пользователей программного комплекса психологической диагностики позволяет дополнить результаты психологического тестирования и адаптировать последовательность и содержание тестов, сохраняя интерес пользователя и снижая влияние внешних раздражителей.

### Обзор существующих подходов

Интерактивные технологии человеко-компьютерного взаимодействия широко применя-

ются в психологическом тестировании [1–3]. Программный инструментарий обычно включает онлайн-тестирование, заполнение опросных листов, дистанционные собеседования, выполнение различных упражнений в игровой форме и т.п. Прохождение тестов реализуется с помощью графических пользовательских интерфейсов [4], получивших широкое распространение и не требующих дополнительного обучения пользователей.

Анализ результатов тестирования может быть проведен с использованием технологий искусственного интеллекта [5, 6]. Известны разные позиции относительно объективности такого анализа, однако развитие искусственных нейронных сетей открывает все новые возможности. В свою очередь, применение технологий виртуальной реальности [7] позволяет повысить привлекательность систем психологического тестирования и обеспечить высокую вовлеченность пользователей.

Персонализация интерфейса как средство поддержки и реабилитации показывает высокую эффективность в современных приложениях [8]. Применение искусственного интеллекта обеспечивает достаточную гибкость и адекватность алгоритмов конфигурирования и персонализации пользовательского интерфейса [9], что расширяет возможности современных программных комплексов. В качестве технических средств персонализации довольно полезно применение систем окулографии (айтрекинга) [10, 11]. В этом случае формирование иммерсивной среды производится с помощью графического пользовательского интерфейса, а отслеживание реакции пользователя – посредством айтрекера или видеокамеры.

Применение интеллектуальных технологий компьютерного зрения в медицине повышает точность диагностики и снижает нагрузку на медицинский персонал [12]. Технологии анализа видео- и машинного обучения позволяют собирать данные об эмоциях пользователей компьютерной техники и использовать эту информацию в дальнейшем для изучения влияния человеческого фактора [13]. Автоматизация сбора и обработки данных с использованием облачных технологий повышает достоверность психологической диагностики [14], а контроль обратной связи [15, 16] сокращает время на психологическое тестирование за счет более адекватного подбора тестов.

Субъективная природа восприятия требует реализации персонифицированных пользовательских интерфейсов, способных адаптиро-

ваться к индивидуальным особенностям пользователя и подстраиваться в случае их изменения. Иммерсивная среда обеспечивает высокое вовлечение пользователя в интерактивное взаимодействие путем формирования корректных последовательностей аудиовизуальных стимулов в результате высокореалистичного моделирования, симуляции и визуализации искусственного окружения в контексте естественного и реального опыта конкретного пользователя.

## Материалы и методы

Программный комплекс психодиагностики с обратной связью на основе компьютерного зрения построен на базе интерактивного веб-интерфейса с возможностью подключения внешней видеокамеры для контроля движения лица и эмоций человека и системы окулографии (айтрекинга). Применение дополнительного аппаратного обеспечения позволяет фиксировать обратную связь от пользователя независимо от процессов функционирования программного комплекса, но синхронизированно со сценариями использования системы. Это обеспечивает контроль вовлеченности и при необходимости позволяет привлекать внимание пользователя.

Комплекс имеет сервис-ориентированную архитектуру. Основной интерфейс построен в виде одностраничного веб-приложения. Доступ к функциональным возможностям программного комплекса обеспечивается веб-сервисами, а основная информация о пользователях, тестах и результатах их прохождения хранится в реляционной БД.

Данные видеокамеры собираются сервисом видеостриминга в потоковом виде и передаются по протоколу пользовательских датаграмм UDP. Для сбора видеоаналитики используется видеоброкер *mediamtx*, реализованный на языке Go. Для распознавания лиц и эмоций применяется технология искусственных нейронных сетей. Для описываемого случая целесообразно использовать конфигурацию сверточной нейронной сети из библиотеки *dlib*. Это оптимальное решение для работы в режиме реального времени, позволяющее точно обрабатывать входные данные видеопотока и выполнять поставленные задачи.

Для автоматизации обнаружения и классификации изображений была использована библиотека *FaceRecognition*, предназначенная для распознавания лиц. К преимуществам использования этой библиотеки можно отнести следу-

ющее: одна из самых быстрых библиотек; открытая лицензия MIT; доступ с языков Python, C++, C#; поддержка операционных систем Linux и Windows; использование библиотеки алгоритмов компьютерного зрения OpenCV. Библиотека позволяет анализировать ориентацию головы на изображении по ключевым точкам лица и производить анализ этих ключевых точек, делая доступной и удобной оценку поворота головы.

Сервис окулографии построен с помощью проводного игрового контроллера Tobii Eye Tracker 5. Устройство позволяет отслеживать положения головы и глаз, взаимодействие с устройством осуществляется через веб-приложение, обеспечивающее синхронизацию потоков видео и окулографии. При этом траектория глаз передается в пакетном режиме на сервер с использованием дополнительного порта. Финальная синхронизация потоков выполняется на сервере.

Управление погружением пользователя в иммерсивную реальность осуществляется путем мониторинга и контроля выполняемых действий в ответ на возникающие события. В качестве таких событий могут рассматриваться воздействия на пользователя в реальном мире, а также аудиовизуальные стимулы, формируемые в иммерсивной среде автоматически. Ответное поведение должно включать цепочки действий, выполненных пользователем посредством интерфейсов человеко-компьютерного взаимодействия и автоматически фиксируемых средствами компьютерной техники.

### **Архитектура разработанного программного комплекса**

Система управления погружением пользователя в иммерсивную реальность была использована в психодиагностике при реализации программного комплекса психологического тестирования в рамках построения дополнительного контура обратной связи. При проведении психологического тестирования и реабилитации необходимо обеспечивать высокую вовлеченность пользователя, для чего предусмотрена персонализация пользовательского интерфейса и тестов. В ходе тестирования врач может корректировать перечень и содержание тестов, ускорять их прохождение или вводить специальные паузы в случае недостатка внимания.

Для этого в состав системы включен комплекс обратной связи на основе системы ком-

пьютерного зрения, включающей видеокамеру и ПО видеоконтроля двигательной активности пациента. Это ПО отслеживает общий объем мимических движений, движений головы, а также идентифицирует текущее эмоциональное состояние пациента с помощью искусственной нейронной сети.

Случаи, например, потери внимания, роста и падения глубины погружения, эмоциональной перегрузки характеризуются соответствующими паттернами двигательной активности. Идентификация этих паттернов производится по результатам обработки временных рядов, описывающих двигательную активность человека. Рассматриваются корреляционные характеристики и результаты аппроксимации временных рядов.

Следует отметить, что характерные паттерны могут быть выделены для типичных состояний, характеризующих динамические изменения вовлеченности пользователя, однако для каждого человека они различаются. В связи с этим имеет смысл взамен классификации пользователей по типам психологической активности рассматривать индивидуальные изменения глубины погружения для каждого отдельного пользователя, сравнивая его текущее состояние с предыдущим типовым.

Структурная схема программного комплекса отображена на рисунке 1. Базовый контур психологического тестирования реализован модулями психодиагностики и реабилитации. Здесь предоставляется возможность прохождения различных тестов и процедур с помощью интерактивных пользовательских интерфейсов и комплексов виртуальной реальности. Результаты сохраняются в базу знаний. Дополнительный контур реализован модулем видеоконтроля, который предназначен для реализации обратной связи на основе системы компьютерного зрения.

В модуле мониторинга содержится информация обо всех пользователях и их активности. По каждому пациенту хранятся данные с результатами прохождения психологических тестирований и реабилитационных мероприятий. В системе реализован функционал, позволяющий формировать отчетные формы с необходимым набором данных, а также графики по полученным результатам.

### **Разработка программного комплекса**

Процесс создания программного комплекса включал разработку технического задания сов-

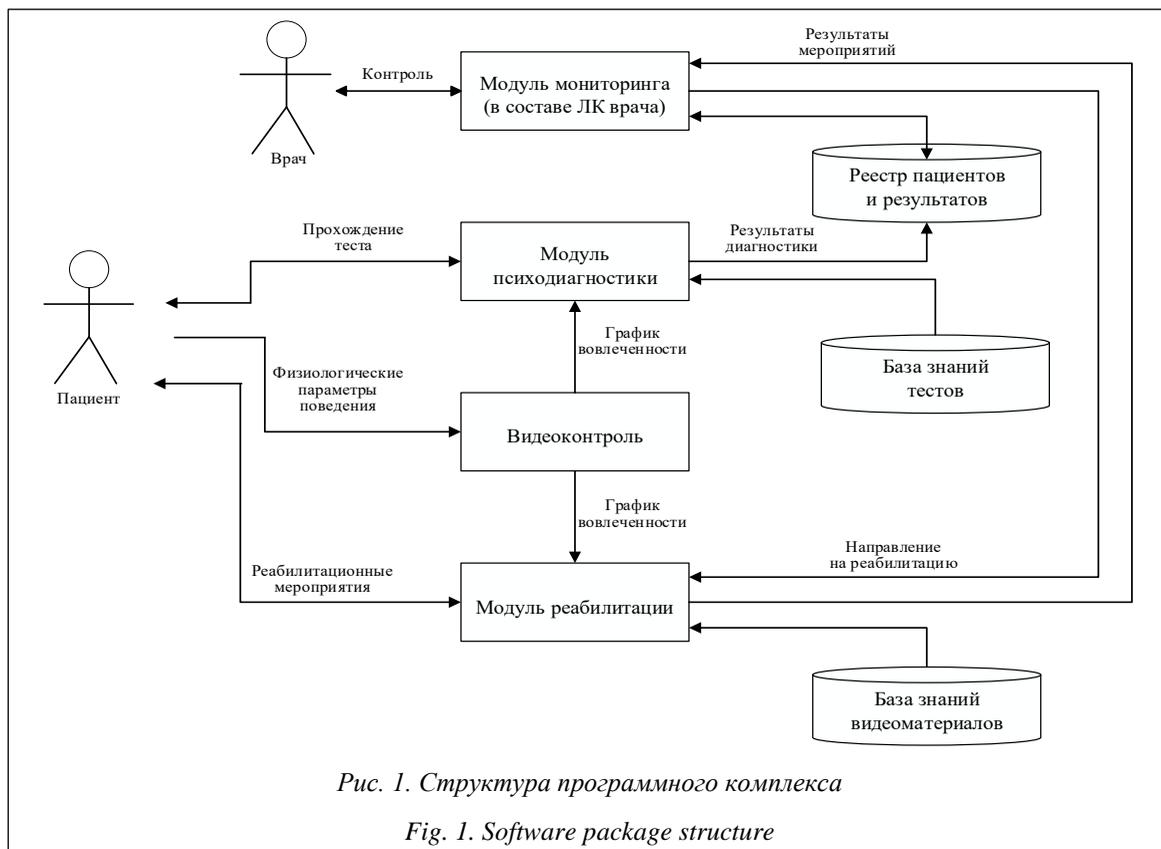


Рис. 1. Структура программного комплекса

Fig. 1. Software package structure

местно с сотрудниками Центра психологии Самарского государственного медицинского университета, проектирование программного комплекса, обучение интеллектуальных компонентов компьютерного зрения, внедрение и апробацию в реальных условиях эксплуатации. Отдельное внимание было уделено созданию обучающего набора данных (датасета) для модели искусственной нейронной сети. Для получения необходимых весов модель обучалась на 5 000 изображений, разделенных на обучающую (80 %) и тестовую (20 %) выборки. Перед началом процесса обучения была проведена подготовка для оптимизации набора данных и обеспечения реакции модели на режим обучения.

Каждое изображение в наборе данных было подвергнуто единым процедурам изменения размера и обрезки для установления единого масштаба и обеспечения равного представления всех образцов. Кроме того, было выполнено тщательное выравнивание черт лица, чтобы центрировать его на каждом изображении и обеспечить стандартизованную перспективу. Для уменьшения возможного влияния неравномерности освещенности использован процесс нормализации света. Это позволило унифицировать уровни освещенности в наборе

данных и тем самым снизить их влияние на процесс обучения модели.

Архитектура программного комплекса приведена на рисунке 2.

Серверная часть комплекса реализована на Django (Python 3.10), пользовательский интерфейс – на ReactJS (TypeScript + JS). Для обратной связи реализованы сервисы распознавания лица (на языке C#), захвата изображения с USB веб-камеры (PHP) и стриминга видео (bluenviron/mediamtx), а также приложение для взаимодействия с айтрекером Tobii Eye Tracker 5 (на языке C#). Базы знаний реализованы с помощью СУБД PostgreSQL 12. Серверная часть работает на Ubuntu 22.04. Клиентская часть адаптивна под все браузеры, например, Chrome, Firefox, Safari.

### Результаты реализации программного комплекса

Программный комплекс психодиагностики представляет последовательность тестов, формируемых для пациента (рис. 3). Реализованные диагностические методики включают модифицированный вариант методики словесно-цветовой интерференции, тест «Социальный интеллект» Гилфорда, тест Бурдона, таблицы

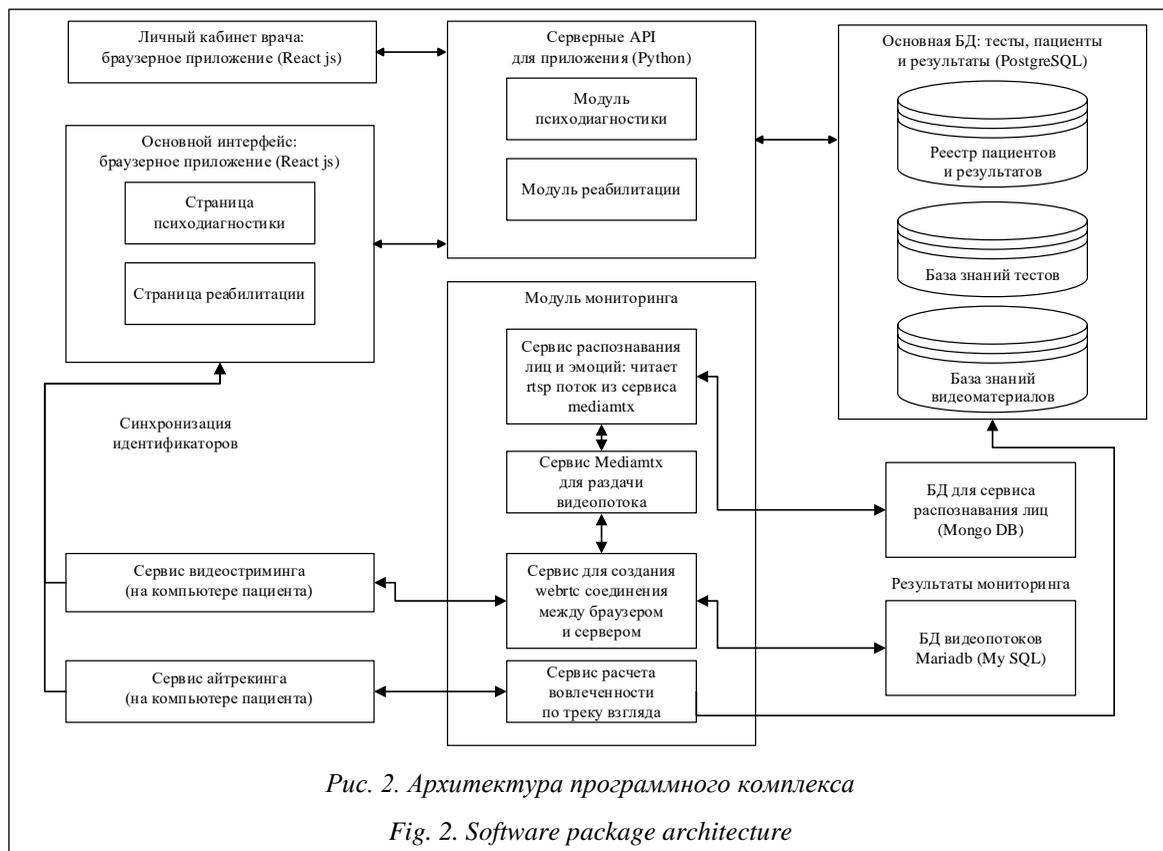


Рис. 2. Архитектура программного комплекса

Fig. 2. Software package architecture

Шульте, опросник эмоциональной эмпатии EETS, методику диагностики самооценки мотивации одобрения Марлоу–Крауна, оценку по Торонтской алекситимической шкале и шкале тревоги Спилбергера. Перечень методик может быть расширен через дополнительные настройки.

Рекомендации формируются для каждого респондента индивидуально в зависимости от результатов ранее пройденных тестов, его активности. В процессе формирования рекомендаций тесты приоритизируются следующим образом: сначала респонденту предлагается пройти тесты из группы «Восприятие» (группа А), затем из группы «Эмоции» (группа В), далее «Личность» (группа С). В первую очередь рекомендуется пройти тесты, которые не были пройдены ранее. Затем пользователю предлагаются тесты, которые были завершены им досрочно (если такие имеются). Если результаты тестирования неудовлетворительные, предлагается пройти реабилитационное мероприятие, а затем заново пройти тест с неудовлетворительными результатами.

На главной странице специалиста, помимо данных о результатах прохождения тестирования, отображается сводная диаграмма для каждого обследуемого (<http://www.swsys.ru/uploaded/image/2024-2/11.jpg>). На диаграмме показаны

графики результатов психодиагностики и вовлеченности (неустойчивость внимания), приведенные к значениям от 0 до 1. На главной странице специалиста также отображается история активности респондентов с возможностью выбора нужного периода.

Сводная диаграмма результатов демонстрирует специалисту зависимость между результатами прохождения тестов и степенью вовлеченности пользователя в процесс их прохождения. В зависимости от группы тестов требуется разная степень вовлеченности. Тесты группы «Восприятие» требуют высокой концентрации респондента, следовательно, качество результатов прохождения теста напрямую зависит от степени вовлеченности. Для тестов группы «Эмоции» в меньшей степени требуется высокая концентрация на интерфейсе системы. Тесты данной группы, как правило, содержат вопросы с несколькими вариантами ответов. Выбранный ответ напрямую отражает истинное состояние пользователя. О качестве прохождения данных тестов могут свидетельствовать общее время прохождения каждого теста, а также оценка внимания пользователя. Тесты группы «Личность» требуют средней вовлеченности в интерфейс системы относительно двух других названных групп.

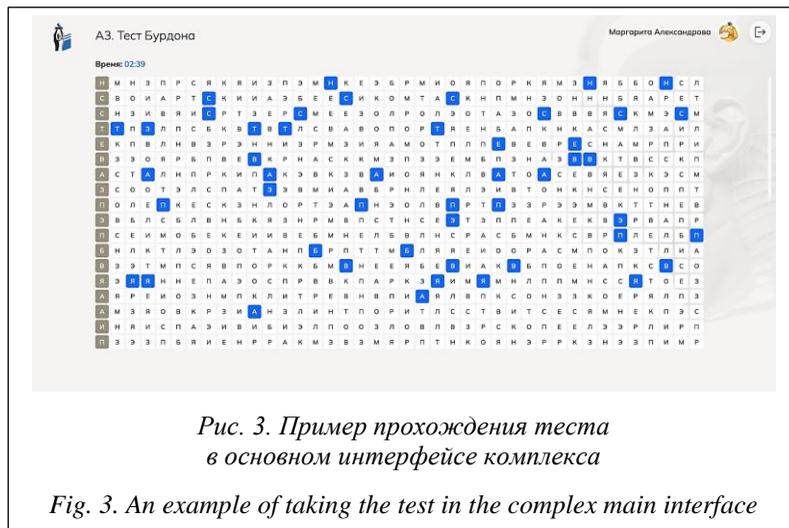


Рис. 3. Пример прохождения теста в основном интерфейсе комплекса

Fig. 3. An example of taking the test in the complex main interface

Сводная диаграмма результатов может быть построена как по одному респонденту и его первичным результатам по прохождению каждого теста, так и по группе респондентов или по нескольким результатам одного респондента, если требуется повторное прохождение тестов для отслеживания динамики изменения результатов (<http://www.swsys.ru/uploaded/image/2024-2/11.jpg>). При повторном прохождении тестов респондентами сводная диаграмма результатов позволяет оценить его качество с учетом общего времени, концентрации пользователей и их активности.

Полученные результаты подтверждают гипотезу о том, что наибольшая концентрация и степень вовлеченности необходимы при прохождении тестов группы «Восприятие». При этом при повторном прохождении психодиагностики тесты данной группы вызывают наибольший интерес у респондентов: отмечаются повышение скорости прохождения тестов, высокая степень концентрации и незначительные потери в качестве результатов.

Следует отметить, что первичное прохождение тестов дает наиболее качественный результат, так как степень вовлеченности респондента достаточно высокая. При повторном прохождении тестов наблюдаются снижение уровня вовлеченности пользователя и, как следствие, незначительное снижение качества тестирования.

### Заключение

Формирование иммерсивной среды с контролем высокой вовлеченности пользователей является важной задачей ПО для автоматизированного психологического тестирования. Реализация предложенных в данной статье модели и системы управления позволяет обеспечить такую возможность и тем самым персонализировать пользовательские интерфейсы с учетом индивидуальных особенностей пользователей. Мониторинг вовлеченности пользователей позволяет верифицировать результаты психодиагностики и контролировать влияние на них внешних факторов.

Приведенная иллюстрация на примере системы психодиагностики показывает, насколько важно собирать и анализировать информацию о результатах использования программных систем пользователями для оценки степени их вовлеченности.

Дальнейшие исследования планируется направить на практическую апробацию предложенных решений и расширение области их внедрения.

### Список литературы

1. Noland R. Intelligence testing using a tablet computer: Experiences with using Q-Interactive. *Training and Education in Professional Psychology*, 2017, vol. 11, no. 3, pp. 156–163. doi: 10.1037/tep0000149.
2. Baker A., Iqbal F., Laila M., Waheed A. Psychology with mahnoor app: Android-based application for self assessment, psychology dictionary, and notes. In: *Mobile Devices and Smart Gadgets in Medical Sciences*, 2020, pp. 214–231. doi: 10.4018/978-1-7998-2521-0.ch010.
3. Войтов В.К., Шепелева Е.А., Гаврилова Е.В., Думин П.Н., Ермаков С.С. Новые средства психологического тестирования // *Моделирование и анализ данных*. 2021. Т. 11. № 1. С. 94–108. doi: 10.17759/mda.2021110107.
4. Dhengre S., Mathur J., Oghazian F., Tan X., Mccomb C. Towards enhanced creativity in interface design through automated usability evaluation. *Proc. ICCS'20*, 2020, pp. 366–369.
5. Irshad S., Azmi S., Begum N. Uses of artificial intelligence in psychology. *J. of Health and Medical Sci.*, 2022, vol. 5, no. 4, pp. 21–30. doi: 10.31219/osf.io/2hx5n.
6. Tiwari A., Chugh A., Sharma A. Uses of artificial intelligence with human-computer interaction in psychology. In: *Intelligent Data-Centric Systems, Innovations in Artificial Intelligence and Human-Computer Interaction in the Digital Era*, 2023, pp. 173–205.
7. Тычков А.Ю., Волкова К.Ю., Киселева Д.В., Родионова Е.А. Обзор систем виртуальной реальности // *Изв. вузов. Поволжский регион. Технич. науки*. 2020. № 2. С. 3–13. doi: 10.21685/2072-3059-2020-2-1.

8. Поцелуйко А.С., Кравец А.Г., Кульцова М.Б. Персонализация интерфейсов мобильных приложений на основе паттернов интерфейсов для людей с ограниченными возможностями // Прикаспийский журнал: управление и высокие технологии. 2019. № 3. С. 17–27. doi: 10.21672/2074-1707.2019.47.3.017-027.

9. Тагирова Л.Ф., Зубкова Т.М. Нейросетевая инструментальная среда для создания персонализированных интерфейсов прикладных программ // Программные продукты и системы. 2023. Т. 36. № 2. С. 212–222. doi: 10.15827/0236-235X.142.212-222.

10. Coco M. Eye-tracking: Measurements and applications. In: Encyclopedia of Behavioral Neuroscience, 2021, pp. 204–214. doi: 10.1016/B978-0-12-819641-0.00158-4.

11. Santhoshikka R., Laranya C.R., Harshavarthini C., Preetha R., Saran Kumar K. Eye tracking and its applications. IJARSET, 2021, vol. 8, no. 8, pp. 126–130. doi: 10.17148/IJARSET.2021.8824.

12. Ильяшенко О.Ю., Лукьянченко Е.Л. Возможности применения компьютерного зрения для аналитики данных в медицине // Изв. Саратовского ун-та. Новая сер. Сер.: Математика. Механика. Информатика. 2022. Т. 22. № 2. С. 224–232. doi: 10.18500/1816-9791-2022-22-2-224-232.

13. Куприянов Р.Б. Применение технологий компьютерного зрения для автоматического сбора данных об эмоциях обучающихся во время групповой работы // Информатика и образование. 2020. № 5. С. 56–63. doi: 10.32517/0234-0453-2020-35-5-56-63.

14. Lazurik V., Styervoyedov, N., Varlamova N. Information processing psychodiagnostic system: Designing and implementation. Eastern-European J. of Enterprise Tech., 2021, vol. 2, pp. 45–54. doi: 10.15587/1729-4061.2021.230042.

15. Иващенко А.В., Александрова М.В., Жейков Д.С., Захарова Е.В., Колсанов А.В. Адаптация интерфейсов виртуальной реальности в приложениях психодиагностики и медицинской реабилитации // Медицинская техника. 2023. № 5. С. 33–35.

16. Иващенко А.В., Александрова М.В. Формирование иммерсивной среды с управляемой глубиной погружения в приложения дополненной реальности // Модели, системы, сети в экономике, технике, природе и обществе. 2023. № 1. С. 111–124. doi: 10.21685/2227-8486-2023-1-7.

Software &amp; Systems

doi: 10.15827/0236-235X.142.193-200

2024, 37(2), pp. 193–200

### Psychodiagnostics software with computer vision feedback

Anton V. Ivaschenko <sup>1</sup>✉, Margarita V. Aleksandrova <sup>2</sup>, Denis S. Zheikov <sup>1</sup>,  
Elena V. Mazankina <sup>1</sup>, Elena V. Zakharova <sup>1</sup>, Aleksandr V. Kolsanov <sup>1</sup>

<sup>1</sup> Samara State Medical University, Samara, 443099, Russian Federation

<sup>2</sup> SEC “Open code”, Samara, 443001, Russian Federation

#### For citation

Ivaschenko, A.V., Aleksandrova, M.V., Zheikov, D.S., Mazankina, E.V., Zakharova, E.V., Kolsanov, A.V. (2024) ‘Psychodiagnostic software with computer vision feedback’, *Software & Systems*, 37(2), pp. 193–200 (in Russ.). doi: 10.15827/0236-235X.142.193-200

#### Article info

Received: 21.09.2023

After revision: 09.01.2024

Accepted: 18.03.2024

**Abstract.** The paper presents psychological diagnostic software that considers the involvement of respondents when taking tests. The research focuses on managing the depth of user immersion in terms of interactive user interfaces. The paper proposes a formal model of an immersive environment intended to implement a system for managing user immersion into human-computer interaction processes by monitoring and controlling the actions performed in response to emerging events. Such events are effects on a user in the real world, as well as audiovisual stimuli generated automatically in an immersive environment. Response behavior includes action chains automatically recorded by computer vision and oculo-graphy means. Unlike analogues, the developed software package for psychological diagnostics and medical rehabilitation allows implementing feedback on monitoring and control of patient involvement. For this purpose, the system includes a computer vision subsystem with a video camera and software for video monitoring of patient’s motor activity. The software tracks the total volume of facial movements, head movements, and also identifies the current emotional state using an artificial neural network. The user interface and tests are personalized to ensure high user involvement. The practical significance of the work is monitoring the involvement of patients when undergoing psychological testing and using this information to personalize medical services. Analysis of involvement of psychological diagnostic software package users makes it possible to supplement psychological testing results and adapt the sequence and content of tests while maintaining user’s interest and reducing the influence of external factors.

**Keywords:** automation of psychological testing, user interfaces, human-computer interaction, immersive reality, accented visualization

#### References

1. Noland, R. (2017) ‘Intelligence testing using a tablet computer: Experiences with using Q-Interactive’, *Training and Education in Professional Psychology*, 11(3), pp. 156–163. doi: 10.1037/tep0000149.
2. Baker, A., Iqbal, F., Laila, M., Waheed, A. (2020) ‘Psychology with mahnoor app: Android-based application for self assessment, psychology dictionary, and notes’, in *Mobile Devices and Smart Gadgets in Medical Sciences*, pp. 214–231. doi: 10.4018/978-1-7998-2521-0.ch010.
3. Voitov, V.K., Shepeleva, E.A., Gavrilova, E.V., Dumin, P.N., Ermakov, S.S. (2021) ‘New psychological testing tools’, *Modelling and Data Analysis*, 11(1), pp. 94–108 (in Russ.). doi: 10.17759/mda.2021110107.

4. Dhengre, S., Mathur, J., Oghazian, F., Tan, X., McComb, C. (2020) 'Towards enhanced creativity in interface design through automated usability evaluation', *Proc. ICCV'20*, pp. 366–369.
5. Irshad, S., Azmi, S., Begum, N. (2022) 'Uses of artificial intelligence in psychology', *J. of Health and Medical Sci.*, 5(4), pp. 21–30. doi: 10.31219/osf.io/2hx5n.
6. Tiwari, A., Chugh, A., Sharma, A. (2023) 'Uses of artificial intelligence with human-computer interaction in psychology', in *Intelligent Data-Centric Systems, Innovations in Artificial Intelligence and Human-Computer Interaction in the Digital Era*, pp. 173–205.
7. Tychkov, A., Volkova, K., Kiseleva, D., Rodionova, E. (2020) 'A review of virtual reality systems', *University Proc. Volga Region. Tech. Sci.*, (2), pp. 3–13 (in Russ.). doi: 10.21685/2072-3059-2020-2-1.
8. Potseluiiko, A.S., Kravets, A.G., Kultsova, M.B. (2019) 'Personalization of mobile application interfaces on the basis of interface patterns for people with special needs', *Caspian J.: Control and High Tech.*, (3), pp. 17–27 (in Russ.). doi: 10.21672/2074-1707.2019.47.3.017-027.
9. Tagirova, L.F., Zubkova, T.M. (2023) 'Neural network tool environment for creating adaptive application program interfaces', *Software & Systems*, 36(2), pp. 212–222 (in Russ.). doi: 10.15827/0236-235X.142.212-222.
10. Coco, M. (2021) 'Eye-tracking: Measurements and applications', in *Encyclopedia of Behavioral Neuroscience*, pp. 204–214. doi: 10.1016/B978-0-12-819641-0.00158-4.
11. Santhoshikka, R., Laranya, C.R., Harshavarthini, C., Preetha, R., Saran Kumar, K. (2021) 'Eye tracking and its applications', *IJARSET*, 8(8), pp. 126–130. doi: 10.17148/IJARSET.2021.8824.
12. Iliashenko, O.Yu., Lukyanchenko, E.L. (2022) 'Possibilities of using computer vision for data analytics in medicine', *Izv. of Saratov University. Mathematics. Informatics*, 22(2), pp. 224–232 (in Russ.). doi: 10.18500/1816-9791-2022-22-2-224-232.
13. Kupriyanov, R.B. (2020) 'Application of computer vision technologies for automatic data collection about emotions of students during group work', *Informatics and Education*, (5), pp. 56–63 (in Russ.). doi: 10.32517/0234-0453-2020-35-5-56-63.
14. Lazurik, V., Styervoyedov, N., Varlamova, N. (2021) 'Information processing psychodiagnostic system: Designing and implementation', *Eastern-European J. of Enterprise Tech.*, 2, pp. 45–54. doi: 10.15587/1729-4061.2021.230042.
15. Ivaschenko, A.V., Aleksandrova, M.V., Zheikov, D.S., Zakharova, E.V., Kolsanov, A.V. (2023) 'Adaptation of virtual reality interfaces in applications of psychodiagnostics and medical rehabilitation', *Medical Equipment*, (5), pp. 33–35 (in Russ.).
16. Ivaschenko, A.V., Aleksandrova, M.V. (2023) 'Development of immersive environment with controlled depth perception in augmented reality applications', *Models, Systems, Networks in Economics, Technology, Nature and Society*, (1), pp. 111–124 (in Russ.). doi: 10.21685/2227-8486-2023-1-7.

**Авторы**

**Ивашенко Антон Владимирович**<sup>1</sup>, д.т.н., профессор, директор Передовой медицинской инженерной школы Самарского государственного медицинского университета, anton.ivashenko@gmail.com  
**Александрова Маргарита Владимировна**<sup>2</sup>, руководитель проектов, margarita.alexandrowa@mail.ru  
**Жейков Денис Сергеевич**<sup>1</sup>, преподаватель, d.s.zhejkov@samsmu.ru  
**Мазанкина Елена Владимировна**<sup>1</sup>, директор Центра психологии Самарского государственного университета, e.v.mazankina@samsmu.ru  
**Захарова Елена Владимировна**<sup>1</sup>, д.псх.н., доцент, зав. кафедрой, e.v.zakharova@samsmu.ru  
**Колсанов Александр Владимирович**<sup>1</sup>, д.м.н., профессор РАН, ректор, a.v.kolsanov@samsmu.ru

**Authors**

**Anton V. Ivaschenko**<sup>1</sup>, Dr.Sc. (Engineering), Professor, Director of the Higher School of Medical Engineering at Samara State Medical University, anton.ivashenko@gmail.com  
**Margarita V. Aleksandrova**<sup>2</sup>, Project Manager, margarita.alexandrowa@mail.ru  
**Denis S. Zheikov**<sup>1</sup>, Lecturer, d.s.zhejkov@samsmu.ru  
**Elena V. Mazankina**<sup>1</sup>, Director of the Center for Psychology at Samara State Medical University, e.v.mazankina@samsmu.ru  
**Elena V. Zakharova**<sup>1</sup>, Dr.Sc. (Psychology), Associate Professor, Head of Chair, e.v.zakharova@samsmu.ru  
**Aleksandr V. Kolsanov**<sup>1</sup>, Dr.Sc. (Medics), Professor of RAS, Rector, a.v.kolsanov@samsmu.ru

<sup>1</sup> Самарский государственный медицинский университет, г. Самара, 443099, Россия

<sup>2</sup> Компания «Открытый код», г. Самара, 443001, Россия

<sup>1</sup> Samara State Medical University, Samara, 443099, Russian Federation  
<sup>2</sup> SEC "Open code", Samara, 443001, Russian Federation

## Автоматическая генерация заданий по результатам анализа использования их банка в интеллектуальной обучающей системе

О.А. Сычев <sup>1</sup>✉, А.А. Прокудин <sup>1</sup>, М.Е. Денисов <sup>1</sup>

<sup>1</sup> Волгоградский государственный технический университет,  
г. Волгоград, 400005, Россия

### Ссылка для цитирования

Сычев О.А., Прокудин А.А., Денисов М.Е. Автоматическая генерация заданий по результатам анализа использования их банка в интеллектуальной обучающей системе // Программные продукты и системы. 2024. Т. 37. № 2. С. 201–212. doi: 10.15827/0236-235X.142.201-212

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 18.10.2023

После доработки: 20.12.2023

Принята к публикации: 22.12.2023

**Аннотация.** Развитие методов генерации учебных заданий и их классификации для применения в учебном процессе позволило реализовать интеллектуальную обучающую систему, способную при необходимости автоматически генерировать задания в фоновом режиме. В статье описана архитектура интеллектуальной обучающей системы с фоновой генерацией заданий, рассмотрена система CompPrehension, реализующая этот принцип. Для решения проблемы использовались методы символического искусственного интеллекта и онтологических рассуждений, объектно-ориентированного проектирования и программирования, а также системного анализа. В работе представлены основные функциональные элементы таких компонентов системы, как тренажер, банк заданий, генератор заданий и домен (предметная область), а также их взаимодействие с БД для хранения заданий и с машинами вывода для их решения. Апробация подхода на банке заданий по определению порядка выполнения операторов в выражении показала способность системы избирательно пополнять несбалансированный банк, исходя из его состава и истории запросов на поиск заданий в нем. Менее чем за пять суток было сгенерировано несколько тысяч заданий на каждый концепт с недостаточным числом заданий. Это дает низкую вероятность получения одинаковых заданий даже при одновременном тестировании сотен студентов. Результаты исследования имеют практическую значимость для учебных заведений, стремящихся совершенствовать обучающий процесс.

**Ключевые слова:** интеллектуальные обучающие системы, генерация вопросов, дистанционное обучение, разработка обучающих систем, CompPrehension, банк заданий

**Введение.** В процесс преподавания входят два фундаментальных цикла обратной связи – цикл обучаемого (малый) и цикл преподавателя (большой). Малый цикл включает попытку обучаемого применить изученный материал (ответ на тестовый вопрос, решение задачи), оценку этого решения преподавателем, выдачу обратной связи о качестве решения и причинах или характере ошибок (при наличии) и подбор следующего задания для наиболее эффективного обучения. Предназначение цикла – дать возможность научиться конкретному обучаемому. Большой цикл осуществляется преподавателем и подразумевает выдачу и контроль заданий (тестов), оценку эффективности их решения, оценку качества банка заданий (вопросов) и выявление недостатков, корректировку банка заданий. Он предназначен для совершенствования учебного процесса и повышения его качества. Однако многие задачи как в малом, так и в большом цикле являются рутинными и трудоемкими, что ограничивает возможности преподавателей работать с конкретными обучаемыми и совершенствовать свои курсы.

Автоматизированные обучающие системы призваны улучшить образовательный процесс

за счет автоматизации части задач – от предоставления информации об изучаемой теме до проверки решения задач и объяснения причин ошибок [1]. При разработке систем адаптивного тестирования и интеллектуальных обучающих систем (ИОС) достигнут большой прогресс в автоматизации цикла обратной связи обучаемого, включая выдачу подсказок, объяснение сделанных ошибок и адаптивный подбор следующей задачи на основе решения обучаемым предыдущих.

Гораздо меньше усилий вложено в автоматизацию цикла обратной связи преподавателя, то есть в анализ и пополнение банка заданий (вопросов), хотя качество заданий является одним из ключевых факторов успешности ИОС. Небольшое количество заданий в банке приводит к тому, что многие обучаемые будут получать одинаковые задания. Это провоцирует обмен правильными ответами и списывание, так что, банки заданий, составленные вручную, плохо масштабируются при росте аудитории. Кроме того, адаптивные обучающие системы должны иметь широкий выбор возможных задач различного уровня сложности на все изучаемые понятия и законы предметной области,

чтобы раскрыть потенциал адаптивных алгоритмов.

Автоматизация замыкания большого цикла обратной связи в процессе преподавания включает автоматическую генерацию учебных заданий, анализ банка заданий и его пополнение. Однако большинство работ по автоматической генерации заданий являются экспериментальными; сгенерированные задания, как правило, невозможно использовать в учебном процессе без редактирования или отбора преподавателем [2]. К тому же в процессе генерации редко учитываются результаты анализа банка заданий.

Новые исследования в области автоматической генерации и классификации учебных заданий открывают возможности их прямого использования без отбора и проверки человеком [3], а также разработки архитектур ИОС. Цель авторов настоящей статьи – разработка и апробация архитектуры ИОС, способной по мере автоматизированного адаптивного обучения (с реализацией малого цикла обратной связи) автоматически замкнуть большой цикл обратной связи: анализировать недостатки существующего банка заданий в ходе его реального применения, создавать запросы на генерацию дополнительных заданий, автоматически генерировать учебные задания и выбирать подходящие для включения в банк и использования в дальнейших учебных сессиях.

### **Обзор автоматизированных обучающих систем для обучения программной инженерии**

Система обучения Problets [4] предоставляет несколько типов заданий на понимание принципов процедурного программирования для трех языков. Создание вопросов происходит на основе шаблонов с рандомизацией, а обратная связь представляет собой контекстные объяснения для любого шага решения. Объяснение ошибок осуществляется через комментарии о том, как должно быть. В распространяемом комплекте Problets содержится, по-видимому, фиксированный набор шаблонов для задач, который не подлежит модификации во время эксплуатации. Пополнение возможно через шаблоны в нотации BNF, вручную задаваемые преподавателем.

В работе [5] представлена ASSISTment Builder – среда создания тренажеров с линейными примерами, нацеленная на сокращение времени формирования навыков и создания во-

просов. Они описывают функцию «вариаблизации», позволяющую создавать шаблоны с переменными, и отмечают неспособность «обобщать схожие проблемы», что приводит к большому количеству работы по принципу «скопируй и немного измени». Поскольку похожие друг на друга учебные задачи полезны для практики начинающим, использование одного и того же обучающего контента для создания группы схожих задач облегчило бы их составление.

Система ASSISTment предлагает обучающемуся два типа направляющей обратной связи: серию текстовых подсказок о том, как решить задачу, и серию вспомогательных вопросов, отображаемых в случае неправильного ответа на «главную задачу». Каждый вспомогательный вопрос имеет несколько подсказок, и каждая задача (вопрос) может отображать заранее подготовленное сообщение как реакцию на известный неправильный ответ. Система ASSISTment предоставляет три варианта организации последовательности учебных задач: фиксированный, случайный и ограниченный заранее определенным набором задач (данные об успеваемости учащегося не используются). Предложенная авторами архитектура предполагает полный контроль над контентом со стороны преподавателя. Банк заданий обновляется вручную. Для помощи создателям контента добавлены возможность комментирования вопросов учащимися и простая статистика наиболее частых неправильных ответов.

Автоматически генерировать задачи на определение результатов выполнения кода, случайным образом изменяя константы и операторы в синтаксическом дереве фрагмента кода на языке Java, позволяет Unlimited Trace Tutor [1]. Для генерации используется фиксированный набор простых шаблонов, содержащих циклы for, while и развилки if (пополнение банка заданий не предполагается). Для получения более сложных фрагментов кода шаблоны могут комбинироваться. При запуске тренировки учащийся может выбрать алгоритмические структуры, которые он хочет изучать, а также уровень сложности. Продолжительность тренировки может быть намеренно ограничена по времени.

Система Automata Tutor v3, описанная в источнике [2], поддерживает автоматическую генерацию для пяти типов задач и нескольких уровней сложности. Для задач на основе графов, таких как регулярные выражения, уровень сложности влияет на структуру генерируемого графа. Для задач на основе кода некоторые слу-

чайные преобразования применяются к созданным вручную базовым программам, разработанным для разных уровней сложности, а затем задачи низкого качества отбрасываются. Составление базовых программ, видимо, является ответственностью разработчика новых видов задач.

Визуальная среда моделирования программ UWhistle [6] предоставляет богатые функциональные возможности для отслеживания каждого мелкого события, происходящего во время выполнения кода. Она может как демонстрировать действия, так и просить учащегося выполнить их вручную, предоставляя текстовые объяснения, когда учащийся ошибается. Система UWhistle содержит фиксированный набор встроенных примеров кода, а также позволяет использовать произвольный код на языке Python в качестве трассируемой программы. Система не осуществляет генерацию и пополнение банка заданий.

Почти все приведенные аналоги предоставляют преподавателю сводку по оценкам учащихся, и ни один из них не предполагает возможности использовать информацию о составе банка заданий и статистику его применения при генерации новых заданий.

### Генерация вопросов и учебных заданий

Автоматическая генерация может создать десятки и сотни тысяч вопросов за относительно короткое время. Для сравнения: банк вопросов ASSISTment содержит всего около 3 000 вопросов, созданных вручную более чем 30 преподавателями, а наименьшее время, затраченное на создание учебного материала, составляет около 40 часов на каждый час обучения [5].

Автоматизация управления учебными заданиями затрагивает два аспекта – создание вопросов и оценка ответов. Наиболее популярный [7] и надежный способ генерации вопросов состоит в использовании шаблона почти готовой задачи с местами для вставки (плейсхолдерами) [8]. Их перечень фиксирован [5] или выбирается случайным образом из заданного диапазона (для числовых величин) [9]. Одни подстановочные параметры могут вычисляться из других в момент формирования вопроса. Этот подход позволяет кратно размножить задания, созданные человеком, но ограничен набором ручных шаблонов, создание которых все еще требует больших затрат. Существуют вариации шаблонного метода, например, в ра-

боте [10] предлагается автоматически извлекать компоненты задания из подготовленного вручную исходного кода.

Другой подход использует перечень правил или дерево решений со случайными переходами для генерации шаблона с последующим заполнением случайными величинами, например, синтез [11] или модификация [1] программ на основе грамматики, модели NLP-трансформеров [12]. Этот подход позволяет генерировать сотни тысяч заданий с приемлемой вариативностью, но преподаватель имеет меньше контроля над структурой и содержанием генерируемых вопросов, а правильная настройка генератора может оказаться сложной задачей. Недостатком этого подхода является то, что для генерации заданий на другую тему или даже другого типа в рамках одной темы требуется разработка новой конфигурации для генератора или создание совершенно нового генератора.

В сложных и/или менее детерминированных предметных областях популярен подход с использованием специальных датасетов [13], или больших баз знаний, онтологий. По теме программирования тоже существуют онтологии, например, CodeOntology для языка Java [14], но они могут отсутствовать для целевого языка программирования или быть недостаточно информативными. Так, в готовом датасете CodeOntology удалена информация о действиях внутри методов, которая была бы полезна в настоящем исследовании. Этот подход предоставляет так же мало контроля над фактическим содержанием вопросов, но, с другой стороны, базируется на разнообразных реальных данных и процесс генерации более прямолинейный и масштабируемый, чем в предыдущем подходе.

Подход с использованием открытых данных – текстов [15] или открытого кода [16], который используется в данной работе, применим ко многим видам учебных вопросов и имеет высокий потенциал благодаря большому объему и разнообразию доступных через Интернет открытых данных, которые могут стать хорошим источником для шаблонов вопросов. Трудности могут возникать из-за разрозненности источников данных, отсутствия в них подходящих метаданных для быстрого поиска и сложностей предобработки и фильтрации сырых данных, к тому же возможен риск потенциально недопустимого контента. С точки зрения наполнения банка вопросов вероятен дисбаланс в разрезе различных существенных харак-

теристик вопросов из-за объективного дисбаланса в реальных данных. Например, попытка найти много разных примеров программного кода с использованием вложенных тернарных операторов может не увенчаться успехом, так как такое редко встречается в профессиональном коде из-за снижения читаемости. Другой пример: оператор `walrus` в Python является допустимым вариантом оператора присваивания и даже рекомендуется к использованию, но в профессиональном коде редко используется из-за непопулярности. Поэтому также необходимо планировать выравнивание численности непредставленных групп примеров.

Автоматически сгенерированные задачи должны быть автоматически проверяемыми [17], что помогает масштабировать все этапы обучения. «Длинные» вопросы с открытым ответом (например, написание программ и эссе), которые включают элемент творчества, проверить полностью автоматически невозможно, так как перечень критериев для определения правильности ответа слишком широк. Хотя есть попытки применить к таким ответам машинное обучение, качество проверок далеко от 100 %, и потому они используются, как правило, только для помощи оценщику-человеку [18].

Новизна данной работы заключается в практической разработке автоматизированной обучающей системы с возможностью автоматической (без участия человека) генерации заданий для пополнения банка знаний по мере необходимости. Разработанные ранее системы позволяли пополнять банк заданий вручную или автоматизированно, но при участии человека.

Практическая значимость исследования заключается в существенном уменьшении затрат преподавателей на поддержку банков заданий в автоматизированных обучающих системах, включая анализ повторений заданий и разработку новых.

### **Метод построения ИОС на основе автоматической генерации заданий**

Для эффективной работы ИОС должна обеспечивать учащихся заданиями для усвоения учебных тем, стремясь снизить до минимума частоту их повторения в рамках выполняемого упражнения, тем самым снижая вероятность списывания и запоминания уже решенного задания. Система, позволяющая автоматически замкнуть оба цикла обратной связи учебного процесса, должна решать в автоматическом режиме следующие задачи:

- подбор задания для решения на основе имеющихся данных о знаниях и умениях обучающегося;
- оценка решения задания;
- определение ошибок обучающегося и их возможных причин с генерацией объясняющих сообщений;
- определение классов заданий, которых недостаточно для эффективной работы;
- генерация учебных заданий, их классификация и отбор по заданным критериям;
- решение сгенерированных задач для определения множества правильных ответов.

Полезными, но необязательными являются генерация подсказок в процессе решения задания, демонстрация примеров решения заданий, ведение обучающего диалога с помощью наводящих вопросов.

При ручном планировании банка заданий в адаптивных обучающих системах очень сложно предсказать потенциальную потребность в заданиях различных категорий, поскольку это требует знания того, какие ошибки и трудности обучаемые будут испытывать чаще и задания какой сложности наиболее эффективны для обучения. Поэтому в процессе эксплуатации адаптивной системы возникает дефицит определенных видов задач. Это приводит к их частому повторению или подбору менее подходящих задач, что снижает качество обучения [19]. Так, при эксплуатации прототипа ИОС *Comp-Prehension* [20] с автоматически сгенерированным банком задач на тему «порядок вычисления выражений в алгоритмических языках высокого уровня» замечен значительный недостаток задач на выполнение операции внутри операций скобочного типа (вызовы функций, обращение к элементу массива) и на правую ассоциативность операций. При возникновении недостатка заданий должен создаваться запрос на генерацию новых заданий и активироваться механизм его выполнения. Ключевым фактором успеха является то, что ИОС обладает необходимой информацией о том, какие задания и насколько часто требуются, и способна самостоятельно пополнять банк заданий при достижении граничных условий, в роли которых могут выступать низкая точность выполнения запросов на поиск задач (нет задачи с подходящими свойствами, берем похожую) или высокая частота использования одной и той же задачи в одной группе студентов.

Запрос на генерацию заданий должен описывать ожидаемый уровень сложности, а также концепты и законы предметной области, необ-

ходимые для решения задачи. Он может также включать в себя перечень недопустимых концептов и законов в случае, если обучаемые изучили тему не полностью. Авторы предлагают абстракцию требований к заданию, которая определяет необходимые свойства задач с тем, чтобы во время проведения упражнения выполнить поиск в банке и выдать учащемуся задание, наиболее отвечающее указанным характеристикам. Это основной способ обеспечить генерацию уникальных упражнений из больших банков заданий, которые не являются обзорными для человека-преподавателя.

Предлагаемая архитектура такой обучающей системы состоит из трех основных компонентов.

- Тренажер. Взаимодействует с обучаемыми и преподавателями и реализует выбор заданий упражнений, их проверку, выдачу объясняющей обратной связи, генерацию подсказок.
- Банк заданий. Хранит задачи и выбирает подходящую случайную задачу по запросу. Ведет статистику использования заданий по упражнениям и может генерировать запросы на добавление заданий определенного типа в случае их недостаточного количества.
- Генератор задач. Производит фоновую генерацию учебных заданий в соответствии с запросами. Построен на основе подхода, использующего открытый программный код для генерации заданий [21].

Возможны различные сценарии использования генератора задач в процессе работы ИОС, а также однократный запуск генератора задач перед началом эксплуатации тренажера для генерации полного банка заданий. Это позволяет при необходимости эксплуатировать тренажер отдельно от генератора, а также дает преподавателю большую уверенность в банке заданий. Такой сценарий поможет в простых

предметных областях, где доля сложных заданий (с участием нескольких различных концептов или правил предметной области) невелика. При создании более комплексных ИОС при однократной генерации велик риск получить несбалансированный банк заданий, в котором некоторые темы будут обойдены вниманием.

Подход с обратной связью заключается в фоновой генерации заданий в зависимости от потребностей тренажера, в котором задания могут добавляться небольшими порциями, закрывая текущую потребность тренажера в задачах на плохо покрытые определенные темы (включая при необходимости сложные задания на комбинации концепций). На рисунке 1 показан процесс работы ИОС, основанной на генерации недостающих задач в фоновом режиме.

С практической точки зрения целесообразно применять оба подхода: для инициализации нового банка можно применить массовую генерацию без конкретных требований, а затем во время учебного процесса вести фоновую генерацию тех заданий, количество которых оказалось недостаточным для эффективного обучения.

Используя предложенную фоновую генерацию заданий ИОС могут обеспечить следующие преимущества перед классическими:

- экономия времени на создании банка заданий («скажи системе, чему надо научиться»);
- сокращение возможности обмена правильными ответами за счет большого количества непрерывно пополняющихся заданий;
- наполнение банка заданий на основе потребностей обучаемых (банк содержит преимущественно те задания, в которых есть потребность);
- улучшение точности подбора заданий под конкретные потребности обучаемых.

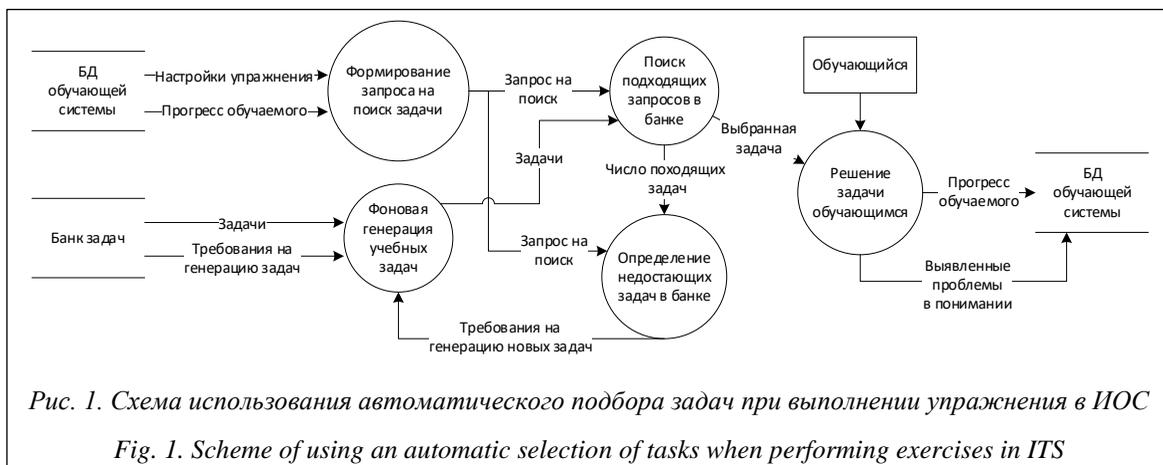


Рис. 1. Схема использования автоматического подбора задач при выполнении упражнения в ИОС

Fig. 1. Scheme of using an automatic selection of tasks when performing exercises in ITS

Этот подход не лишен недостатков. Например, можно выделить следующие:

- трудоемкость процесса написания генератора заданий для сложных предметных областей;
- отсутствие со стороны преподавателя контроля за автоматическим подбором заданий: недоверие преподавателя к действиям системы, которые он не контролирует, может тормозить ее внедрение;
- несоответствие стратегии подбора задания в ИОС стилю обучения преподавателя (в идеале ИОС должна поддерживать различные стратегии обучения).

На основе предложенного подхода доработана мультидоменная ИОС CompPrehension путем внедрения поддержки генерации заданий, в том числе на основании анализа потребностей системы. Архитектура разработанной системы показана на рисунке 2. Частичная диаграмма классов основных компонентов системы представлена на рисунке 3: ввиду большого количества классов в системе CompPrehension (208 классов) на диаграмме отражены только основные классы и методы компонентов, принимающие участие в процессах подбора вопросов, оценки ответов на вопросы и генерации новых заданий.

Центральный компонент системы, поддерживающий ее образовательные функции, – тренажер. Он организует выполнение всех ключевых процессов малого цикла обратной связи

в процессе обучения. Тренажер использует четыре основных базовых компонента, для которых возможно подключение альтернативных реализаций в виде плагинов, – домен (предметная область), инкапсулирующий все, что связано с особенностями конкретной предметной области, педагогическая стратегия, определяющая порядок выдачи задания, виды обратной связи с обучаемым, а также способы реакции на успехи и неудачи обучаемого, бэкенд (используется для интеграции внешних машин вывода, рассуждающих при решении задачи) и фронтенд (способ взаимодействия с пользователем, как правило, обучаемым). Домен содержит также модули, необходимые для генерации заданий, три остальных базовых компонента используются только тренажером.

Кроме того, в домене, как в формальном представлении предметной области, содержится все необходимое для создания вопросов, оценки ответов на них и предоставления обратной связи (рис. 3, компонент «Абстрактный домен»). Принимая решение об оценке правильности ответа обучаемого, домен использует бэкенд для запуска машины вывода и формальных рассуждений на основе правил предметной области, которым обучает система. Домен генерирует обратную связь обучаемому, включающую объяснения ошибок и подсказки. В настоящее время в системе CompPrehension реализованы домены «Порядок вычисления выражения» и «Алгоритмические структуры»,

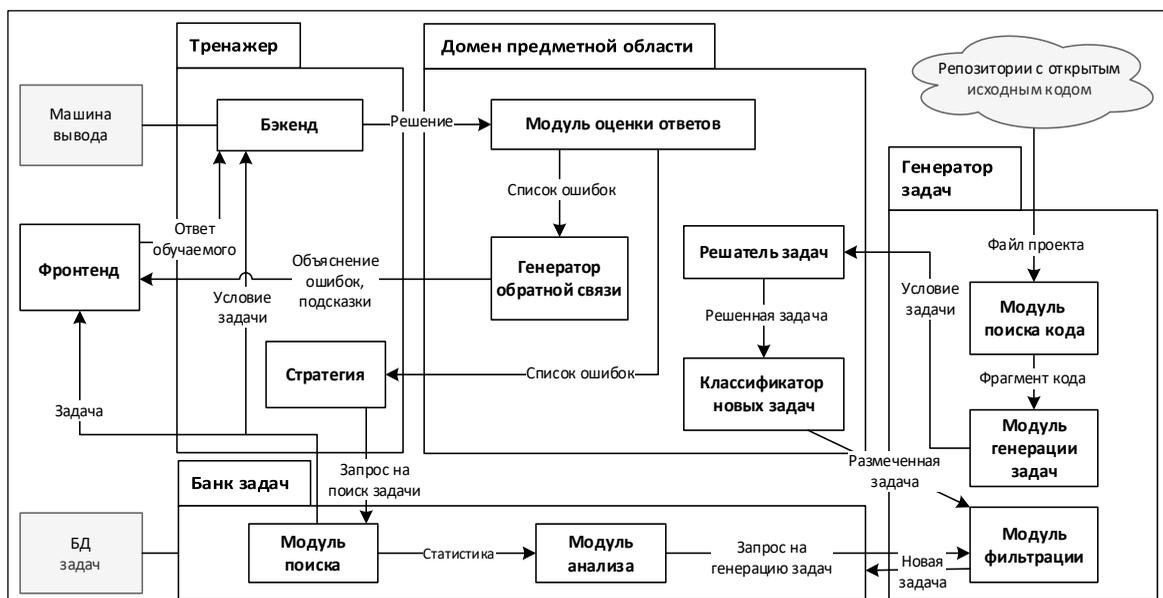


Рис. 2. Основные компоненты обучающей системы CompPrehension с генератором заданий

Fig. 2. Main components of the CompPrehension tutoring system with a learning-problem generator

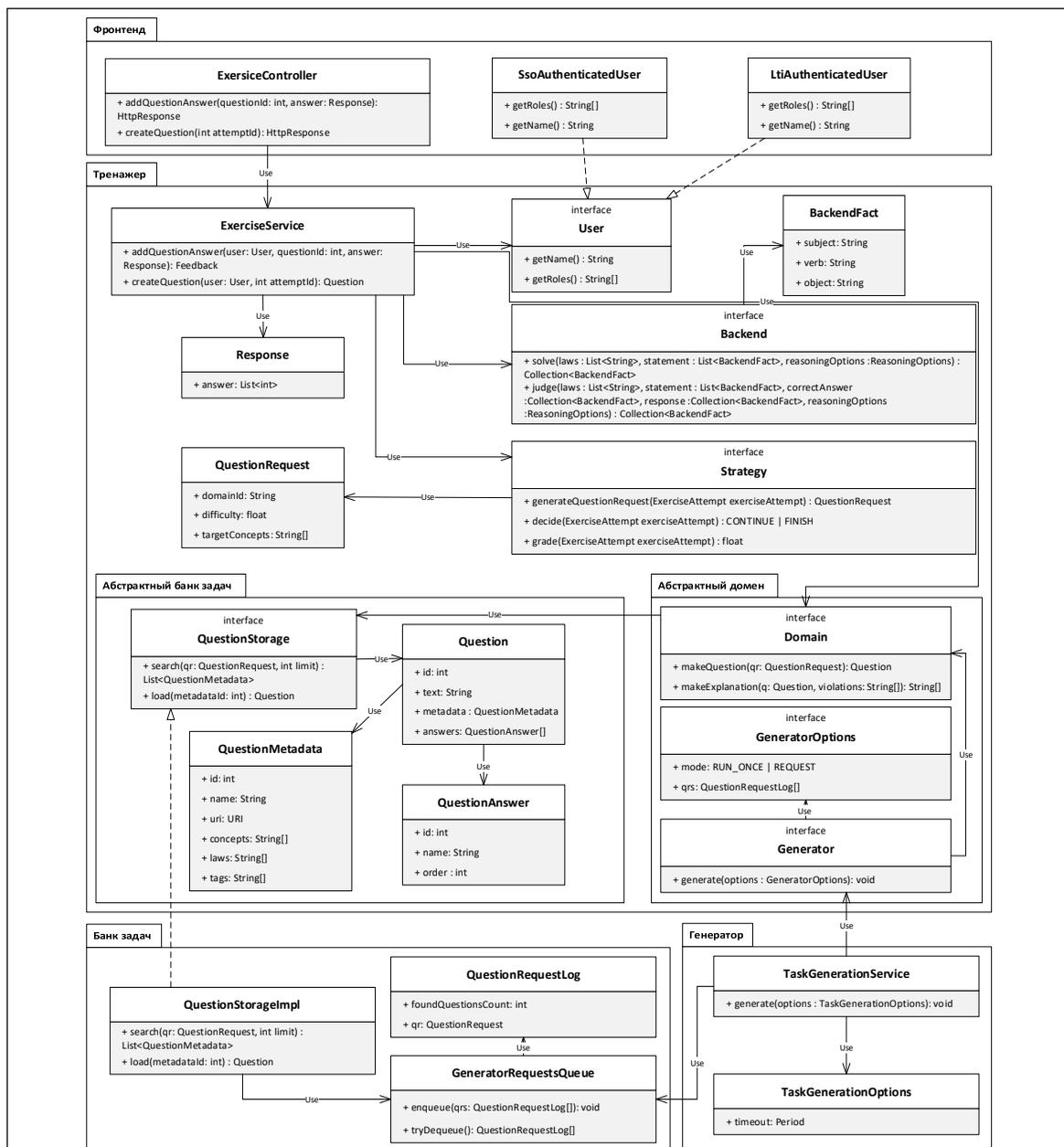


Рис. 3. Частичная диаграмма классов *CompPrehension* с генератором заданий  
 Fig. 3. Partial class diagram of *CompPrehension* with a learning-problem generator

идет разработка доменов «Области видимости», «Выражения доступа» и «Порядок прилагательных в английском языке».

Стратегия отделяет педагогический подход к обучению от особенностей предметной области. Она определяет требования к следующему заданию упражнения, учитывая как настройки этого упражнения, так и данные об успехах и ошибках обучающегося, а также решает, когда упражнение можно считать завершенным (рис. 3, интерфейс *Strategy* и класс *QuestionRequest*). Стратегия вычисляет оценку за ответ

и упражнения на основе данных домена и определяет, какие виды обратной связи доступны обучающему. Стратегии могут представлять фиксированное количество заданий (например, в целях оценки знаний на экзамене) или быть адаптивными к ответам обучающегося.

Интерфейс *Backend* (рис. 3) инкапсулирует взаимодействие с машинами вывода – логическими решателями, которые используют правила и факты, предоставляемые доменами. Такие интерфейсы используются для получения правильных ответов на задания домена и

определяют ошибки в ответах учащихся. Примерами подключаемых машин вывода могут служить Apache Jena, Prolog и др. Использование возможностей символьного искусственного интеллекта позволяет учить машину рассуждать так же, как и обучаемый, и находить объяснения ошибок, которые повышают доверие к решениям системы согласно концепции объяснимого искусственного интеллекта.

Публичный интерфейс тренажера – фронтенд – инкапсулирует в себе основные сценарии взаимодействия с ним пользователей. Он дает возможность организовать взаимодействие с тренажером как через собственный сайт, так и с помощью сторонних систем дистанционного обучения (поддерживается протокол LTI – Learning Tools Interoperability), а в перспективе позволит работать через мобильные приложения и чат-боты (рис. 2, компонент «Фронтенд»).

Банк задач позволяет хранить и наполнять базу учебных задач, а также выполнять поиск задач по определенным критериям. Концептуально банк задач состоит из модуля хранения и модуля анализа.

Модуль хранения содержит условия и метаданные для каждого учебного задания. Для удобства обновления банка данные каждого задания хранятся в одном файле, а его метаданные (учебные свойства и служебная информация) занесены в индексированную SQL-таблицу. Модуль позволяет выполнять быстрый поиск по большому количеству заданий средствами языка запросов SQL.

Модуль анализа проводит мониторинг статистики поисковых запросов и найденного для них количества подходящих вопросов. В случае нахождения запросов, число подходящих заданий для которых ниже порогового, а в результатах есть часто используемые задания, создается запрос на генерацию (рис. 3, классы GeneratorRequestQueue и QuestionRequestLog).

Генератор учебных заданий является отдельным компонентом и работает в фоновом режиме. Он предназначен для автоматизации большого цикла обратной связи обучающей системы и непосредственно взаимодействует с банком заданий и доменом. Реализованы два подхода к генерации – первоначальная генерация (по ручному триггеру) и автоматическая (по запросу на генерацию). Когда генератор получает запрос на генерацию, он передает его соответствующему домену. Домен анализирует запрос и самостоятельно принимает решение о начале генерации, процессом которой он

управляет (рис. 2, компоненты «Генератор» и «Абстрактный домен»).

Для существующих доменов, посвященных изучению программирования, реализован процесс генерации, в котором в качестве источника данных для шаблонов задач используется открытый программный код, размещенный на портале github.com. Из загружаемых API-репозитория отбираются файлы на целевом языке программирования, из которых извлекаются необходимые для генерации заданий фрагменты (например, отдельные выражения или функции). Фрагменты кода преобразуются в шаблоны заданий. Домен расширяет шаблон до готового задания или набора заданий, рассчитывает правильный ответ (множество ответов) для него и вычисляет учебные свойства задания, в частности, отделяет фактически используемые в процессе решения концепты от просто фигурирующих в условии.

Преподаватель, использующий такую систему, может создать упражнение очень быстро, если для него разработан домен: достаточно указать свойства желаемого упражнения. Для этого все концепты и правила предметной области (домена) делятся на три класса: целевые (которые должны развиваться или оцениваться в рамках данного упражнения), запрещенные (обучаемый не должен их знать на данном этапе) и допустимые (не являются целью упражнения, но их знание ожидается от обучаемых – например, в результате изучения предыдущих тем). Задаются ожидаемая сложность упражнения, от которой зависит, например, среднее количество различных концептов и правил в одном задании, а также сложность структуры объектов в условии задачи. В общем случае простые упражнения будут чередоваться с заданиями на различные целевые концепты (правила), а сложные – комбинировать много концептов (правил) в каждом задании упражнения. Кроме того, выбирается желаемая педагогическая стратегия из доступных в системе (например, фиксированное число заданий или адаптивное выполнение до достижения системой уверенности в знаниях обучаемого). Преподаватель может опробовать различные режимы настроек с помощью просмотра созданного упражнения.

Таким образом, задав настройки на одном экране, преподаватель может получить упражнение с требуемыми свойствами, при этом, если в банке недостаточно задач с нужными свойствами, они будут добавлены по мере необходимости.

Замыкание большого цикла обратной связи требует серьезной работы от разработчиков домена и генератора задач для каждой предметной области, но эта работа проводится однократно для каждого нового вида задания. Она снимает большую нагрузку по составлению заданий с пользователей-преподавателей.

### Результаты апробации подхода и дискуссия

Для проверки предлагаемого подхода исследовано поведение генератора заданий. Основной целью эксперимента являлась проверка способности выполнять балансировку банка заданий. Для этого путем первичной генерации получен несбалансированный банк заданий, который затем был дополнен в процессе фоновой генерации при использовании в учебном процессе Волгоградского государственного технического университета.

В настройках поиска репозитория генератора указан язык программирования Си, размер репозитория – от 50 КБ до 100 МБ, исключены клоны репозитория и выполнена сортировка по убыванию пользовательского рейтинга.

В ходе начального заполнения получено 5 655 заданий на порядок вычисления выражений. Затем тренажер эксплуатировался в учебном процессе и накапливал статистику по видам требуемых заданий: сохранено 23 запроса на поиск задания, среди которых 11 суще-

ственно различных по составу целевых концептов, для двух из которых найдено менее 500 заданий, для девяти – менее 5 000 заданий. После этого была запущена фоновая генерация.

На первом этапе порог начала генерации и уровень остановки были заданы равными 500 заданиям. Согласно этой настройке в генерации участвовали только два запроса, которым соответствовало менее 200 заданий (см. таблицу). После работы генератора около суток для каждого из двух запросов найдено достаточно заданий, запросы помечены генератором как удовлетворенные и перемещены в архив, после чего генерация была остановлена.

На втором этапе, чтобы задействовать больше запросов, был увеличен порог генерации необходимых заданий в 10 раз, из-за чего девять запросов начали обрабатываться. Два запроса из предыдущего этапа в генерации не участвовали, так как они были архивированы, а новых запросов с недостатком не поступало (эти цели обучения не были актуализированы в процессе эксплуатации). Время работы генератора на втором этапе составило 4–5 суток. Для большинства запросов успешно сгенерировалось достаточно заданий, чтобы дополнить количество подходящих до 5 000. Эти запросы также были архивированы. Два запроса остались не полностью удовлетворенными, однако количество заданий по ним было сильно увеличено (см. таблицу). За более длительное время генерация может быть доведена до конца.

### Результаты дополнительной догенерации заданий

#### Results of additional learning-problem generation

Семейство целевых концептов	Заданий в банке до генерации	Количество обработанных репозитория	Заданий добавлено в банк	Заданий в банке после генерации
<i>Первый этап (для порога в 500 заданий)</i>				
Присваивания с обновлением (+ = &= )	143	24	361	504
Инкремент ( ++ -- )	168	32	351	519
<i>Второй этап (для порога в 5 000 заданий)</i>				
Оператор (=)	2 073	37	2 985	5 058
Оператор (,)	972	*114	2 960	*3 932
Арифметика (+ - * / %)	2 086	33	2 937	5 023
Побитовые (&   >> <<)	762	*92	2 897	*3 659
Индексация ([ ])	2 520	33	2 508	5 028
Сравнения (< == != >)	2 824	35	2 278	5 102
Доступ к полю (. ->)	2 985	17	2 034	5 019
Указатели (* _ ->)	3 131	14	1 981	5 112
Логические (! &&   )	3 407	43	1 597	5 004

Примечание: \* – генерация была остановлена вручную.

За время работы генератора было скачано 1 398 репозиторий, 526 из которых генератор смог использовать, чтобы сгенерировать и отправить в банк тренажера 14 479 новых заданий на порядок вычисления выражений для шести запросов. Число заданий в банке возросло с 5 655 до 20 134. Из одного репозитория было сгенерировано от 1 до 640 заданий. Наиболее редко встречающиеся задания включали такие ситуации (правила), как наличие оператора внутри скобочного оператора (обращение к элементу массива, вызов функции) и двух подряд идущих операторов с одинаковым приоритетом и правой ассоциативностью.

### Заключение

В статье описан подход к построению ИОС. Проанализированы подходы к генерации учебных заданий и требования к системе, позволяющей автоматически генерировать учебные задания и использовать их в процессе обучения.

Описана архитектура ИОС с фоновой генерацией заданий в соответствии с возникающей необходимостью, рассмотрена система CompPrehension, реализующая этот принцип на практике. В ходе апробации системы в учебном процессе университета изначально сгенерированный несбалансированный банк заданий был дополнен заданиями по темам, на которые заданий было мало, что позволило избежать выдачи одинаковых заданий даже при значительном количестве обучаемых.

В дальнейшем планируется расширить список поддерживаемых предметных областей, улучшить точность поиска заданий в банке и реализовать генерацию новых заданий в областях, не связанных с обучением программированию. Результаты исследования имеют практическую значимость для разработки ИОС и учебных заведений, которые стремятся повысить массовость и разнообразие учебных задач.

### Список литературы

1. Qi R., Fossati D. Unlimited trace tutor: Learning code tracing with automatically generated programs. Proc. SIGCSE, 2020, pp. 427–433. doi: 10.1145/3328778.3366939.
2. D'Antoni L., Helfrich M., Kretinsky J., Ramneantu E., Weininger M. Automata tutor v3. In: LNTCS. Proc. CAV, 2020, vol. 12225, pp. 3–14. doi: 10.1007/978-3-030-53291-8\_1.
3. Sychev O., Prokudin A., Denisov M. Generation of code tracing problems from open-source code. Proc. SIGCSE, 2023, pp. 875–881. doi: 10.1145/3545945.3569774.
4. Kumar A.N. Long term retention of programming concepts learned using a software tutor. In: LNPSE. Proc. ITS, 2020, vol. 12149, pp. 382–387. doi: 10.1007/978-3-030-49663-0\_46.
5. Razzaq L., Patvarczki J., Almeida S.F., Vartak M., Feng M., Heffernan N.T., Koedinger K.R. The ASSISTment builder: Supporting the life cycle of tutoring system content creation. IEEE Transactions on Learning Tech., 2009, vol. 2, no. 2, pp. 157–166. doi: 10.1109/TLT.2009.23.
6. Juričić V. Software visualization in education. INFUTURE2019. Proc. INFUTURE2019: Knowledge in the Digital Age, 2020, pp. 216–220. doi: 10.17234/INFUTURE.2019.26.
7. Kurdi G., Leo J., Parsia B., Sattler U., Al-Emari S. A systematic review of automatic question generation for educational purposes. IJAIED, 2019, vol. 30, pp. 121–204. doi: 10.1007/s40593-019-00186-y.
8. Rusak G., Yan L. Unique exams: Designing assessments for integrity and fairness. Proc. SIGCSE, 2021, pp. 1170–1176. doi: 10.1145/3408877.3432556.
9. Mosbeck M., Hauer D., Jantsch A. VELs: VHDL E-learning system for automatic generation and evaluation of per-student randomized assignments. Proc. NORCAS Conf.: SoC, 2018, pp. 1–7. doi: 10.1109/NORCHIP.2018.8573455.
10. Жуков И.А., Костюк Ю.Л. Предметно-ориентированный язык для генерации заданий из исходных текстов программ // Вестн. НГУ. Сер. Информационные технологии. 2022. Т. 20. № 1. С. 18–27.
11. Ade-Ibijola A. Syntactic generation of practice novice programs in python. In: CCIS. Proc. SACLA, 2019, vol. 963, pp. 158–172. doi: 10.1007/978-3-030-05813-5\_11.
12. Kulshreshtha D., Shayan M., Belfer R., Reddy S., Serban I.V., Kochmar E. Few-shot question generation for personalized feedback in intelligent tutoring systems. In: PAIS, 2022, vol. 351, pp. 17–30. doi: 10.3233/FAIA220062.
13. Das B., Majumder M., Phadikar S., Sekh A.A. Automatic question generation and answer assessment: A survey. RPTEL, 2021, vol. 16, art. 5. doi: 10.1186/s41039-021-00151-1.
14. Atzeni M., Atzori M. CodeOntology: RDF-ization of source code. In: LNISA. Proc. ISWC, 2017, vol. 10588, pp. 20–28. doi: 10.1007/978-3-319-68204-4\_2.
15. Maheen F., Asif M., Ahmad H. et al. Automatic computer science domain multiple-choice questions generation based on informative sentences. PeerJ Comput. Sci., 2022, vol. 8, art. e1010. doi: 10.7717/PEERJ-CS.1010.
16. Sychev O., Pensky N., Prokudin A. Generating expression evaluation learning problems from existing program code. Proc. ICALT, 2022, pp. 183–187. doi: 10.1109/ICALT55010.2022.00061.
17. Hahn M.G., Navarro S.M.B., De La Fuente V., Burgos D. A systematic review of the effects of automatic scoring and automatic feedback in educational settings. IEEE Access, 2021, vol. 9, pp. 108190–108198. doi: 10.1109/ACCESS.2021.3100890.

18. Bernius J.P., Krusche S., Bruegge B. Machine learning based feedback on textual student answers in large courses. *Computers and Education: Artificial Intelligence*, 2022, vol. 3, art. 100081. doi: 10.1016/j.caeai.2022.100081.
19. Narayanan S., Kommuri V.S., Subramanian S.N., Bijlani K. Question bank calibration using unsupervised learning of assessment performance metrics. *Proc. ICACCI*, 2017, pp. 19–25. doi: 10.1109/ICACCI.2017.8125810.
20. Sychev O., Anikin A., Penskov N., Denisov M., Prokudin A. CompPrehension – model-based intelligent tutoring system on comprehension level. In: *LNPSE. Proc. ITS*, 2021, vol. 12677, pp. 52–59. doi: 10.1007/978-3-030-80421-3\_6.
21. Prokudin A., Sychev O., Denisov M. Learning problem generator for introductory programming courses. *Software Impacts*, 2023, vol. 17, art. 100519. doi: 10.1016/j.simpa.2023.100519.

Software &amp; Systems

doi: 10.15827/0236-235X.142.201-212

2024, 37(2), pp. 201–212

### Intelligent tutoring system with automatic assignment generation according to the results of using an assignment bank

Oleg A. Sychev , Artem A. Prokudin <sup>1</sup>, Mikhail E. Denisov <sup>1</sup><sup>1</sup> Volgograd State Technical University, Volgograd, 400005, Russian Federation

#### For citation

Sychev, O.A., Prokudin, A.A., Denisov, M.E. (2024) 'Intelligent tutoring system with automatic assignment generation according to the results of using an assignment bank', *Software & Systems*, 37(2), pp. 201–212 (in Russ.). doi: 10.15827/0236-235X.142.201-212

#### Article info

Received: 18.10.2023

After revision: 20.12.2023

Accepted: 22.12.2023

**Abstract.** An educational process includes two feedback loops: learner feedback directs actual learning for a given individual (small loop), while teacher feedback about the quality of the bank of learning problems (big loop) allows improving a learning process in time, increasing the number of assignments which are often used repeatedly and so learners can share solutions. Automating the big feedback loop is an important step in developing intelligent tutoring systems, which grants course authors relief from routine work on creating and enhancing learning assignment banks. Modern advances in methods of learning assignment generation and their classification for using in a learning process allows implementing an intelligent tutoring system, which can generate new learning assignments in the background mode if necessary. The paper describes the architecture of an intelligent tutoring system with background learning assignment generation and the CompPrehension system, which implements the mentioned architecture. We describe all major functional system components: a training aid, a learning assignment bank, an assignment generator, a domain; and their interaction with the assignment database and external semantic reasoners for solving learning assignments. Practical evaluation of the described approach showed that the system can selectively improve the initial unbalanced learning assignment bank, taking into account the history of assignment requests. In less than 5 days, the system has generated several thousand additional learning assignments for each concept with low learning assignment count. This gives low probability of receiving the same assignments even when testing hundreds of students. The study results are significant for educational institutions seeking to improve the learning process.

**Keywords:** intelligent tutoring systems, question generation, online learning, learning system development, CompPrehension, learning assignment bank

#### References

1. Qi, R., Fossati, D. (2020) 'Unlimited trace tutor: Learning code tracing with automatically generated programs', *Proc. SIGCSE*, pp. 427–433. doi: 10.1145/3328778.3366939.
2. D'Antoni, L., Helfrich, M., Kretinsky, J., Ramneantu, E., Weininger, M. (2020) 'Automata tutor v3', in *LNTCS. Proc. CAV*, 12225, pp. 3–14. doi: 10.1007/978-3-030-53291-8\_1.
3. Sychev, O., Prokudin, A., Denisov, M. (2023) 'Generation of code tracing problems from open-source code', *Proc. SIGCSE*, pp. 875–881. doi: 10.1145/3545945.3569774.
4. Kumar, A.N. (2020) 'Long term retention of programming concepts learned using a software tutor', in *LNPSE. Proc. ITS*, 12149, pp. 382–387. doi: 10.1007/978-3-030-49663-0\_46.
5. Razzaq, L., Patvarczki, J., Almeida, S.F., Vartak, M., Feng, M., Heffernan, N.T., Koedinger, K.R. (2009) 'The ASSISTment builder: Supporting the life cycle of tutoring system content creation', *IEEE Transactions on Learning Tech.*, 2(2), pp. 157–166. doi: 10.1109/TLT.2009.23.

6. Juričić, V. (2020) 'Software visualization in education. INFUTURE2019', *Proc. INFUTURE2019: Knowledge in the Digital Age*, pp. 216–220. doi: 10.17234/INFUTURE.2019.26.
7. Kurdi, G., Leo, J., Parsia, B., Sattler, U., Al-Emari, S. (2019) 'A systematic review of automatic question generation for educational purposes', *IJAIED*, 30, pp. 121–204. doi: 10.1007/s40593-019-00186-y.
8. Rusak, G., Yan, L. (2021) 'Unique exams: Designing assessments for integrity and fairness', *Proc. SIGCSE*, pp. 1170–1176. doi: 10.1145/3408877.3432556.
9. Mosbeck, M., Hauer, D., Jantsch, A. (2018) 'VELS: VHDL E-learning system for automatic generation and evaluation of per-student randomized assignments', *Proc. NORCAS Conf.: SoC*, pp. 1–7. doi: 10.1109/NORCHIP.2018.8573455.
10. Zhukov, I.A., Kostyuk, Yu.L. (2022) 'A domain-specific language for generating tasks from programs source code', *Bull. of NSU. Ser.: Inform. Tech.*, 20(1), pp. 18–27 (in Russ.). doi: 10.25205/1818-7900-2022-20-1-18-27.
11. Ade-Ibijola, A. (2019) 'Syntactic generation of practice novice programs in python', in *CCIS. Proc. SACLA*, 963, pp. 158–172. doi: 10.1007/978-3-030-05813-5\_11.
12. Kulshreshtha, D., Shayan, M., Belfer, R., Reddy, S., Serban, I.V., Kochmar, E. (2022) 'Few-shot question generation for personalized feedback in intelligent tutoring systems', in *PAIS*, 351, pp. 17–30. doi: 10.3233/FAIA220062.
13. Das, B., Majumder, M., Phadikar, S., Sekh, A.A. (2021) 'Automatic question generation and answer assessment: A survey', *RPTTEL*, 16, art. 5. doi: 10.1186/s41039-021-00151-1.
14. Atzeni, M., Atzori, M. (2017) 'CodeOntology: RDF-ization of source code', in *LNISA. Proc. ISWC*, 10588, pp. 20–28. doi: 10.1007/978-3-319-68204-4\_2.
15. Maheen, F., Asif, M., Ahmad, H. et al. (2022) 'Automatic computer science domain multiple-choice questions generation based on informative sentences', *PeerJ Comput. Sci.*, 8, art. e1010. doi: 10.7717/PEERJ-CS.1010.
16. Sychev, O., Penskoy, N., Prokudin, A. (2022) 'Generating expression evaluation learning problems from existing program code', *Proc. ICALT*, pp. 183–187. doi: 10.1109/ICALT55010.2022.00061.
17. Hahn, M.G., Navarro, S.M.B., De La Fuente, V., Burgos, D. (2021) 'A systematic review of the effects of automatic scoring and automatic feedback in educational settings', *IEEE Access*, 9, pp. 108190–108198. doi: 10.1109/ACCESS.2021.3100890.
18. Bernius, J.P., Krusche, S., Bruegge, B. (2022) 'Machine learning based feedback on textual student answers in large courses', *Computers and Education: Artificial Intelligence*, 3, art. 100081. doi: 10.1016/j.caeai.2022.100081.
19. Narayanan, S., Kommuri, V.S., Subramanian, S.N., Bijlani, K. (2017) 'Question bank calibration using unsupervised learning of assessment performance metrics', *Proc. ICACCI*, pp. 19–25. doi: 10.1109/ICACCI.2017.8125810.
20. Sychev, O., Anikin, A., Penskoy, N., Denisov, M., Prokudin, A. (2021) 'CompPrehension – model-based intelligent tutoring system on comprehension level', in *LNPSE. Proc. ITS*, 12677, pp. 52–59. doi: 10.1007/978-3-030-80421-3\_6.
21. Prokudin, A., Sychev, O., Denisov, M. (2023) 'Learning problem generator for introductory programming courses', *Software Impacts*, 17, art. 100519. doi: 10.1016/j.simpa.2023.100519.

**Авторы**

**Сычев Олег Александрович**<sup>1</sup>, к.т.н.,  
доцент, oasychev@gmail.com

**Прокудин Артем Александрович**<sup>1</sup>,  
аспирант, prokudin@vstu.ru

**Денисов Михаил Евгеньевич**<sup>1</sup>,  
аспирант, denisov@vstu.ru

**Authors**

**Oleg A. Sychev**<sup>1</sup>, Dr.Sc. (Engineering),  
Associate Professor, oasychev@gmail.com

**Artem A. Prokudin**<sup>1</sup>,  
Postgraduate Student, prokudin@vstu.ru

**Mikhail E. Denisov**<sup>1</sup>,  
Postgraduate Student, denisov@vstu.ru

<sup>1</sup> Волгоградский государственный  
технический университет,  
г. Волгоград, 400005, Россия

<sup>1</sup> Volgograd State Technical University,  
Volgograd, 400005,  
Russian Federation

## Инжиниринг цифрового тренажера для обучения операторов формования листового стекла

В.П. Мешалкин<sup>1,2</sup>, Т.Б. Чистякова<sup>2</sup>, Д.Ю. Петров<sup>3,4</sup>✉

<sup>1</sup> Российский химико-технологический университет имени Д.И. Менделеева, г. Москва, 125047, Россия

<sup>2</sup> Санкт-Петербургский государственный технологический институт (технический университет), г. Санкт-Петербург, 190013, Россия

<sup>3</sup> Институт проблем точной механики и управления РАН, г. Саратов, 410028, Россия

<sup>4</sup> Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского, г. Саратов, 410012, Россия

### Ссылка для цитирования

Мешалкин В.П., Чистякова Т.Б., Петров Д.Ю. Инжиниринг цифрового тренажера для обучения операторов формования листового стекла // Программные продукты и системы. 2024. Т. 37. № 2. С. 213–220. doi: 10.15827/0236-235X.142.213-220

### Информация о статье

Группа специальностей ВАК: 2.3.3

Поступила в редакцию: 31.03.2023

После доработки: 27.01.2024

Принята к публикации: 26.02.2024

**Аннотация.** В статье проведен анализ формования листового стекла флоат-методом, определены виды нештатных ситуаций технологического процесса. Выполнен кластерный анализ множества нештатных ситуаций, каждая из которых описана как теоретико-множественная модель. Для упрощения интерфейса цифрового тренажера разработаны модели бизнес-процессов для обучения и аттестации операторов. На основе анализа существующих бизнес-процессов предприятия в нотации BPMN 2.0 разработаны логико-информационные модели перспективных бизнес-процессов Разработка тестов и Аттестация сотрудников. При разработке архитектуры, режимов функционирования и программно-информационного обеспечения цифрового тренажера использована методика model-based system engineering, реализованная в среде визуального моделирования IBM Rhapsody на основе нотаций UML и SysML. Разработаны диаграммы вариантов использования и диаграммы классов цифрового тренажера с использованием языка UML, а также структура БД системы обучения, включающая критерии группировки ситуаций, характеристики ситуаций, функции пользователей системы, и структура ПО. Реализованы основные функции – обучение и тестирование операторов, ведение БД персонала цифрового тренажера, составление отчетов по результатам тестирования, редактирование БД по аварийным ситуациям на основе предварительного анализа опыта технологов и операторов. Цифровой тренажер разработан на языке PascalABC, для записи результатов использована текстовая БД. Разработан пользовательский интерфейс цифрового тренажера.

**Ключевые слова:** цифровой тренажер, обучение, бизнес-процесс, формование листового стекла, оператор, диаграмма вариантов использования, диаграмма классов, нештатная ситуация

**Благодарности.** Работа проведена в ИПТМУ РАН при выполнении госзадания Минобрнауки РФ, тема № FFNM-2022-0010 (разработка диаграмм, структуры и ПО цифрового тренажера), и при поддержке гранта РНФ, проект № 21-79-30029 (описание бизнес-процессов)

**Введение.** Для снижения риска ошибочных действий персонала при авариях и нештатных ситуациях на производстве необходимы специальные цифровые тренажеры, позволяющие операторам после обучения на них своевременно выбирать и выполнять регламентированные действия по переводу *технологического процесса* (ТП) из аварийного в штатный режим [1]. Цифровизированное управление рисками и безопасностью при проектировании энерго- и ресурсоэффективных химических производств обеспечивается за счет использования методов оптимизации показателей надежности [2]. На основе системного анализа опыта действий операторов в нештатных ситуациях (ухудшение качества, аварийные ситуации) разрабатываются математические модели

для их обнаружения на ранних стадиях [3, 4]. Математическое моделирование ТП производства листового стекла осуществляют специализированные программные комплексы GFM (Glass Furnace Model), NOGRID CFD и другие, а также универсальные программные комплексы Логос, ANSYS, COMSOL Multiphysics и т.п. Для повышения квалификации операторов необходимо использовать специальные цифровые тренажеры, интегрированные с информационными системами предприятия [5–7].

Высокоэффективное функционирование предприятий стекольной промышленности обеспечивается сокращением отходов формируемого стекла и затрат на производство при повышении качества выпускаемой продукции [8]. Увеличение количества отходов стекла во многом

связано с возникновением нештатных ситуаций, заканчивающихся аварийным обрывом ленты стекла [9], в результате производство несет убытки. Внедрение новых технологий и техники выполняют при капитальном ремонте производства листового стекла. Это обеспечивает переход производства на новый уровень производительности и обуславливает необходимость систематически повышать квалификацию технологов и операторов.

Перспективным направлением в области обучения и контроля знаний технологического персонала химических предприятий является использование цифровых тренажеров и инструментов Индустрии 4.0 [10, 11].

Для обучения операторов разработан цифровой тренажер. Определены параметры штатных режимов и основные нештатные ситуации, возникающие при формовании ленты стекла флоат-методом в ванне с расплавом олова.

Каждая нештатная ситуация описана как теоретико-множественная модель  $E = \langle N, C, I, J, R, S \rangle$ , где  $N$  – название аварийной ситуации;  $C$  – множество признаков аварийной ситуации;  $I$  – множество изображений пульта оператора, отображающих развитие аварийной ситуации с течением времени;  $J$  – ванна расплава общего вида;  $R$  – множество причин возникновения аварийной ситуации;  $S$  – последовательность действий оператора по ликвидации данной аварийной ситуации.

Для облегчения процессов обучения и аттестации операторов формования ленты стекла и упрощения интерфейса цифрового тренажера множество аварийных ситуаций разделено на подмножества (группы). Такое разделение проведено на основе кластерного анализа с использованием пакета прикладных программ NCSS. Получены пять кластеров (групп) нештатных ситуаций (см. таблицу).

Разработаны логико-информационные модели бизнес-процессов в текущем состоянии в нотации Business Process Model and Notation (BPMN) 2.0. На рисунке 1 в нотации BPMN 2.0 показаны логико-информационные модели бизнес-процессов Разработка тестов и Аттестация сотрудников. Представлены исполнители единиц деятельности в этих процессах – Главный технолог, Эксперт-технолог, Администратор комплекса для аттестации, Кадровик, Оператор формования листового стекла. Бизнес-процессы взаимодействуют с БД: 1С.Кадры, АСУТП, База тестов, Результаты обучения. Стартовыми событиями являются Утверждение нового технического регламента, Заверше-

ние разработки новых тестов, Прием оператора формования листового стекла на работу, Наступление срока переаттестации сотрудников.

Анализ логико-информационной модели бизнес-процессов показал, что тренажер должен функционировать в трех режимах: Оператор, Администратор, Эксперт. Выбор режима осуществляется на основе меню прав выбранного имени пользователя и пароля.

В режиме Оператор предусмотрены обучение операторов и их аттестация. Аттестация проводится двумя способами: по изображению ситуации и ее описанию оператор должен определить название ситуации, причины возникновения (для нештатных) и последовательность действий по переводу процесса в штатный режим или для заданной ситуации определяется последовательность действий конкретного специалиста с учетом синхронизации по времени его действий с действиями других специалистов в данном ТП.

В режиме Администратор осуществляются ведение БД пользователей (обучаемого персонала) и формирование отчетов по аттестации.

В режиме Эксперт заполняется и редактируется БД аварийных ситуаций, возникающих при формовании листового стекла методом флоат-процесса, с учетом изменения регламентов ТП и анализа практического опыта действий по устранению нештатных ситуаций и дефектов.

При разработке ПО цифрового тренажера использована методика MBSE (model-based system engineering), позволяющая выполнять управление требованиями, проектирование, анализ, верификацию и валидацию на протяжении всего жизненного цикла. Методика MBSE реализована в среде визуального моделирования IBM Rhapsody на основе нотаций UML и SysML.

На основе анализа требований определены варианты использования цифрового тренажера для разработки тестов и базы тестов, редактирования тестов, подготовки обучения и тестирования, анализа обучения и тестирования, отчета об обучении и тестировании, разработки базы обучения персонала, обучения оператора, тестирования оператора. С цифровым тренажером взаимодействуют акторы: оператор, эксперт, администратор, ПО 1С.Кадры и АСУТП.

На рисунке 2 показана диаграмма вариантов использования цифрового тренажера для обучения операторов формования листового стекла.

На основе анализа диаграммы вариантов использования разработана структура классов

## Кластеры нештатных ситуаций

## Clusters of emergency situations

Номер кластера	Основная причина неисправности	Нештатная ситуация
1	Прекращение подачи средств обеспечения формования листового стекла (электроэнергия, природный газ, вода, защитная атмосфера ванны с расплавом олова)	Кратковременное отключение электроэнергии
		Длительное отключение электроэнергии
		Полное прекращение подачи электроэнергии
		Отключение газа
		Прекращение подачи защитной атмосферы ванны с расплавом олова
2	Отказы устройств формования стекла	Отскок одного последнего бортоформирующего ролика
		Отскок одного непоследнего бортоформирующего ролика
		Поломка дозирующего шибера
		Поломка отсекающего шибера
		Отключение нагревателя
		Течь холодильника
		Остановка лера печи отжига
3	Прилипание стекла к ванне с расплавом олова	Налипание ленты стекла на бортоформирующий ролик
		Залипание ленты под бортоформирующий ролик
		Залипание стекломассы ко дну
		Залипание ленты стекла к окружке ванны
4	Инородные включения в разливе стекла	Крупное включение по краю ленты
		Крупное включение по центру ленты
		Камень в выработочном канале
5	Нестабильное состояние разлива стекла	Нестабильное состояние лужи

цифрового тренажера: Users – обеспечивает хранение информации о пользователях; Tests – содержит информацию о тестах; Test – подкласс для одного теста; Group – позволяет управлять пользователями; Attestation – обеспечивает процессы обучения и тестирования; Create\_test – обеспечивает создание новых тестов. Диаграмма классов показана на рисунке 3. Для каждого пользователя определяются атрибуты: группа, имя, пароль, производство, ТП и режим ТП.

Для группировки ситуаций определены параметры: производство, ТП, режим ТП, группа ситуаций. Каждая нештатная ситуация в кластере описывается следующими характеристиками: название ситуации, множество признаков ситуации, графическое изображение на пульте оператора и общий вид ТП, отображаю-

щие развитие этой ситуации во времени, множество причин возникновения этой ситуации, последовательность действий персонала по обработке ситуации с учетом должностных обязанностей.

Структура функциональных подсистем цифрового тренажера представлена на рисунке 4.

Оператор обучается на основе последовательного изучения всех нештатных ситуаций, для каждой из которых демонстрируются пульт оператора и общий вид ТП с развитием ситуации во времени, список причин, признаков ситуации, действий по ее устранению, последовательность действий всех специалистов, участвующих в заданной ситуации, с учетом синхронизации по времени и длительности выполнения каждого действия (<https://swsys.ru/uploaded/image/2024-2/Meshalkin.html>).

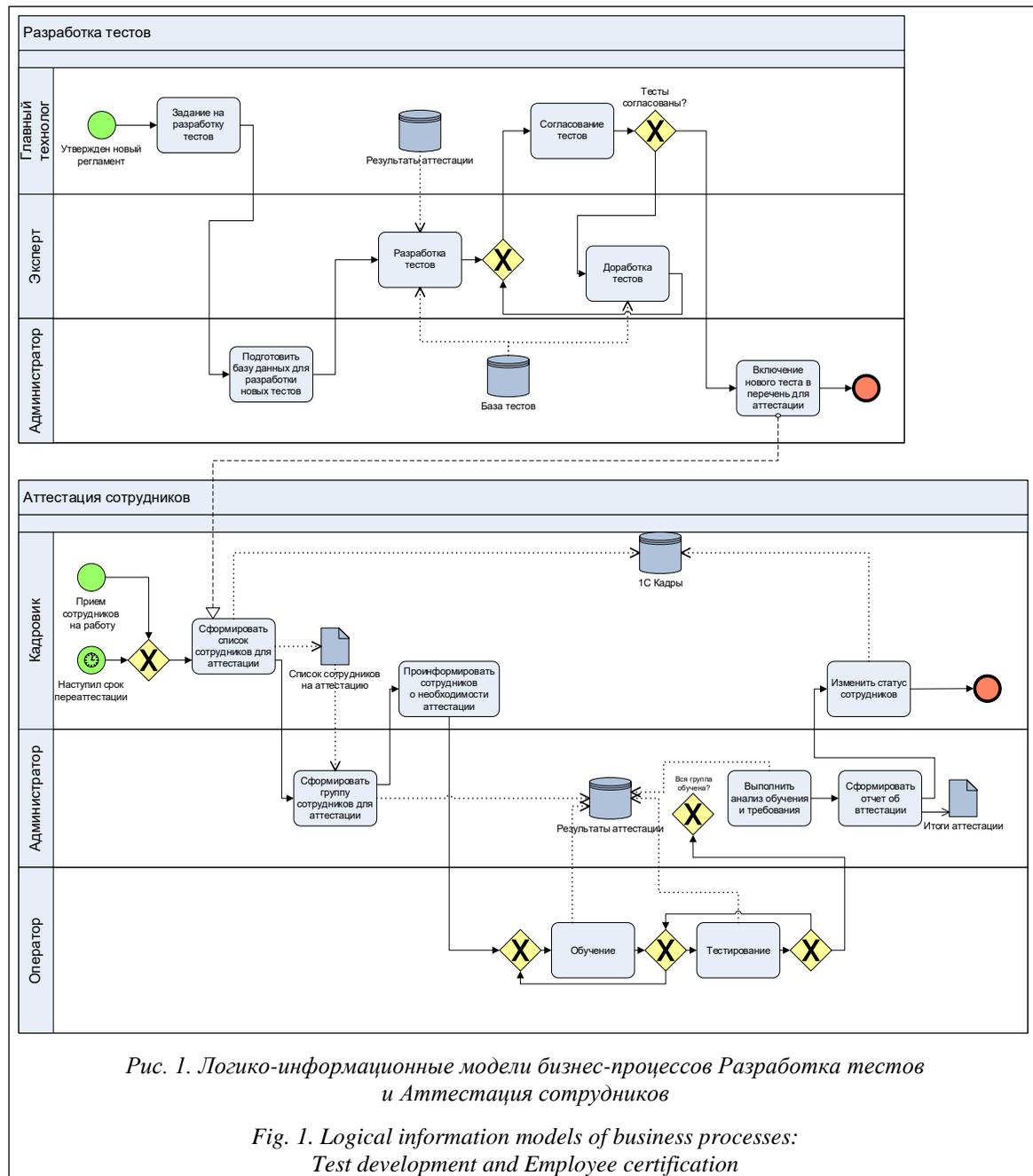


Рис. 1. Логико-информационные модели бизнес-процессов Разработка тестов и Аттестация сотрудников

Fig. 1. Logical information models of business processes: Test development and Employee certification

Тренажер осуществляет аттестацию операторов двумя способами: по изображению ситуации и ее описанию определяются название ситуации, причины возникновения (для нештатных) и последовательность действий по изменению ситуации; для заданной ситуации определяется последовательность действий конкретного специалиста с учетом синхронизации по времени его действий с действиями других специалистов в данном ТП (<http://www.swsys.ru/uploaded/image/2024-2/4.jpg>).

При предотвращении аварийной ситуации или ликвидации ее последствий важное значе-

ние имеет своевременное воздействие на ТП каждого из должностных лиц. Оператор обязан знать последовательность и длительность собственных действий и синхронизировать их с другими специалистами, работающими с данной нештатной ситуацией. Проконтролировать уровень знаний специалиста можно с использованием формы Синхронизация действий. Из списка действий необходимо своевременно выбрать истинные действия в правильной последовательности и в соответствии с должностью.

Цифровой тренажер реализован на языке PascalABC, а для хранения результатов исполь-

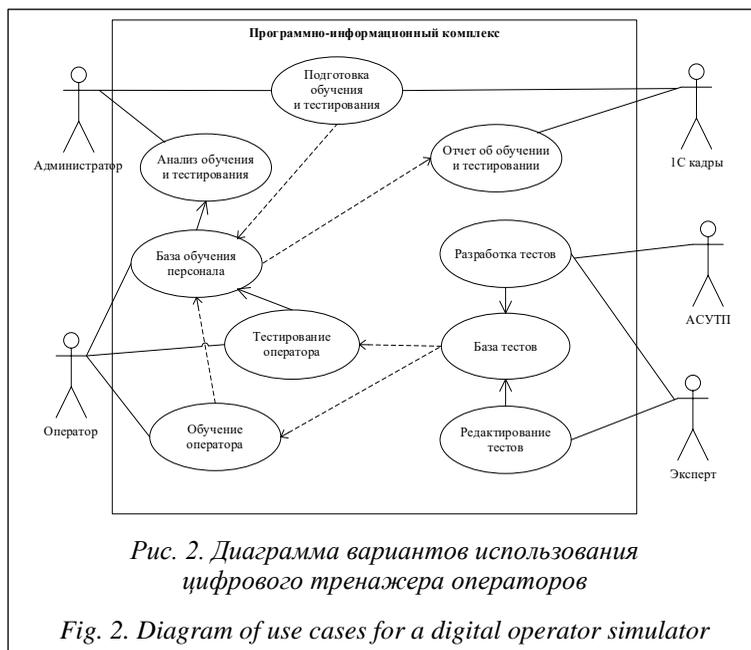


Рис. 2. Диаграмма вариантов использования цифрового тренажера операторов

Fig. 2. Diagram of use cases for a digital operator simulator

ченной продукции, при этом оно характеризуется большим количеством нештатных ситуаций. Для управления таким сложным ТП необходимы квалифицированные операторы, которые по характерным признакам могут идентифицировать нештатные ситуации на ранних этапах возникновения, понимать вызывающие их причины, выполнять регламентированные действия по их устранению.

Обучение и аттестацию операторов стекольного производства обеспечивает предложенный цифровой тренажер. Комплексное применение подходов к созданию тренажера на основе проектирования бизнес-процессов и разработки структуры обеспечения с использованием универсального языка моделирования SysML позволило разработать цифровой тренажер для обучения операторов формирования листового стекла действиям в нештатных ситуациях.

Логико-информационные модели бизнес-процессов в программном комплексе Business Studio дают возможность описать взаимодействие участников процесса повышения квали-

зована текстовая БД. В системе реализовано дерево развития ситуаций, обеспечивающее дополнительный режим обучения, в котором могут быть объединены несколько последовательных аварийных ситуаций.

**Заключение**

При производстве листового стекла предъявляются жесткие требования к качеству полу-

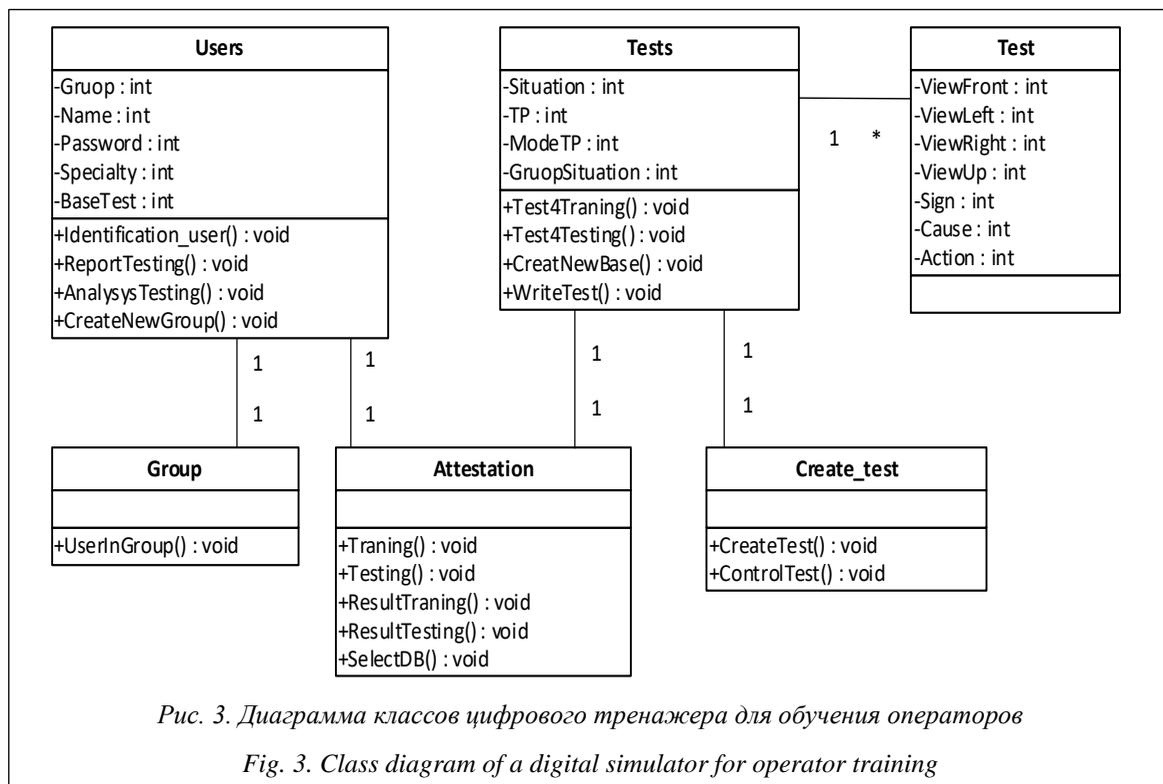


Рис. 3. Диаграмма классов цифрового тренажера для обучения операторов

Fig. 3. Class diagram of a digital simulator for operator training

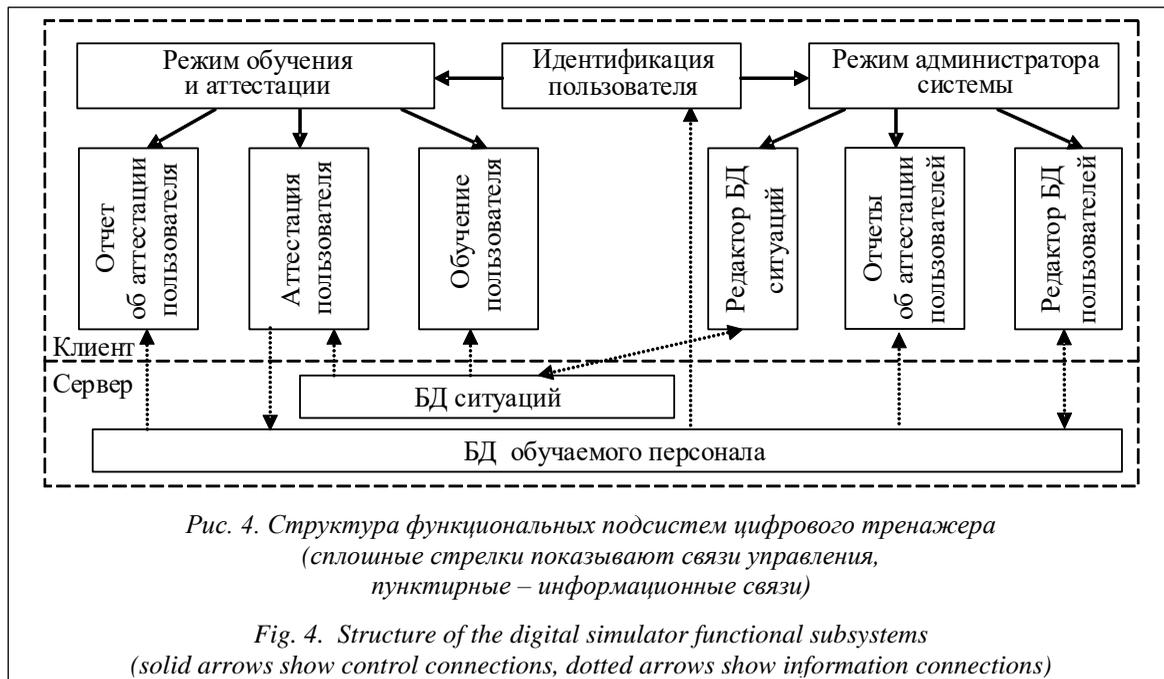


Рис. 4. Структура функциональных подсистем цифрового тренажера (сплошные стрелки показывают связи управления, пунктирные – информационные связи)

Fig. 4. Structure of the digital simulator functional subsystems (solid arrows show control connections, dotted arrows show information connections)

фикации. Эти модели в нотации BPMN 2.0 стали основой для разработки диаграммы вариантов использования в среде визуального моделирования IBM Rhapsody на языке SysML. Структура описания ситуаций ТП формования листового стекла обеспечивает разработку описания ТП варки стекла, его отжига и других процессов, необходимых для обучения персонала. Цифровой тренажер позволяет использовать в качестве обучающего материала анимацию и видеоролики. Это повышает наглядность процесса обучения и улучшает понимание ситуаций, возникающих в различных ТП. Разработанный цифровой тренажер является компонентом внедряемой в настоящее время на предприятии единой информационной системы предприятия. Тренажер обеспечивает возмож-

ность его использования для повышения квалификации производственного персонала промышленных объектов аналогичного класса (производство полимерной пленки, линолеума, бумаги и др.).

В процессе подготовки информации об аварийных ситуациях ТП для наполнения БД цифрового тренажера специалисты предприятия решают главную задачу – более полное понимание истинных причин возникновения нештатных ситуаций, а значит, и разработку новых, менее затратных способов их устранения. Перспективными направлениями развития цифрового тренажера являются разработка функций экспертной системы реального времени и обеспечение поддержки принятия решений производственным персоналом.

#### Список литературы

1. Marcano L., Haugen F.A., Sannerud R., Komulainen T. Review of simulator training practices for industrial operators: How can individual simulator training be enabled? *Safety Sci.*, 2019, vol. 115, pp. 414–424. doi: 10.1016/j.ssci.2019.02.019.
2. Meshalkin V.P., Dovi V.G., Bobkov V.I., Belyakov A.V., Butusov O.B. et al. State of the art and research development prospects of energy and resource-efficient environmentally safe chemical process systems engineering. *Mendelev Communications*, 2021, vol. 31, no. 5, pp. 593–604. doi: 10.1016/j.mencom.2021.09.003.
3. Фураев Д.Н., Чистякова Т.Б. Компьютерная система для обучения ресурсосберегающему проектированию процессов вторичной переработки нефти // *Изв. СПбГТИ(ТУ)*. 2021. № 59. С. 101–106. doi: 10.36807/1998-9849-2021-59-85-101-106.
4. Koteleva N., Kuznetsov V., Vasileva N. A simulator for educating the digital technologies skills in industry. Part one. Dynamic simulation of technological processes. *Appl. Sci.*, 2021, vol. 11, no. 22, art. 10885. doi: 10.3390/app112210885.
5. Moreland J., Toth K., Fang Y., Block M. et al. Interactive simulators for steel industry safety training. *Steel Research Int.*, 2019, vol. 90, no. 4, art. 1800513. doi: 10.1002/srin.201800513.
6. Жуков И.В., Хабаров М.Д., Харазов В.Г. Компьютерные тренажеры для тестирования и разработки систем усовершенствованного управления технологическим процессом // *Изв. СПбГТИ(ТУ)*. 2017. № 40. С. 111–114.
7. Богомолов Б.Б., Быков Е.Д., Меньшиков В.В., Зубарев А.М. Организационно-технологическое моделирование химико-технологических систем // *Теоретические основы химической технологии*. 2017. Т. 51. № 2. С. 221–229. doi: 10.1134/S0040579517010043.

8. Han J., Li L., Wang J., Chen Sh., Liu Ch., Li C. Simulation and evaluation of float glass furnace with different electrode positions. *J. of the American Ceramic Society*, 2022, vol. 105, no. 2, pp. 7097–7110. doi: 10.1111/jace.18700.
9. Петров Д.Ю. Архитектура информационной системы управления жизненным циклом цифрового двойника для непрерывного производства // Изв. СПбГИ(ТУ). 2021. № 57. С. 98–104.
10. Жуков И.В., Харазов В.Г. Результаты поэтапной модернизации и эксплуатации усовершенствованной системы управления (APC-системы) // Изв. СПбГИ(ТУ). 2017. № 41. С. 105–112.
11. Hashim A.H.H., Hussein B.N. Role of operator training simulator (OTS) in capability building towards the fourth Industrial Revolution (IR 4.0). *IOP Conf. Ser.: Mater. Sci. Eng.*, 2020, vol. 778, art. 012044. doi: 10.1088/1757-899X/778/1/012044.

Software &amp; Systems

doi: 10.15827/0236-235X.142.213-220

2024, 37(2), pp. 213–220

### Digital simulator for training sheet glass forming operators

Valery P. Meshalkin <sup>1,2</sup>, Tamara B. Chistyakova <sup>2</sup>, Dmitry Yu. Petrov <sup>3,4</sup>✉<sup>1</sup> Mendeleev University of Chemical Technology of Russian Federation, Moscow, 125047, Russian Federation<sup>2</sup> St. Petersburg State Technological Institute (Technical University), St. Petersburg, 190013, Moscow, 125047, Russian Federation<sup>3</sup> Institute for Precision Mechanics and Control Problems of the Russian Academy of Sciences, Saratov, 410028, Russian Federation<sup>4</sup> Saratov State University, Saratov, 410012, Russian Federation

#### For citation

Meshalkin, V.P., Chistyakova, T.B., Petrov, D.Yu. (2024) 'Digital simulator for training sheet glass forming operators', *Software & Systems*, 37(2), pp. 213–220 (in Russ.). doi: 10.15827/0236-235X.142.213-220

#### Article info

Received: 31.03.2023

After revision: 27.01.2024

Accepted: 26.02.2024

**Abstract.** The paper analyzes flat glass forming by a float method and determines the types of abnormal technological process situations. It shows a cluster analysis of many emergency situations; each situation is described as a set-theoretic model. To simplify a digital simulator interface, there are developed business process models for operator training and certification. The paper shows logical and information models of promising business processes Test development and Employee certification developed in BPMN 2.0 notation based on the analysis of existing business processes of an enterprise. When developing the architecture and modes of operation and software and information support of the digital simulator, the authors used a model-based system engineering methodology implemented in the IBM Rhapsody visual modeling environment based on UML and SysML notations. The authors developed use case diagrams and class diagrams of the digital simulator using the UML language, a software structure, as well as a training system database structure that includes criteria for grouping situations, situation characteristics, functions of system users. The main implemented functions are the following: training and testing of operators; maintaining a database of digital simulator personnel; compiling reports on test results; editing an emergency database based on a preliminary analysis of technologists' and operators' experience. The digital simulator is designed in the PascalABC language, the results are recorded in a text database. The user interface of the digital simulator has been also developed.

**Keywords:** digital simulator, training, business process of forming flat glass, operator, use case diagram, class diagram, emergency situation

**Acknowledgements.** The work was performed at the IPTMU RAS when fulfilling the state assignment of the Ministry of Science and Higher Education of the Russian Federation (topic no. FFM-2022-0010); the work was supported by a grant from the Russian Science Foundation according to the project no. 21-79-30029

#### References

1. Marcano, L., Haugen, F.A., Sannerud, R., Komulainen, T. (2019). 'Review of simulator training practices for industrial operators: How can individual simulator training be enabled?', *Safety Sci.*, 115, pp. 414–424. doi: 10.1016/j.ssci.2019.02.019.
2. Meshalkin, V.P., Dovì, V.G., Bobkov, V.I., Belyakov, A.V., Butusov, O.B. et al. (2021) 'State of the art and research development prospects of energy and resource-efficient environmentally safe chemical process systems engineering', *Mendeleev Communications*, 31(5), pp. 593–604. doi: 10.1016/j.mencom.2021.09.003.

3. Furaev, D.N., Chistyakova, T.B. (2021) 'Computer system for teaching resource-saving design of oil recycling processes', *Bull. of SPSIT(TU)*, (59), pp. 101–106 (in Russ.). doi: 10.36807/1998-9849-2021-59-85-101-106.
4. Koteleva, N., Kuznetsov, V., Vasileva, N. (2021) 'A simulator for educating the digital technologies skills in industry. Part one. Dynamic simulation of technological processes', *Appl. Sci.*, 11(22), art. 10885. doi: 10.3390/app112210885.
5. Moreland, J., Toth, K., Fang, Y., Block, M. et al. (2019) 'Interactive simulators for steel industry safety training', *Steel Research Int.*, 90(4), art. 1800513. doi: 10.1002/srin.201800513.
6. Zhukov, I.V., Habarov, M.D., Kharazov, V.G. (2017) 'Computer simulators for testing and optimization of technological solutions in refining', *Bull. of SPSIT(TU)*, (40), pp. 111–114 (in Russ.).
7. Bogomolov, B.B., Bykov, E.D., Men'shikov, V.V., Zubarev, A.M. (2017) 'Organizational and technological modeling of chemical process systems', *Theoretical Foundations of Chemical Eng.*, 51(2), pp. 238–246 (in Russ.). doi: 10.1134/S0040579517010043.
8. Han, J., Li, L., Wang, J., Chen, Sh., Liu, Ch., Li, C. (2022) 'Simulation and evaluation of float glass furnace with different electrode positions', *J. of the American Ceramic Society*, 105(2), pp. 7097–7110. doi: 10.1111/jace.18700.
9. Petrov, D.Yu. (2021) 'Architecture of the information system for life cycle management of digital twin for continuous processes', *Bull. of SPSIT(TU)*, (57), pp. 98–104 (in Russ.).
10. Zhukov, I.V., Kharazov, V.G. (2017) 'The results of the gradual upgrading and operation of the advanced control system (APC system)', *Bull. of SPSIT(TU)*, (41), pp. 105–112 (in Russ.).
11. Hashim, A.H.H., Hussein, B.N. (2020) 'Role of operator training simulator (OTS) in capability building towards the fourth Industrial Revolution (IR 4.0)', *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 778, art. 012044. doi: 10.1088/1757-899X/778/1/012044.

**Авторы****Мешалкин Валерий Павлович**<sup>1, 2</sup>,академик РАН, д.т.н., профессор, директор,  
vpmeshalkin@gmail.com**Чистякова Тамара Балабековна**<sup>2</sup>,

д.т.н., профессор, зав. кафедрой, chistb@mail.ru

**Петров Дмитрий Юрьевич**<sup>3, 4</sup>,к.т.н., доцент, старший научный сотрудник,  
зав. кафедрой, iac\_sstu@mail.ru**Authors****Valery P. Meshalkin**<sup>1, 2</sup>, Academician of RAS,Dr.Sc. (Engineering), Professor, Director,  
vpmeshalkin@gmail.com**Tamara B. Chistyakova**<sup>2</sup>, Dr.Sc. (Engineering),

Professor, Head of Chair, chistb@mail.ru

**Dmitry Yu. Petrov**<sup>3, 4</sup>, Cand. of Sci. (Engineering),Associate Professor, Senior Researcher,  
Head of Chair, iac\_sstu@mail.ru<sup>1</sup> Российский химико-технологический университет имени Д.И. Менделеева,

г. Москва, 125047, Россия

<sup>2</sup> Санкт-Петербургский государственный технологический институт (технический университет), г. Санкт-Петербург, 190013, Россия<sup>3</sup> Институт проблем точной механики

и управления РАН, г. Саратов,

410028, Россия

<sup>4</sup> Саратовский национальный исследовательский

государственный университет имени

Н.Г. Чернышевского, г. Саратов, 410012, Россия

<sup>1</sup> Mendeleev University of Chemical Technology of Russian Federation, Moscow, 125047, Russian Federation<sup>2</sup> St. Petersburg State Technological Institute (Technical University),

St. Petersburg, 190013, Russian Federation

<sup>3</sup> Institute for Precision Mechanics and Control Problems of the Russian Academy of Sciences, Saratov, 410028, Russian Federation<sup>4</sup> Chernyshevsky National Research

State University of Saratov,

Saratov, 410012, Russian Federation

## Обеспечение полноты и актуальности пространственной информации в распределенных вычислительных системах больших организаций

В.Д. Трубин <sup>1</sup>✉, С.В. Павлов <sup>1</sup>, О.И. Христуло <sup>1</sup>

<sup>1</sup> Уфимский университет науки и технологий (УУНиТ), г. Уфа, 450076, Россия

### Ссылка для цитирования

Трубин В.Д., Павлов С.В., Христуло О.И. Обеспечение полноты и актуальности пространственной информации в распределенных вычислительных системах больших организаций // Программные продукты и системы. 2024. Т. 37. № 2. С. 221–229. doi: 10.15827/0236-235X.142.221-229

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 15.06.2023

После доработки: 22.02.2024

Принята к публикации: 06.03.2024

**Аннотация.** Применение информационных технологий для сбора, управления и обработки больших объемов разнородной информации об организациях с целью управления ими давно стало естественной практикой. Поиск эффективных методов хранения и обработки информации о вычислительной инфраструктуре организации остается актуальной задачей и сегодня. В данной работе авторы исследуют использование геоинформационных технологий для обработки пространственной и атрибутивной информации об объектах вычислительной инфраструктуры организации на примере университета как сложной распределенной структуры. Одной из важнейших задач является централизованное хранение полных и актуальных пространственных данных с разграничением доступа пользователей из разных подразделений к различным частям БД для обеспечения безопасности данных. Результатом работы стал механизм формирования распределенной пространственной структуры данных, предназначенной для хранения геоданных с учетом пространственного распределения отдельных подразделений организации. Приведена схема взаимодействия различных подразделений организации с базой геоданных. Результаты исследований могут найти применение при разработке структуры хранения геоданных об объектах компьютерной инфраструктуры, а также в любой другой сфере деятельности организации. Разработанная для геоинформационной системы университета подсистема автоматизированного обеспечения достоверности и актуальности пространственных данных позволяет ускорить обработку пространственной информации в БД, исключить дублирование устаревших данных и тиражирование актуальных, оставляя при этом различные формы представления этой информации для каждой из подсистем в рамках одной геоинформационной системы.

**Ключевые слова:** геоинформационная система, пространственная БД, атрибутивная БД, сложная распределенная система, вычислительная инфраструктура, методы хранения и обработки данных

**Введение.** С применением информационных технологий управление сложными распределенными организациями существенно упростилось. Вся деятельность организации в той или иной степени проходит через информационные системы, предназначенные для автоматизации (бухгалтерский учет, оповещение о приказах руководства, управление деятельностью отдельных подразделений и т.д.). При проектировании подобных систем необходимо решать множество задач, например, в каком виде должна быть информация, пропускаемая через систему, какие методы использовать для управления этой информацией и как ее эффективно обрабатывать.

Задача существенно усложняется, когда система проектируется для сложной распределенной организации, что делает систему тоже сложной и распределенной. Ранее авторами уже рассматривались некоторые подходы к анализу, структурированию и организации разнородных данных для сложных распределенных информационных систем [1]. В частности,

применение геоинформационных технологий для обработки пространственной и атрибутивной информации об объектах организации с целью повышения эффективности ее функционирования.

Эксплуатация ресурсов информационных технологий организации, в свою очередь, также является задачей по управлению сложной распределенной системой, поскольку компьютерная инфраструктура включает в себя все многообразие вычислительных ресурсов (серверы, стационарные компьютеры, мобильные устройства, средства виртуализации), устройств ввода-вывода информации (принтеры, мониторы, сканеры, многофункциональные устройства), объединенных в вычислительную сеть с помощью специальных средств (линии связи, маршрутизаторы, коммутаторы и др.). Огромное количество и разнообразие компьютерной техники повышают сложность задачи сбора, обработки, планирования технического обслуживания и ремонтных работ, а актуальность задачи хранения информации о вычислительной

технике подтверждается существующими исследованиями и разработками в этой области [2–4]. Однако типовая подобная система не использует пространственное расположение техники для поддержки принятия решений. Для обеспечения качественного создания и функционирования этой инфраструктуры существенное значение имеет информация о местоположении и взаимном расположении ее отдельных компонентов, а также о расположении сопутствующей инфраструктуры (планы помещений, сеть электроснабжения и др.) и основных ее пользователей (потребителей), что приводит к необходимости автоматизированной обработки такой информации, называемой пространственной [1], и обуславливает создание пространственной БД.

В качестве объекта исследования был взят Уфимский государственный авиационный технический университет (далее – университет) как сложная распределенная система. Университет включает множество подразделений, каждое из которых оперирует собственными массивами данных. Для хранения и обработки этих разнородных данных в виде пространственных и атрибутивных записей разработана *геоинформационная система* (ГИС) университета. Помимо собственно данных и атрибутивной информации, она включает в себя решаемые задачи и пользователей системы (как людей, так и алгоритмы управления), в совокупности образующие сложную структуру современного управления университетом. Исследования в этой области [5] идентифицируют подобную систему как информационно-управляющую (ИУС), однако также не учитывают пространственное расположение ее объектов как дополнительный инструмент управления. Представленная на рисунке 1 структура хранения и использования пространственных данных отражает некоторые типовые пользовательские роли в подобной распределенной системе, обеспечивающей автоматизацию определенных видов вузовской деятельности. Рассмотрим более детально компьютерную инфраструктуру университета как основной инструмент обработки информации.

#### **Обеспечение полноты и актуальности компьютерной инфраструктуры**

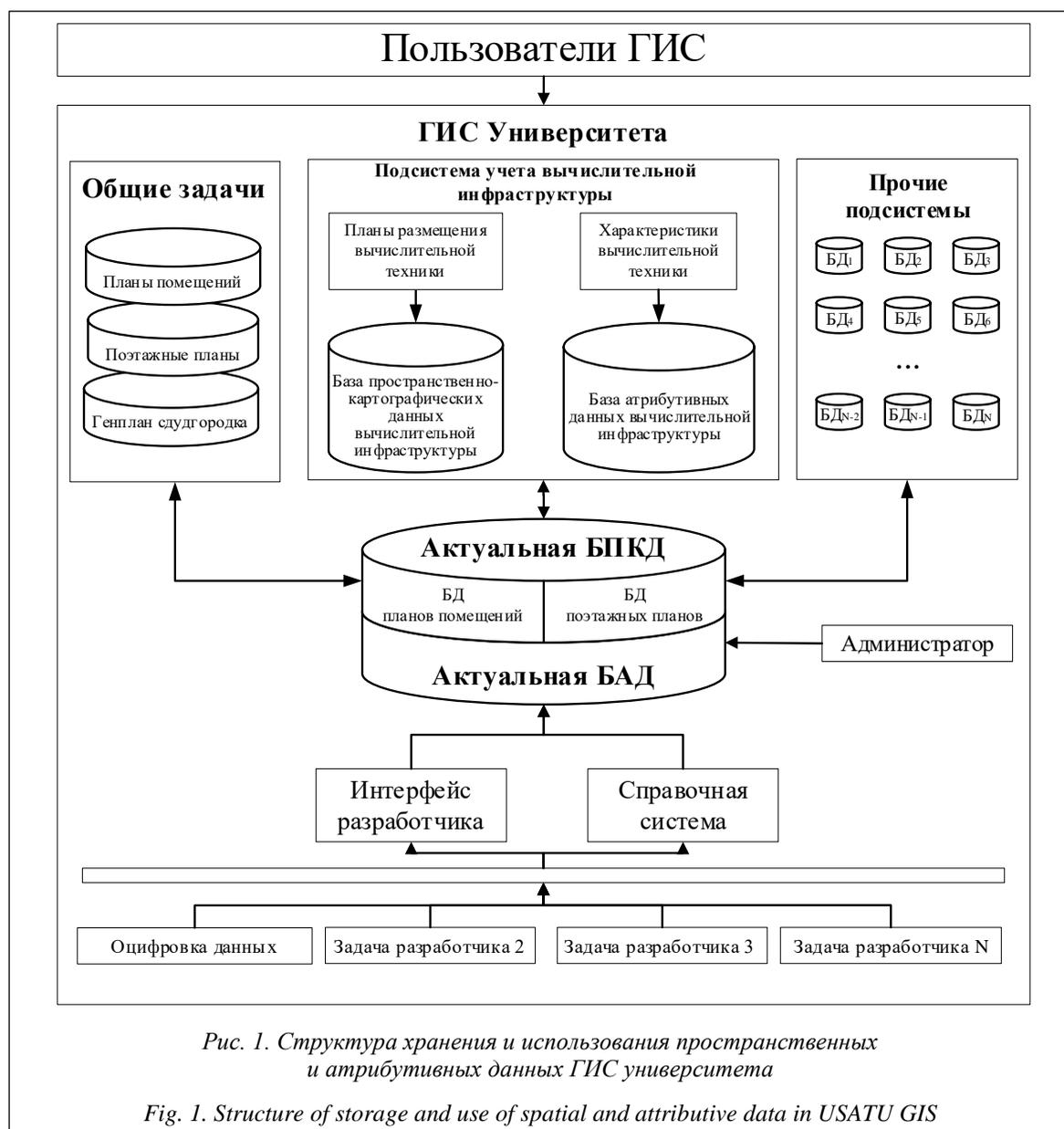
Разрабатываемая ГИС должна содержать в себе пространственные данные, удовлетворяющие запросам и требованиям различных организационных структур, покрывающих зна-

чительную часть деятельности университета. В качестве надежных источников этих данных выступают официальные чертежи, планы корпусов и схемы коммуникации, однако даже в таком случае не исключена вероятность обнаружения неточностей при интеграции данных из двух разных источников.

Это связано с отсутствием мероприятий по своевременной актуализации чертежей, упростить организацию которых может помочь наличие единого источника информации в виде ГИС. Подобная проблема характерна и для атрибутивной информации, полученной на основе анализа бухгалтерского учета технического оснащения помещений университета [6]. Поэтому, помимо официальных источников данных, в организации баз пространственных и атрибутивных данных используются материалы, полученные на основе полевых исследований, а также чертежи и планы, не имеющие юридической силы, но отражающие более актуальную информацию об окружающей действительности. В частности, именно эти документы представляют собой основной источник данных для наполнения БД помещений, необходимых для функционирования большинства подсистем в составе ГИС университета (<http://www.swsys.ru/uploaded/image/2024-2/13.jpg>).

Полнота данных как основной критерий разнообразия функциональных возможностей информационной системы [5] является показателем ее пригодности к последующему применению, то есть к анализу инфраструктуры университета. Каждой из существующих подсистем ГИС необходим собственный набор данных, иногда пересекающийся с другими подсистемами в виде полей в атрибутивных таблицах или в одинаковых слоях пространственных объектов.

Актуальность данных обеспечивается за счет своевременного обновления БД системы. В ГИС университета актуальность данных поддерживается пользователями системы на уровне своих подсистем. Сотрудник подразделения, данные которого устарели, обновляет информацию об объектах своего интереса, используя инструментарий ГИС. Для внесения этих изменений можно использовать условные обозначения на специальных слоях, но в проекте предполагается разработка специального инструментария. Эти изменения отправляются на сервер кафедры ГИС, курирующей проект, где администратор системы, в свою очередь, подтверждает внесение предложенных изменений в централизованную БД. В дальнейшем со-



трудники других подразделений, которые используют ту же информацию, получат обновленные данные с сервера. С точки зрения организации такой распределенной БД имеет смысл рассматривать реляционную СУБД как основной механизм управления этими данными. Необходимо сохранять целостность и непротиворечивость данных в системе. Помимо этого, подразумевается многопользовательская работа с обозначенными данными, а обеспечить их корректность при такой работе позволяет механизм транзакций реляционных СУБД. Все изменения и новая информация вносятся в БД человеком вручную, и в быстрой обработке поступающих данных нет необходимости [7, 8].

Под защищенностью понимается централизованное хранение полных и актуальных данных с разграничением доступа пользователей разных подразделений к разным участкам БД [9]. Предоставление доступа ко всей базе пространственных данных сотрудникам подразделений не только излишне, но и небезопасно с точки зрения сохранности данных в актуальном виде [10].

Одной из важнейших задач, решаемых ГИС университета, является учет имущества, связанного с вычислительными системами учреждения. Информация об объектах вычислительной инфраструктуры является важным аспектом, определяющим политику развития университета в целом. Для эффективного

управления информационной системой университета недостаточно физических характеристик существующих рабочих станций, серверного оборудования и сетевых каналов. Различным структурным подразделениям необходима специфическая информация об инфраструктуре вычислительных коммуникаций, состоянии вычислительных машин, установленном ПО и т.п. В дальнейшем эти сведения могут быть использованы подразделением для последующей обработки в целях администрирования сетей, безопасности, анализа аппаратного обеспечения с последующей закупкой нового оборудования и пр.

Существуют программные продукты, реализующие централизованный опрос и учет всей вычислительной техники [11] в рамках отдельно взятой локальной сети, однако ни один из них не использует пространственные данные в качестве дополнительного атрибута для последующего анализа. Помимо трудностей при хранении пространственной информации совместно с большим объемом атрибутивных данных, возникает вопрос обеспечения актуальности непосредственно пространственных данных, поскольку в отличие от разного рода географических объектов и недвижимости пространственное расположение точечных объектов в выбранной исследовательской области может изменяться достаточно часто.

Потребителями этой информации являются все кафедры университета. Каждая кафедра несет ответственность за свои вычислительные мощности.

На текущий момент для предоставления требуемого набора данных сущности, которая его запрашивает, необходимо собрать всю возможную информацию в единую БД, после чего разработчики ГИС университета приводят ее в понятный для конечного пользователя вид.

Одним из наиболее востребованных современных подходов к разработке подобных распределенных ГИС с возможностью многопользовательского редактирования данных в едином хранилище является организация геопортала [12]. Однако такой подход подразумевает использование внутреннего инструментария геопортала для редактирования данных, что не всегда удобно для пользователя без определенных знаний о ГИС. Поэтому приведение информации и методов ее использования в удобный для конечного пользователя вид также является основополагающим критерием при разработке ГИС распределенной организации для специалистов разных профилей.

Описывается как атрибутивная, так и пространственная информация. Основная проблема, с которой предстоит столкнуться разработчикам ГИС при анализе и работе с этими данными, – противоречивость и разное качество данных из различных источников. Для ее решения необходимо разработать дополнительную подсистему проверки полноты и актуальности вводимых данных.

Сопоставление пространственной информации в такой подсистеме должно проводиться вручную, поскольку разные источники информации могут по-разному выделить один и тот же объект в пределах помещений. В этом случае критерием отбора является источник, чья информация наиболее полная и актуальная. Такими источниками могут выступать максимально приближенные к дате сбора пространственных данных официальная документация архитектурных планов этажей помещений, инвентарная опись имущества отдельных подразделений, акты закупок и иные виды документов, подтверждающие наличие и утверждающие характеристики рабочего оборудования. Информация из этих источников должна быть синхронизирована с другими подсистемами, работающими с этой же пространственной информацией, таким образом, чтобы при удалении или добавлении новых объектов инфраструктуры изменения отражались в каждой из подсистем. Таким образом, первичное наполнение актуальными данными будет проводиться на основе документации и ее ручной проверки на соответствие реальному положению вещей на момент сбора, а их дальнейшая актуализация по мере необходимости будет обеспечена в рамках работы ГИС сотрудниками подразделений, ответственными за конкретную информацию.

Синхронизация атрибутивных данных в настоящее время не производится, поскольку каждому из подразделений необходим свой набор данных. Одна и та же информация представляется конечному пользователю в разном виде, что приводит к ее дублированию в БД. В качестве решения этой проблемы с целью оптимизации БД возможно программно реализовать слияние информации из разных источников в единую атрибутивную таблицу, после чего размножить ее на каждый пространственный объект. Однако в таком случае информация в БД будет избыточной и должна дополнительно обрабатываться методами нормализации [13]. Общая сводная атрибутивная таблица должна генерироваться на основе данных

из разных подсистем, чтобы предоставлять эту информацию по требованию пользователей. Подразделения же должны работать со своими данными и иметь возможность вносить в них изменения только в своих подсистемах, не влияя на полную информацию об объектах. Таким образом будет обеспечиваться актуальность данных в системе. Цель реализации описанного подхода – разработка механизма формирования структуры распределенных пространственных данных и подсистемы автоматизированного обеспечения достоверности и актуальности.

### Механизм формирования структуры распределенных пространственных данных

Для эффективной организации пространственных данных нескольких подразделений и управления ими предлагается сформировать базу геоданных, учитывающую территориальную распределенность отдельных подразделений организации (рис. 2).

В качестве потока данных подсистемы подразделений организации из БД принимают витрину данных как специализированное хранилище, обслуживающее единственное направление деятельности организации или подразделения. Достоинством применения витрины является относительная простота семантики данных в пределах одной витрины [14], что упростит работу с подсистемами сотрудникам разной квалификации.

Используя математический аппарат, можно представить БД ГИС университета в виде сово-

купности множеств пространственных и атрибутивных данных разных подразделений:

$$BD_0 = \{BPKD, BAD\}, \tag{1}$$

$$BPKD = \{K^i\}, i = \overline{1, n},$$

где BPKD – база пространственно-картографических данных; BAD – база атрибутивных данных;  $K^i$  – совокупность карт  $i$ -го подразделения;  $n$  – количество подразделений, использующих ГИС университета в качестве источника пространственных данных.

Объединим две записи в одну:

$$BD_0 = \sum_{i=1}^n BD_i, \tag{2}$$

где  $BD_0$  – центральная база геоданных информационной системы;  $BD_i$  – база геоданных  $i$ -го подразделения организации.

Однозначно можно выделить из БД отдельных подразделений базу геоданных, содержащую поэтажные планы организации ( $BD_3$ ). Из этих данных состоит база геоданных любого подразделения, и в репликации этих данных нет необходимости. В итоге база геоданных каждого подразделения будет состоять из БД помещений ( $BD_{pi}$ ) и БД специализированной информации ( $BD_i$ ). Таким образом, БД  $i$ -го подразделения можно определить как

$$BD_i = BD_3 + BD_{pi} + BD_i. \tag{3}$$

В таком случае выражение (1) будет иметь вид

$$BD_0 = BD_3 + \sum_{i=1}^n (BD_i + BD_{pi}). \tag{4}$$

Следует учитывать, что существует область данных, необходимая нескольким подразделениям одновременно, в частности, именно БД

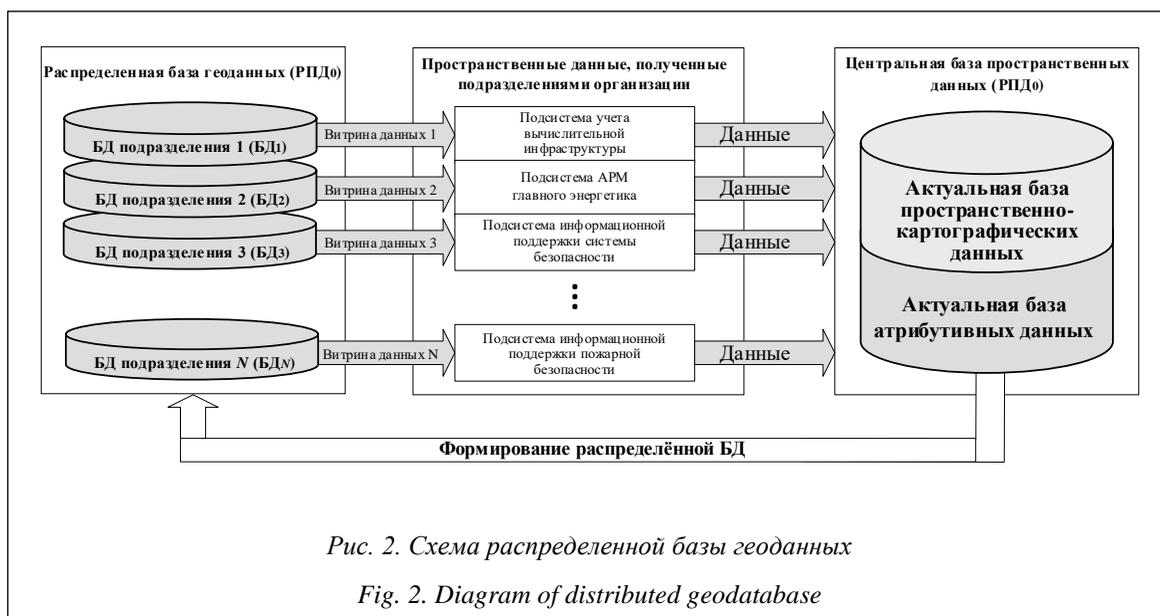


Рис. 2. Схема распределенной базы геоданных

Fig. 2. Diagram of distributed geodatabase

помещений. В качестве примера возьмем помещения кафедры ГИС, БД которой ( $БД_{п.каф.ГИС}$ ) будет состоять из границ помещений, закрепленных за кафедрой, инвентаря этих помещений и линий коммуникаций, включающих в себя вычислительные и электросети. Помимо кафедры ГИС, доступ к данным об этих объектах необходим бухгалтерии, инженерам электрических коммуникаций и управлению информационных технологий. Учитывая это, необходимы выражения, описывающие такие общие базы геоданных для двух подразделений:

$$БД_{nik}^{\Delta} = БД_{ni} \cap БД_{nk}, \quad i = \overline{1, n}; \quad k = \overline{1, n}; \quad i \neq k, \quad (5)$$

для трех подразделений:

$$\begin{aligned} \overline{БД_{nijk}} &= БД_{ni} \cap БД_{nk} \cap БД_{nj} = \\ &= БД_{nik}^{\Delta} \cap БД_{njk}^{\Delta} \cap БД_{nij}^{\Delta}, \end{aligned} \quad (6)$$

где  $i, j, k = \overline{1, n}; i \neq j \neq k$ .

Структура распределенной базы геоданных представлена на рисунке 3.

Выражение (6) справедливо для случая, когда данные одних и тех же помещений доступны трем подразделениям, однако в системе возможна ситуация, когда  $n = i$ , то есть количество подразделений, которым необходимы данные по помещению, равно максимально возможному количеству подразделений. Для решения именно этой задачи предусмотрен механизм обеспечения достоверности и актуальности данных.

Его основными функциями являются отправка единого информационного потока с данными о помещениях во все БД подразделений, которым это необходимо, и принятие произведенных в одной из распределенных БД изменений, отправка их в централизованную БД и дальнейшее распространение уже отредактированной информации.

Для обеспечения контроля над безопасностью при эксплуатации системы предусмотрены механизмы репликации журналов тран-

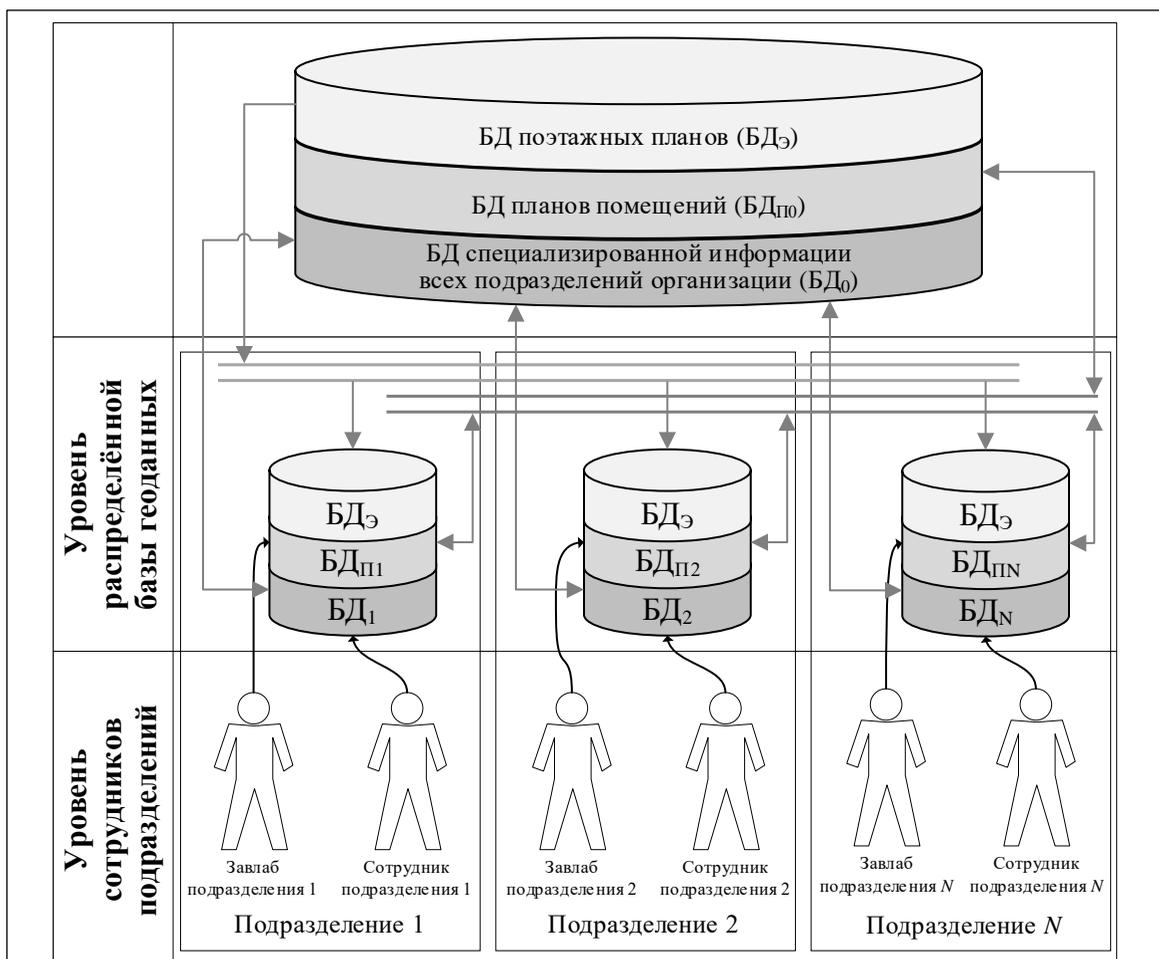


Рис. 3. Структура взаимодействия подразделений с распределенной базой геоданных

Fig. 3. Interaction of departments with a distributed geodatabase

закций и таблиц пользователей на уровнях центральной и распределенной баз геоданных. На уровне центральной БД в реляционной СУБД хранится информация о пользователях, их ролях и привилегиях в отношении чтения и изменения пространственных и атрибутивных данных. Эта информация передается в БД подразделений для авторизации и аутентификации в системе встроенными средствами СУБД и ГИС. Добавление новых учетных записей выполняется администратором системы в центральной базе. Журналы транзакций, выполняющие функцию аудита системы, напротив, реплицируются из БД подразделений в центральную. Поскольку перед применением внесенных изменений новые данные отправляются на проверку администратору, репликация записей журнала транзакций происходит только после одобрения со стороны центральной БД. Если же изменения отклоняются, запись о них остается в журнале БД подразделения до тех пор, пока замечания администратора не будут устранены и транзакция новых данных в центральную базу не будет успешно завершена. Для каждой БД подразделения в центральной базе хранятся соответствующая полная копия журнала транзакций, а также полный журнал центральной БД, что создает избыточность, однако необходимо для внедрения дополнительного функционала работы с системой и для ее дальнейшего развития.

Резервное копирование выполняется для центральной БД посредством создания инкрементальных копий на сервере кафедры ГИС с определенной периодичностью методами обслуживающей СУБД. На момент создания резервной копии все действия с данными должны быть завершены, чтобы исключить возможность возникновения ошибок. Резервное копирование распределенных БД выполнять нецелесообразно ввиду наличия общей копии, хранящейся в центральной БД. Возможен вариант с созданием локальных копий БД каждого отдельно взятого подразделения, что упростит процедуру восстановления данных в случае их разрушения, однако потребует гораздо боль-

шего объема хранилища, а это может негативно сказаться на дальнейшем развитии проекта вследствие его масштабирования по ходу разработки новых подсистем.

### Заключение

Полнота данных обеспечивается за счет одновременной работы с одними и теми же данными разными подразделениями сложной распределенной организации, где каждое отвечает за полноту собственных данных. Данные из разных подразделений хранятся централизованно, что при полном доступе к центральной БД позволяет получить большой объем данных, описывающих организацию максимально полно.

Актуальность данных обеспечивается за счет того же разграниченного доступа. Каждое подразделение актуализирует собственный набор данных в системе по мере ее обновления в организации, а администратор системы подтверждает принимаемые в центральной БД изменения, что гарантирует достоверность внесенных изменений и актуальность полученных в системе данных.

Защищенность с точки зрения сохранности данных от изменений со стороны обеспечивается за счет разграниченного доступа к нескольким БД, формируемых системой. Доступ к информации одной предметной области предоставляется лишь тем подразделениям, работа которых связана с этой предметной областью, что гарантирует отсутствие нежелательных изменений атрибутивных или пространственных данных работниками других подразделений.

Упомянутая подсистема автоматизированного обеспечения достоверности и актуальности пространственных данных позволяет ускорить процесс обработки пространственной информации в ГИС УГАТУ на предмет исключения дублирования устаревших данных и тиражирования актуальных данных, при этом оставляя разные формы представления информации для каждой из подсистем, функционирующих в рамках одной ГИС.

### Список литературы

1. Брекоткина Е.С., Павлов А.С., Павлов С.В., Трубин В.Д., Христодуло О.И. Информационная поддержка управления компьютерной инфраструктурой сложных распределенных систем // Вестн. компьютерных и информационных технологий. 2022. Т. 19. № 3. С. 23–32. doi: 10.14489/vkit.2022.03.pp.023-032.
2. Горбунов И.Д., Антаев М.П., Гаптуллазянова Г.И. Информационная система учета компьютерной техники // Междунар. форум Kazan Digital Week-2022: сб. матер. 2022. С. 172–180.
3. Ахунова И.Г., Козлов А.Н. Автоматизация учета компьютерной и копировально-множительной техники // Инженерное обеспечение в реализации социально-экономических и экологических программ АПК: матер. науч.-практич. конф. 2020. С. 345–349.

4. Миних Е.А., Шамбер О.Ю. Программная реализация базы данных приложения для ведения учета заявок обслуживания компьютерной техники в центре информационных технологий вуза // Форум молодых ученых. 2019. № 1-2. С. 853–856.
5. Попов Ф.А., Ануфриева Н.Ю. Особенности и проблемы построения сложных интегрированных ИУС для целей комплексного управления деятельностью университета // Информация и образование: границы коммуникаций INFO: матер. Междунар. науч.-практич. конф. 2020. № 12. С. 31–33.
6. Брекоткина Е.С., Павлов А.С., Павлов С.В., Христовуло О.И., Трубин В.Д. Анализ и синтез структуры пространственных данных для управления сложными распределенными системами // ИТиС: тр. науч. конф. 2020. С. 120–124.
7. Никишин Д.А. Сопоставление особенностей представления геоданных в картографии и в геоинформатике // Системы и средства информатики. 2018. Т. 28. № 2. С. 60–74. doi: 10.14357/08696527180205.
8. Мацук М.А., Кяхяри И.И., Алтанец Я.В. Реляционные и нереляционные системы БД // Modern Science. 2021. № 3-2. С. 519–524.
9. Брекоткина Е.С., Павлов С.В., Трубин В.Д., Христовуло О.И. Управление проектом создания сложной геоинформационной системы на основе специализированного справочника // Системная инженерия и информационные технологии. 2021. Т. 3. № 2. С. 66–73. doi: 10.54708/26585014\_2021\_32666.
10. Яготинцева Н.В., Татарникова Т.М. Защита данных геоинформационных систем // Информационные технологии и системы: управление, экономика, транспорт, право. 2019. № 1. С. 131–133.
11. Наймушина О.Э. Организация учета компьютерной техники и программного обеспечения // МНИЖ. 2018. № 4. С. 120–122.
12. Ямашкин А.А., Ямашкин С.А. Геопорталы как инструмент доступа к облачным хранилищам данных о метеосистемах // Изв. Саратов. ун-та. Нов. сер. Сер. Науки о Земле. 2022. Т. 22. № 4. С. 251–261. doi: 10.18500/1819-7663-2022-22-4-251-261.
13. Заимов А.И., Ронжин А.Е. Современные методы проектирования баз данных // АВСЭ. 2020. № 11. С. 510–514.
14. Некрасов А.А., Гаврилов С.О., Беленькая М.Н. Средства создания хранилищ данных // ТИТ. 2021. Т. 8. № 1. С. 75–80.

### Ensuring spatial information completeness, relevance and security in distributed computing systems of large organizations

Vladislav D. Trubin <sup>1</sup>✉, Sergey V. Pavlov <sup>1</sup>, Olga I. Khristodulo <sup>1</sup>

<sup>1</sup> Ufa University of Science and Technology, Ufa, 450076, Russian Federation

#### For citation

Trubin, V.D., Pavlov, S.V., Khristodulo, O.I. (2024) 'Ensuring spatial information completeness, relevance and security in distributed computing systems of large organizations', *Software & Systems*, 37(2), pp. 221–229 (in Russ.). doi: 10.15827/0236-235X.142.221-229

#### Article info

Received: 15.06.2023

After revision: 22.02.2024

Accepted: 06.03.2024

**Abstract.** Nowadays, it is a matter of course to use information technologies for collecting, managing and processing large volumes of heterogeneous information about organizations to manage them. The search for effective methods of storing and processing information on an organization computing infrastructure in order to manage that infrastructure more effectively remain an urgent task, as system administrators take an individual approach to each system. This paper investigates the use of GIS technologies to process spatial and attributive information about organization computing infrastructure objects on the example of a university as a complex distributed structure. One of the most important tasks is centralized storage of full and actual spatial data with access rights differentiation for users from different departments to different parts of a database to ensure data security. The result is a mechanism for forming a distributed spatial data structure designed to store geodata, taking into account the spatial distribution of organization individual subdivisions. The paper presents a diagram of interaction between different divisions of an organization with a geodatabase. The resulted research can be used for developing a storage structure of geodata about computer infrastructure objects, and any other area of organization activity. The subsystem of automated ensuring spatial data reliability and relevance that was developed for university GIS allows accelerating the processing of spatial information in a database eliminating outdated data duplication and current data replication, while leaving different forms of presenting this information for each GIS subsystem.

**Keywords:** GIS, spatial database, attributive database, complex distributed system, computing infrastructure, data storage and processing methods

## References

1. Brekotkina, E.S., Pavlov, A.S., Pavlov, S.V., Trubin, V.D., Khristodulo, O.I. (2022) 'Information support for managing the computer infrastructure of complex distributed systems', *Herald of Computer and Information Technologies*, 19(3), pp. 23–32 (in Russ.). doi: 10.14489/vkit.2022.03.pp.023-032.
2. Gorbunov, I.D., Antaev, M.P., Gaptullazyanova, G.I. (2022) 'Computer equipment accounting information system', *Proc. Int. Forum Kazan Digital Week-2022*, pp. 172–180 (in Russ.).
3. Akhunova, I.G., Kozlov, A.N. (2020) 'Automation accounting of computer and copying and multiplying equipment', *Proc. Sci-Pract. Conf. Engineering Support in Implementing Socio-Economic and Environmental Programs of the Agro-Industrial Complex*, pp. 345–349 (in Russ.).
4. Minikh, E.A., Shamber, O.Yu. (2019) 'Software implementation of the application database for keeping records of computer maintenance applications at the university information technology center', *Forum of Young Scientists*, (1-2), pp. 853–856 (in Russ.).
5. Popov, F.A., Anufrieva, N.Yu. (2020) 'Features and problems of construction of complex integrated information and control systems (ICS) for the purposes of integrated management of the university', *Proc. Int. Sci.-Pract. Conf. Information and Education: Borders of Communication*, (12), pp. 31–33 (in Russ.).
6. Brekotkina, E.S., Pavlov, A.S., Pavlov, S.V., Khristodulo, O.I., Trubin, V.D. (2020) 'Analysis and synthesis of spatial data structure for managing complex distributed systems', *Proc. ITaS*, pp. 120–124 (in Russ.).
7. Nikishin, D.A. (2018) 'Comparison of the features of the representation of geodata in cartography and geoinformatics', *Systems and Means of Informatics*, 28(2), pp. 60–74 (in Russ.). doi: 10.14357/08696527180205.
8. Matsuk, M.A., Kyakhyari, I.I., Altanets, Ya.V. (2021) 'Relational and Non-Relational Database Systems', *Modern Science*, (3-2), pp. 519–524 (in Russ.).
9. Brekotkina, E.S., Pavlov, S.V., Trubin, V.D., Khristodulo, O.I. (2021) 'Project management for creating a complex geographic information system based on a specialized reference book', *Systems Eng. and Inform. Tech.*, 3(2), pp. 66–73 (in Russ.). doi: 10.54708/26585014\_2021\_32666.
10. Yagotinceva, N.V., Tatarnikova, T.M. (2019) 'Data protection geographic information systems', *Inform. Tech. and Systems: Management, Economics, Transport, Law*, (1), pp. 131–133 (in Russ.).
11. Naymushina, O.E. (2018) 'Organization of computer engineering and software accounting', *Int. Research J.*, (4), pp. 120–122 (in Russ.).
12. Yamashkin, A.A., Yamashkin, S.A. (2022) 'Geoportals as a tool for access to cloud storage of metageosystems data', *Izv. of Saratov University. Earth Sci.*, 22(4), pp. 251–261 (in Russ.). doi: 10.18500/1819-7663-2022-22-4-251-261.
13. Zaimov, A.I., Ronzhin, A.E. (2020) 'Modern Methods for Database Design', *Actual Issues of the Modern Economics*, (11), pp. 510–514 (in Russ.).
14. Nekrasov, A.A., Gavrilov, S.O., Belenkaya, M.N. (2021) 'Overview of data warehouse development tools', *TIT*, 8(1), pp. 75–80 (in Russ.).

## Авторы

**Трубин Владислав Дмитриевич**<sup>1</sup>, аспирант,  
ассистент кафедры, trubin.vd@ugatu.su  
**Павлов Сергей Владимирович**<sup>1</sup>,  
д.т.н., профессор, psvgis@mail.ru  
**Христодуло Ольга Игоревна**<sup>1</sup>, д.т.н.,  
профессор, o-hristodulo@mail.ru

<sup>1</sup> Уфимский университет науки и технологий  
(УУНиТ), г. Уфа, 450076, Россия

## Authors

**Vladislav D. Trubin**<sup>1</sup>, Postgraduate Student,  
Teaching Assistant, trubin.vd@ugatu.su  
**Sergey V. Pavlov**<sup>1</sup>, Dr.Sc. (Engineering),  
Professor, psvgis@mail.ru  
**Olga I. Khristodulo**<sup>1</sup>, Dr.Sc. (Engineering),  
Professor, o-hristodulo@mail.ru

<sup>1</sup> Ufa University of Science and Technology,  
Ufa, 450076, Russian Federation

## Безопасный обмен файлами на основе сетей доверия и сертификатов открытых ключей с помощью разработанного приложения

П.Б. Хорев <sup>1</sup>✉, Д.А. Лосев <sup>1</sup>

<sup>1</sup> Национальный исследовательский университет «Московский энергетический институт», г. Москва, 111250, Россия

### Ссылка для цитирования

Хорев П.Б., Лосев Д.А. Безопасный обмен файлами на основе сетей доверия и сертификатов открытых ключей с помощью разработанного приложения // Программные продукты и системы. 2024. Т. 37. № 2. С. 230–237. doi: 10.15827/0236-235X.142.230-237

### Информация о статье

Группа специальностей ВАК: 2.3.5

Поступила в редакцию: 29.09.2023

После доработки: 18.12.2023

Принята к публикации: 25.12.2023

**Аннотация.** В статье анализируются существующие программные средства обмена файлами (в том числе облачные хранилища, программы-мессенджеры, электронная почта), выделены их недостатки. Для устранения этих недостатков авторами разработан проект клиент-серверного приложения, основанного на построении сети доверия с помощью сертификатов открытых ключей ее участников. В проекте использованы алгоритмы симметричной и асимметричной криптографии, методы построения сетей доверия и применения сертификатов открытых ключей для обеспечения подлинности, конфиденциальности и целостности данных. Реализованы серверная часть и клиентское мобильное приложение, позволяющие пользователям обмениваться файлами внутри одного конкретного файлового хранилища. Загружаемые файлы имеют электронные подписи их создателей и хранятся на сервере в зашифрованном виде. При разработке приложения использовалась технология, облегчающая его перенос на другие операционные платформы. Клиентское приложение создано и протестировано на устройствах под управлением операционной системы Android. Оно обладает необходимой для файлообменника функциональностью и обеспечивает подлинность, целостность и конфиденциальность передаваемых файлов. Это позволяет использовать разработанное приложение как группами частных пользователей, так и в корпоративных информационных системах.

**Ключевые слова:** клиент-серверное приложение, сертификаты открытых ключей, сеть доверия, безопасный обмен файлами

**Введение.** Задача обеспечения безопасного хранения и обмена файлами встала практически сразу после появления Интернета. Основное внимание при этом уделялось и уделяется до сих пор защите данных при их хранении на сервере с помощью шифрования, разграничения доступа к файлам на уровне пользователей и физической защиты сервера. Безопасность загрузки и выгрузки файлов с сервера и на сервер обычно обеспечивается использованием протокола HTTPS, являющегося объединением протокола передачи гипертекста HTTP и протоколов безопасной передачи данных SSL/TLS. Протокол HTTPS предполагает обязательное подтверждение подлинности (аутентификацию) сервера, а аутентификация клиента может быть дополнительной опцией, которая требует наличия у клиента так называемого сертификата открытого ключа, выданного удостоверяющим центром, которому доверяет сервер. Аутентификация клиента может быть осуществлена сервером после установки защищенного соединения с помощью одного из традиционных методов (например, по имени и паролю пользователя). Надежность такой аутентификации пол-

ностью определяется длиной, сложностью, регулярной сменой и безопасным использованием пароля пользователя.

В качестве примеров решений для обмена данными между пользователями Интернета можно рассмотреть облачные хранилища, программы-мессенджеры и электронную почту. Однако эти средства не всегда удобны и безопасны. Например, в социальной сети «ВКонтакте» есть ограничения на расширение файлов и на их размер, облачные хранилища имеют ограниченное дисковое пространство, а приложения электронной почты могут блокировать передачу исполнимых файлов (в том числе внутри архивов). Кроме того, использование ссылок на файлы в облачных хранилищах может привести к утечке информации, если эти ссылки станут известны посторонним лицам. Доступ к аккаунтам пользователей мессенджеров могут получить другие лица в результате человеческой ошибки владельцев аккаунтов или использования ими недостаточно стойкого к подбору пароля.

В ряде работ (например, [1, 2]) описываются угрозы безопасности данных при их обмене

между пользователями глобальной сети, анализируются достоинства и недостатки применяемых программных решений. В них не предлагаются методы устранения отмеченных недостатков, а для обеспечения безопасности данных рекомендуется использовать защищенные протоколы передачи файлов (HTTPS, SFTP и др.), наложение ограничений на минимальную длину и сложность паролей пользователей, двухфакторную аутентификацию пользователей (например, с помощью пароля и кода из SMS-сообщения на номер телефона, связанного с аккаунтом пользователя, как в сервисе «Яндекс. Почта»).

Однако эти решения имеют ряд недостатков:

- шифрование и контроль целостности файлов обеспечиваются только при их передаче, но не при хранении на стороне сервера;
- не гарантируется реальная регулярная смена пароля пользователя;
- двухфакторная аутентификация может быть неудобной для пользователя (например, из-за увеличения времени на обмен файлами);
- существует риск распространения недостоверной информации посторонними лицами, подобравшими или угадавшими пароль, получившими доступ к мобильному устройству пользователя и (или) применяемому им клиентскому приложению.

В связи с этим возникает необходимость в создании более защищенных файлообменников, которые позволят пользователям безопасно обмениваться файлами с сохранением удобств, предоставляемых другими подобными средствами. Предлагаемое решение основано на построении сети доверия пользователей файлообменника с применением возможностей, предоставляемых современной криптографией. Разработанное приложение имеет серверную и клиентскую части и допускает простой перенос на другие операционные платформы. Возможными применениями разработанного приложения могут быть организация и проведение учебного процесса в университетах, а также обмен конфиденциальными файлами между частными лицами и в информационных системах малого и среднего бизнеса.

#### **Методы создания сетей доверия и использования сертификатов открытых ключей**

В основе обеспечения безопасного хранения и обмена файлами в сети Интернет лежат методы современной криптографии. Традици-

онная симметричная криптография используется для обеспечения конфиденциальности данных путем их шифрования с помощью одного, известного отправителю и получателю данных секретного ключа (его безопасная передача от отправителя к получателю является основной проблемой симметричной криптографии). Симметричные криптосистемы используются для шифрования данных любого объема при их хранении и передаче.

Асимметричная криптография использует два ключа – открытый (публичный) и закрытый, который хранится у владельца и никому не должен передаваться. Открытый ключ используется для шифрования секретного ключа симметричного шифрования, которым защищены передаваемые данные, и для проверки электронной подписи под ними. Закрытый ключ применяется для расшифрования секретного ключа симметричного шифрования и вычисления электронной подписи.

Подлинность открытого ключа (его принадлежность конкретному владельцу двух ключей асимметричного шифрования) подтверждается сертификатом открытого ключа, выданным удостоверяющим центром (центром сертификации). Список доверенных удостоверяющих центров должен вестись у каждого пользователя сети Интернет на каждом его устройстве.

Для создания сети доверия пользователей файлообменника могут применяться методы и средства инфраструктуры открытых ключей (Public Key Infrastructure, PKI) [3], открытого стандарта OpenPGP (Open Pretty Good Privacy) [4] и его свободно распространяемой реализации GnuPG (GNU Privacy Guard) [5], технологии Blockchain [6].

Инфраструктура открытых ключей предоставляет возможность создавать, распространять, использовать сертификаты открытых ключей, которые позволяют аутентифицировать данные и пользователей, управлять ими, а также контролировать доступ к защищенным ресурсам в сети. В основе PKI лежит асимметричная криптография. Сертификаты открытого ключа пользователя, выданные доверенными центрами сертификации, являются ключевым компонентом PKI, поскольку удостоверяют подлинность ключа. Сертификат содержит информацию о пользователе, его открытом ключе, сроке действия сертификата и другую информацию, необходимую для аутентификации.

Достоинства PKI:

- обеспечение подлинности и целостности файлов с помощью их электронной подписи;

- обеспечение конфиденциальности файлов с помощью их шифрования;

- аутентификация пользователей с помощью проверки знания ими закрытых ключей, связанных с открытыми ключами в их сертификатах;

- совместимость с различными программными системами за счет использования международных стандартов X.509 и PKCS (Public Key Cryptography Standards);

- возможность управлять доступом пользователей к файлам, применяя различные уровни прав доступа.

Недостатки PKI:

- сложность создания, настройки и управления (особенно для частных лиц и организаций, не имеющих достаточных компетенций в сфере информационных технологий);

- сложность безопасного обновления сертификатов и их отзыва до истечения срока действия;

- риск компрометации закрытых ключей пользователей (особенно при их генерации на сервере);

- необходимость в дополнительных затратах на приобретение лицензионного ПО.

Открытый стандарт OpenPGP предоставляет инфраструктуру для создания ключей шифрования и управления ими, а также для защиты электронной почты и других данных. OpenPGP использует асимметричную криптографию и пары открытого и закрытого ключей пользователей для защиты их данных. GnuPG – бесплатная реализация OpenPGP, позволяющая пользователям шифровать и подписывать свои данные и сообщения, используя стандарты асимметричной криптографии. GnuPG – проект GNU, доступный бесплатно в большинстве ОС, включая Linux, MacOS и Windows.

Сертификаты открытого ключа в OpenPGP содержат информацию о владельце ключа, включая его имя и адрес электронной почты, дату создания сертификата и др. Они также могут содержать информацию о том, кто подтвердил имя владельца ключа (возможно, что сертификат подтвержден более чем одним лицом).

Достоинства OpenPGP и GnuPG:

- обеспечение подлинности, целостности и конфиденциальности файлов;

- отказ от использования доверенных удостоверяющих центров (центров сертификации) для выпуска сертификатов открытых ключей пользователей;

- открытость исходного кода с реализацией.

Недостатки OpenPGP и GnuPG:

- сложность использования для частных лиц и небольших организаций;

- сложность построения надежной сети доверия (в том числе и из-за транзитивного доверия, которое может привести к недостаточно надежному подтверждению подлинности открытого ключа участника сети доверия).

Блокчейн – технология создания и использования распределенной БД, информация в которой хранится в виде блоков, соединенных между собой цепочкой. Каждый блок содержит некоторую информацию (например, о завершенной транзакции) и имеет уникальный идентификатор, называемый хешем. Одной из основных особенностей этой технологии является децентрализация: информация хранится не на централизованном сервере, а на многих компьютерах, называемых сетевыми узлами. Каждый узел содержит копию всей цепочки, что обеспечивает безопасность и надежность хранения данных.

Сеть доверия на основе технологии Blockchain строится с помощью алгоритма консенсуса, который позволяет участникам сети достигать единства относительно состояния распределенного хранилища.

Достоинства технологии Blockchain:

- децентрализация файлового хранилища и управления сетью доверия;

- обеспечение подлинности, целостности и конфиденциальности файлов.

Недостатки технологии Blockchain:

- сложность использования частными лицами и небольшими организациями;

- сложность масштабирования;

- возможность дополнительных временных задержек при обработке транзакций из-за того, что новые блоки добавляются после выполнения сложных математических вычислений;

- необходимость достижения консенсуса участников сети доверия.

В таблице приведены результаты сравнения рассмотренных технологий.

PKI имеет широкое распространение, часто используется в корпоративной среде, но требует дополнительных затрат на управление ключами и сертификатами. OpenPGP и GnuPG предоставляют свободу и гибкость, не требуют централизованного управления, но могут быть сложными для неопытных пользователей. Блокчейн может обеспечить безопасность и де-

## Сравнение технологий создания сетей доверия

## Comparison of technologies for creating trust networks

Методы и средства	Шифрование данных	Аутентификация пользователей	Электронная подпись данных	Сложность использования частными лицами и малыми предприятиями	Необходимость поддерживать список доверенных удостоверяющих центров
PKI	Да	Да	Да	Да	Да
OpenPGP	Да	Да	Да	Да	Нет
GnuPG	Да	Да	Да	Нет	Нет
Blockchain	Нет	Да	Да	Да	Нет

централизацию, но требует больших затрат на вычислительные ресурсы и может быть неэффективным из-за большого объема вычислений при добавлении новых блоков в цепочку.

В предлагаемом решении будут объединены преимущества технологий PKI и GnuPG. Эти технологии обеспечивают безопасность, децентрализацию и гибкость, а также доступны широкому кругу пользователей.

Алгоритм работы файлообменника следующий.

1. Регистрация пользователя на сервере (с заданием имени пользователя, его пароля, генерацией пары ключей электронной подписи и выдачей сертификата открытого ключа).

2. Идентификация и аутентификация пользователя по уникальному имени (логину) и паролю, удовлетворяющему ограничениям на его минимальную длину и сложность.

3. Авторизация пользователя (получение доступа в файловые хранилища других пользователей, которые ранее были ему предоставлены).

4. Прием или отклонение запросов в друзья или приглашений в файловые хранилища, полученных от других пользователей.

5. Отправка запросов в друзья другим пользователям или удаление пользователей из списка своих друзей.

6. Создание файловых хранилищ с указанием минимального уровня доверия, при котором другой пользователь сможет получить к нему доступ (значения от 0 до 1). Для расчета уровня доверия для нового пользователя, получающего доступ в файловое хранилище, используется теорема Байеса:  $P(H | E) = \frac{P(E | H)P(H)}{P(E)}$ .

Значение априорной вероятности  $P(H)$  принимается равным 0,9, если пользователь находится в друзьях у владельца хранилища, и 0,8

в противном случае. Значение условной вероятности  $P(E|H)$  принимается равным уровню доверия владельца хранилища. Маргинальная вероятность  $P(E)$  определяется на основе уровня доверия всех друзей пользователя, допущенных в его файловое хранилище:  $P(E) = \sum P(E|H_i)P(H_i)$ .

7. Загрузка и выгрузка файлов в хранилище. При выгрузке файла в хранилище пользователя файл шифруется и снабжается электронной подписью с помощью закрытого ключа владельца хранилища. При загрузке файла из хранилища другого пользователя проверяется подпись под файлом с помощью сертификата открытого ключа этого пользователя и файл расшифровывается.

8. Завершение сеанса работы пользователя.

## Серверная часть файлообменника

Представим БД приложения, которая состоит из таблиц. В таблице users содержится информация о зарегистрированных пользователях, в storages – о файловых хранилищах, созданных пользователями, в friend\_requests\_users – о запросах пользователей в друзья, в friend\_users – о пользователях, являющихся друзьями, в storage\_invites – о приглашениях в файловые хранилища пользователей, в storage\_users – о пользователях файловых хранилищ. Сертификаты открытых ключей всех пользователей хранятся в хранилище на сервере, выполняющем в приложении роль удостоверяющего центра, а закрытый ключ пользователя и сертификат его открытого ключа – на устройстве пользователя (закрытый ключ – в защищенной части его памяти). Для управления БД используется система PostgreSQL.

Взаимодействие серверной части файлообменника с БД реализовано с помощью фрейм-

ворка Exposed [7]. Разработаны модели и сущности для каждой таблицы из БД. Все возможные запросы к серверу разделены на три части: запросы для регистрации и авторизации пользователя, для получения информации о пользователях, для получения информации о файловых хранилищах.

Примеры запросов:

- certificate – получение сертификата сервера, необходимого для установления защищенного соединения по протоколу SSL/TLS;
- registration – регистрация нового пользователя с уникальным именем;
- users – получение списка всех зарегистрированных пользователей приложения;
- addFriendRequest – добавление запроса в друзья;
- storageInvites – получение приглашений конкретного пользователя в файловые хранилища;
- createStorage – создание файлового хранилища.

Запросы передаются в формате json.

### Клиентская часть приложения

Проектирование клиентской части приложения проводилось на основе принципов разделения задач, управления пользовательским интерфейсом на основе моделей данных, единого достоверного источника информации (Single Source of Truth, SSoT) [8], однонаправленного потока данных (Unidirectional Data Flow, UDF) [9]. С учетом этих принципов в приложении были выделены три уровня: пользовательского интерфейса, отображающего данные приложения на экране, домена для упрощения и повторного использования взаимодействия между интерфейсом и данными и уровень данных, содержащий бизнес-логику приложения и предоставляющий его данные.

Для обеспечения кроссплатформенности разрабатываемого приложения использован инструмент Kotlin Multiplatform [10], который позволяет разработчикам создавать приложения для разных платформ (Android, iOS, Web и других) с применением общего кода на языке Kotlin. Для пробного тестирования разработано клиентское мобильное приложение для ОС Android с использованием фреймворка Jetpack Compose.

Пользовательский интерфейс клиентского приложения состоит из пяти частей: экраны авторизации и регистрации пользователя, домашний экран, экраны поиска пользователей, со-

здания и выбора файлового хранилища, загрузки и выгрузки файлов.

При проведении тестирования разработанного файлообменника использовались как реальные мобильные устройства (Xiaomi Mi 6 с процессором Snapdragon 835 и 6 Гб оперативной памяти, работающий под управлением ОС Android 9), так и их эмуляторы (Google Pixel 6 Pro с процессором Google Tensor и 12 Гб оперативной памяти под управлением ОС Android 12, Google Pixel 7 Pro с процессором Google Tensor 2 и 12 Гб оперативной памяти под управлением ОС Android 13). Эмуляторы выполнялись на персональном компьютере с процессором AMD Ryzen 7 5800x3d и 64 Гб оперативной памяти, работающем под управлением ОС Windows 11 Pro.

Для тестирования были созданы 20 пользователей, для каждого из которых создавалось файловое хранилище. Случайным образом между пользователями распределялись приглашения к добавлению в друзья и к доступу в файловые хранилища. Для загрузки на сервер и выгрузки с него использовался файл (RAR-архив) размером 94,3 Мб. Время выгрузки для разных устройств и эмуляторов составляло от 12 до 20 секунд, а время загрузки – от 17 до 25 секунд. Увеличение времени загрузки файла связано с особенностью работы серверной части приложения: при выгрузке время считается до окончания выгрузки файла на сервер, после чего файл шифруется, но это время не включается в результат, так как операция производится вне соединения с клиентской частью. При загрузке файла клиент ожидает окончания расшифрования файла перед его передачей с сервера.

Результаты проведенного тестирования показали, что приложение работоспособно и позволяет пользователям создавать собственные сети доверия для безопасного и быстрого обмена файлами между собой. Разработанное приложение предназначено прежде всего для корпоративных информационных систем бюджетных организаций и предприятий малого и среднего бизнеса, а также для частных лиц.

### Заключение

В статье представлены результаты разработки приложения для обмена файлами между пользователями в глобальной сети, обеспечивающего конфиденциальность, подлинность и целостность передаваемых файлов. Разработанное приложение объединяет преимущества

инфраструктуры открытых ключей (PKI) и сетей доверия типа GnuPG. При регистрации пользователя создаются пара его асимметричных ключей и сертификат открытого ключа (в формате стандарта X.509), а закрытый ключ сохраняется в защищенной от несанкционированного доступа части памяти его устройства. В создаваемых на сервере хранилищах файлы находятся в зашифрованном виде, а при передаче между сервером и клиентом используется протокол безопасной передачи данных HTTPS.

По сравнению с другими решениями для обмена файлами между пользователями Интернета разработанное приложение имеет ряд преимуществ.

- Использует более надежный способ подтверждения подлинности и целостности передаваемых файлов (каждый файл снабжается электронной подписью отправителя, которая проверяется получателем с помощью сертификата открытого ключа отправителя, выданного сервером сети доверия). Подбор закрытого ключа подписи в отличие от подбора пароля (при использовании подтверждения подлинности по имени и паролю) является практически не решаемой сегодня задачей (при правильном выборе длины ключа). Использование сервера сети доверия в качестве удостоверяющего центра для выдачи сертификатов открытых ключей пользователей файлообменника избавляет участников сети доверия от необходимости обращения в сторонние удостоверяющие центры и ведения на всех своих устройствах списков доверенных удостоверяющих центров.

- Обеспечивает конфиденциальность файлов с помощью их симметричного шифрования не только при передаче файлов по сети, но и при хранении на сервере, что минимизирует риски, связанные с похищением файлов пользователей с сервера.

- Использует сертификаты открытых ключей в формате общепринятого международного стандарта X.509, что позволяет применять

для работы с ними имеющиеся в открытом доступе инструментальные программные средства.

- Позволяет управлять допуском к файловому хранилищу пользователя сети доверия других пользователей сети с помощью рассчитываемых по теореме Байеса значений уровней доверия.

Таким образом, предложенное решение обеспечивает большую безопасность и дополнительную функциональность по сравнению с передачей файлов с помощью других программных средств. Выбор инструментальных средств разработки позволил существенно уменьшить трудоемкость переноса разработанного клиентского приложения на другие операционные платформы.

Разработанное для обмена файлами приложение может применяться при организации и проведении учебного процесса в университете (например, при выдаче студентам индивидуальных заданий и приеме от них отчетов, содержащих и программный код). Также оно может использоваться как группами частных пользователей Интернета, так и пользователями корпоративных информационных систем предприятий малого и среднего бизнеса.

Направлениями развития разработанного файлообменника для устранения имеющихся в текущей версии недостатков могут быть замена аутентификации пользователей сети доверия по имени и паролю аутентификацией с помощью их закрытых ключей и одного из криптографических протоколов с нулевым разглашением; создание клиентских приложений для других операционных платформ; использование только криптоалгоритмов из российских стандартов; расчет уровня доверия к пользователю при предоставлении ему доступа в файловое хранилище с помощью разных методов для выбора наилучшего; реализация хранения закрытых ключей пользователей не в защищенной части памяти их мобильных устройств, а на внешних устройствах.

#### Список литературы

1. Дикий Д.И., Артемьева В.Д. Протокол передачи данных MQTT в модели удаленного управления правами доступа для сетей Интернета // Науч.-технич. вестн. информационных технологий, механики и оптики. 2019. Т. 19. № 1. С. 109–117. doi: 10.17586/2226-1494-2019-19-1-109-117.
2. Hou R., Ren G., Zhou C., Yue H., Liu H., Liu J. Analysis and research on network security and privacy security in ubiquitous electricity Internet of things. Computer Communications, 2020, vol. 158, pp. 64–72. doi: 10.1016/j.comcom.2020.04.019.
3. Kent D., Cheng B.H.C., Siegel J. Assuring vehicle update integrity using asymmetric public key infrastructure (PKI) and public key cryptography (PKC). SAE Int. J. Transp. Cyber. & Privacy, 2019, vol. 2, no. 2, pp. 141–158. doi: 10.4271/11-02-02-0013.
4. Yeoh W.-Z., Teh J.S., Chen J. Automated enumeration of block cipher differentials: An optimized branch-and-bound GPU framework. JISA, 2022, vol. 65, art. 103087. doi: 10.1016/j.jisa.2021.103087.

5. Ray P.P., Chowhan B., Kumar N., Almogren A. BioTHR: Electronic health record servicing scheme in IoT-blockchain ecosystem. *IEEE Internet of Things J.*, 2021, vol. 8, no. 13, pp. 10857–10872. doi: 10.1109/JIOT.2021.3050703.
6. Mukherjee P., Pradhan C. Blockchain 1.0 to Blockchain 4.0 – The evolutionary transformation of blockchain technology. In: *Blockchain Tech.: Applications and Challenges*. ISRL, 2021, vol. 203, pp. 29–49. doi: 10.1007/978-3-030-69395-4\_3.
7. Candas S., Muschner Ch., Buchholz S. et al. Code exposed: Review of five open-source frameworks for modeling renewable energy systems. *Renewable and Sustainable Energy Reviews*, 2022, vol. 161, art. 112272. doi: 10.1016/j.rser.2022.112272.
8. Mulyana E., Fakhri G. Network automation with a single source of truth in a heterogeneous environment. *Int. J. on Electrical Eng. and Informatics*, 2022, vol. 14, no. 1, pp. 92–100. doi: 10.15676/ijeei.2022.14.1.6.
9. Воротицкий Ю.И., Румас Р.А. Архитектура аппаратно-программного средства однонаправленной передачи данных в компьютерных сетях // Докл. БГУИР. 2023. Т. 21. № 3. С. 96–101. doi: 10.35596/1729-7648-2023-21-3-96-101.
10. Кузнецова С.В. Особенности кроссплатформенной разработки мобильных приложений с использованием Xamarin // Тр. МАИ. 2022. № 125. С. 1–27. doi: 10.34759/trd-2022-125-21.

Software &amp; Systems

doi: 10.15827/0236-235X.142.230-237

2024, 37(2), pp. 230–237

### Secure file sharing based on trust networks and public key certificates using a developed application

Pavel B. Khorev <sup>1</sup>✉, Dmitry A. Losev <sup>1</sup><sup>1</sup> National Research University "Moscow Power Engineering Institute",  
Moscow, 111250, Russian Federation

#### For citation

Khorev, P.B., Losev, D.A. (2024) 'Secure file sharing based on trust networks and public key certificates using a developed application', *Software & Systems*, 37(2), pp. 230–237 (in Russ.). doi: 10.15827/0236-235X.142.230-237

#### Article info

Received: 29.09.2023

After revision: 18.12.2023

Accepted: 25.12.2023

**Abstract.** The paper analyzes the shortcomings of existing software tools for file sharing (including cloud storage, messenger programs, e-mail). To eliminate the shortcomings, the authors of the paper have developed a project of a client-server application based on building a trust network using public key certificates of its participants. The project uses symmetric and asymmetric cryptography algorithms, methods for building trust networks and using public key certificates to ensure the authenticity, confidentiality and integrity of data. The implemented server part and the client mobile application allow users to exchange files with each other within one specific file storage. Uploaded files have electronic signatures of their creators and are stored on the server in encrypted form. When developing the application, the authors used the technology that facilitates its portability to other operating platforms. The client application is built and tested on devices running the Android operating system. The developed application has the functionality necessary for file hosting and ensures the authenticity, integrity and confidentiality of transferred files. This allows it to be used both by groups of private users and in corporate information systems.

**Keywords:** client-server application, public key certificates, trust network, secure files sharing

#### References

1. Dikii, D.I., Artemeva, V.D. (2019) 'MQTT data protocol in remote access control management model for internet networks', *Sci. and Tech. J. of Inform. Tech., Mech. and Optics*, 19(1), pp. 109–117 (in Russ.). doi: 10.17586/2226-1494-2019-19-1-109-117.
2. Hou, R., Ren, G., Zhou, C., Yue, H., Liu, H., Liu, J. (2020) 'Analysis and research on network security and privacy security in ubiquitous electricity Internet of things', *Computer Communications*, 158, pp. 64–72. doi: 10.1016/j.comcom.2020.04.019.
3. Kent, D., Cheng, B.H.C., Siegel, J. (2019) 'Assuring vehicle update integrity using asymmetric public key infrastructure (PKI) and public key cryptography (PKC)', *SAE Int. J. Transp. Cyber. & Privacy*, 2(2), pp. 141–158. doi: 10.4271/11-02-02-0013.

4. Yeoh, W.-Z., The, J.S., Chen, J. (2022) 'Automated enumeration of block cipher differentials: An optimized branch-and-bound GPU framework', *JISA*, 65, art. 103087. doi: 10.1016/j.jisa.2021.103087.
5. Ray, P.P., Chowhan, B., Kumar, N., Almogren, A. (2021) BIoTHR: Electronic health record servicing scheme in IoT-blockchain ecosystem. *IEEE Internet of Things J.*, 8(13), pp. 10857–10872. doi: 10.1109/JIOT.2021.3050703.
6. Mukherjee, P., Pradhan, C. (2021) 'Blockchain 1.0 to Blockchain 4.0 – The evolutionary transformation of blockchain technology', in *Blockchain Tech.: Applications and Challenges. ISRL*, 203, pp. 29–49. doi: 10.1007/978-3-030-69395-4\_3.
7. Candas, S., Muschner, Ch., Buchholz, S. et al. (2022) 'Code exposed: Review of five open-source frameworks for modeling renewable energy systems', *Renewable and Sustainable Energy Reviews*, 161, art. 112272. doi: 10.1016/j.rser.2022.112272.
8. Mulyana, E., Fakhri, G. (2022) 'Network automation with a single source of truth in a heterogeneous environment', *Int. J. on Electrical Eng. and Informatics*, 14(1), pp. 92–100. doi: 10.15676/ijeei.2022.14.1.6.
9. Varatnitski, Y.I., Rumas, R.A. (2023) 'Architecture of hardware and software for unidirectional data transmission in computer networks', *Doklady BGUIR*, 21(3), pp. 96–10 (in Russ.). doi: 10.35596/1729-7648-2023-21-3-96-101.
10. Kuznetsova, S.V. (2022) 'Features of cross-platform mobile applications development using Xamarin', *Proc. of MAI*, (125), pp. 1–27 (in Russ.). doi: 10.34759/trd-2022-125-21.

**Авторы**

**Хорев Павел Борисович**<sup>1</sup>, к.т.н., доцент,  
профессор, pbkh@yandex.ru  
**Лосев Дмитрий Алексеевич**<sup>1</sup>,  
магистрант, lda-1028@mail.ru

**Authors**

**Pavel B. Khorev**<sup>1</sup>, Cand. of Sci. (Engineering),  
Associate Professor, Professor, pbkh@yandex.ru  
**Dmitry A. Losev**<sup>1</sup>, Graduate Student,  
lda-1028@mail.ru

<sup>1</sup> Национальный исследовательский университет  
«Московский энергетический институт»,  
г. Москва, 111250, Россия

<sup>1</sup> National Research University  
"Moscow Power Engineering Institute",  
Moscow, 111250, Russian Federation

## Исследование эффективности алгоритма саранчи в задачах сегментации изображений

Д.Ю. Запорожец<sup>1</sup>, Ю.А. Кравченко<sup>1</sup>, Н.В. Кулиева<sup>1</sup>,  
М.М. Семенова<sup>1</sup>, Э.В. Кулиев<sup>1</sup>✉

<sup>1</sup> Южный федеральный университет, г. Таганрог, 347922, Россия

### Ссылка для цитирования

Запорожец Д.Ю., Кравченко Ю.А., Кулиева Н.В., Семенова М.М., Кулиев Э.В. Исследование эффективности алгоритма саранчи в задачах сегментации изображений // Программные продукты и системы. 2024. Т. 37. № 2. С. 238–246. doi: 10.15827/0236-235X.142.238-246

### Информация о статье

Группа специальностей ВАК: 1.2.1

Поступила в редакцию: 02.10.2023

После доработки: 23.01.2024

Принята к публикации: 07.02.2024

**Аннотация.** Современные методы обработки изображений играют ключевую роль во множестве приложений, включая медицинскую диагностику, робототехнику и компьютерное зрение. Важной задачей является сегментация изображений, то есть выделение на них объектов и областей. В данной статье проведен анализ применения для решения этой задачи трех биоинспирированных алгоритмов: пчелиной колонии, светлячкового и алгоритма саранчи. Сделан обзор существующих методов сегментации изображений, описаны их ограничения в контексте современных требований высокой точности и эффективности. Особое внимание уделено описанию механизмов и принципов работы алгоритмов пчелиной и светлячковой колоний, их характеристикам, преимуществам и недостаткам при решении задач сегментации. Проведен анализ применения алгоритма саранчи в данных задачах, включая его адаптацию и экспериментальное исследование на разнообразных изображениях. Адаптация алгоритма включает внедрение специализированных механизмов для улучшения точности и эффективности поиска оптимального решения, что позволяет преодолеть ограничения существующих методов. Представлены экспериментальные исследования, которые подтверждают превосходство алгоритма саранчи над другими биоинспирированными подходами и демонстрируют его эффективность на изображениях. Преимущество предложенного алгоритма саранчи состоит в сегментации изображений и открывает новые перспективы для его применения в различных областях обработки данных. Особое внимание уделено таким аспектам алгоритма саранчи, как гибкость, масштабируемость и способность адаптироваться к сложным задачам сегментации. Это делает его особенно ценным для применения в различных областях, требующих высокой точности и эффективности обработки изображений.

**Ключевые слова:** сегментация изображений, биоинспирированный алгоритм, алгоритм саранчи, пчелиный алгоритм, светлячковый алгоритм, моделирование поведения саранчи

**Благодарности.** Исследование выполнено в Южном федеральном университете за счет гранта РФФИ № 23-21-00089, <https://rscf.ru/project/23-21-00089/>

**Введение.** В современном информационном обществе обработка и анализ изображений играют важную роль во множестве областей, включая медицину, робототехнику, компьютерное зрение и многие другие. Особенно актуальна задача сегментации изображений, заключающаяся в выделении отдельных объектов или областей [1]. Эффективная сегментация изображений имеет решающее значение для получения более точных данных, а также для дальнейшего анализа и распознавания объектов.

Научная проблема и противоречие заключаются в необходимости повышения точности и эффективности существующих методов сегментации, особенно в условиях разнообразных и сложных сцен [2, 3].

Цель данного исследования – анализ и сравнение биоинспирированных алгоритмов, в частности, алгоритма саранчи, для оптимизации задачи сегментации изображений.

Задачи исследования включают обзор существующих методов сегментации, адаптацию алгоритма саранчи к этой задаче и анализ его эффективности по сравнению с другими биоинспирированными алгоритмами.

Экспериментальное исследование проведено на наборе изображений различной природы и разрешения, что позволяет всесторонне оценить преимущества и недостатки каждого метода.

### Обзор существующих исследований

Алгоритм саранчи и его применение описаны в различных работах по сегментации изображений.

Авторами исследования [2] разработан метод сегментации изображений саранчи, который акцентирует внимание на выборе группы цветовых признаков. Методика особенно важна для идентификации саранчи на фоне

сложных и переменных окружающих условий. Исследователи использовали различные цветовые пространства для извлечения характеристик цвета как саранчи, так и фона. Затем применяли модель машины опорных векторов (SVM) для классификации изображений саранчи. Подход дополнен алгоритмом Оцу для уточнения сегментации, что позволило достичь более точных результатов по сравнению с методами, использующими одиночные цветовые признаки. Это исследование предоставляет научную основу для точного и эффективного обнаружения районов встречи саранчи и их плотности, что является ключевым для современного мониторинга и раннего предупреждения о катастрофах, связанных с этими насекомыми.

Новаторский алгоритм для автоматического выбора классов пикселей в задачах сегментации изображений представлен в [3]. Алгоритм назван Locust Search (LS) и основан на моделировании поведения стай саранчи. Отличаясь от большинства существующих эволюционных алгоритмов, LS избегает концентрации особей в оптимальных позициях, что помогает избежать таких проблем, как преждевременная конвергенция к субоптимальным решениям и ограниченный баланс между исследованием и эксплуатацией. Алгоритм был проверен на нескольких эталонных функциях и изображениях, демонстрируя эффективность в плане точности и надежности. Это исследование вносит значительный вклад в область биоинспирированных алгоритмов и их применение в компьютерном зрении.

Основным недостатком алгоритма саранчи, как и большинства эволюционных алгоритмов, является проблема достижения субоптимальных результатов, выявленная в работе [4]. Она обусловлена недостаточным балансом между исследованием и эксплуатацией в стратегиях поиска, что приводит к ряду ограничений в применении алгоритма, а именно:

- как и другие эволюционные алгоритмы, алгоритм саранчи может неэффективно распределять ресурсы между исследованием новых областей поискового пространства и эксплуатацией уже найденных решений;
- высокая вычислительная сложность алгоритма саранчи может ограничивать его применимость в задачах с ограниченными ресурсами или требующих быстрого времени отклика;
- эффективность алгоритма саранчи сильно зависит от правильной настройки его параметров, что может быть непростой задачей в зависимости от конкретного случая использования.

Для преодоления этих ограничений были предложены различные улучшения и модификации алгоритма. Например, в работе [5] представлен улучшенный вариант алгоритма саранчи (LS-II), который включает в себя более точные модели движения саранчи, основанные на вероятностных методах, что позволяет улучшить баланс между исследованием и эксплуатацией.

### Обзор существующих методов сегментации изображений

Сегментация изображений является важной задачей в области компьютерного зрения и обработки изображений, направленной на разделение изображения на семантически связанные регионы [4]. В работе представлен обзор двух методов сегментации на основе эволюционных алгоритмов – пчелиного и светлячкового.

*Пчелиный алгоритм* – это биологически инспирированный эволюционный метод оптимизации, основанный на поведении пчелиных колоний [6]. Он применяется для решения различных задач оптимизации и поиска, включая сегментацию изображений. Его основная идея заключается в имитации поведения пчелиных рабочих, собирающих пыльцу и нектар, для нахождения оптимальных решений.

Пчелиный алгоритм для сегментации изображений обычно представляет решение как множество сегментов и определяет функции приспособленности на основе характеристик цвета, текстуры и геометрии [6, 7]. Пчелы-разведчики исследуют пространство решений, обмениваясь информацией о качестве найденных сегментаций и влияя друг на друга.

Алгоритм выделяется своей способностью к эффективному поиску и адаптации к сложным ландшафтам пространства решений. Он хорошо подходит для задач, где требуется баланс между исследованием (поиском новых решений) и эксплуатацией (оптимизацией известных решений). Однако его эффективность сильно зависит от настройки параметров, и он может быть вычислительно затратным.

*Светлячковый алгоритм* – еще один метод оптимизации, заимствованный у живой природы [8]. Он применяется в задачах оптимизации, где решение может быть представлено как точка в многомерном пространстве. В контексте сегментации изображений алгоритм используется для поиска оптимальных параметров сегментации, таких как количество сегментов, пороги схожести и др. Основная идея

заключается в том, чтобы светлячки (потенциальные решения) двигались к более ярким, что соответствует улучшению качества сегментации [9].

Алгоритм хорошо применим в оптимизации и локализации глобальных экстремумов в многомерных пространствах. Особенно полезен в приложениях, где необходимо найти глобальное оптимальное решение. Но, как и пчелиный алгоритм, он требует тщательной настройки параметров и может быть неэффективным при решении проблем с высокой размерностью.

Оба алгоритма представляют интересные методы для решения задачи сегментации изображений, основаны на природных аналогиях и способны исследовать множество потенциальных решений.

Традиционные подходы включают методы, основанные на пороговом значении, кластеризации, а также использующие границы и текстуры. Эти методы хорошо работают при стабильных условиях освещения и высоком качестве изображений, но часто терпят неудачу при изменчивости условий или наличии шума [10].

Эволюционные алгоритмы, включая пчелиный и светлячковый, представляют собой новые подходы к сегментации. Пчелиный алгоритм основан на стратегии поиска пищи у пчел. Он использует концепцию исследования и эксплуатации, позволяя эффективно анализировать пространство решений. Светлячковый алгоритм основан на привлекательной динамике световых сигналов. Эти методы показывают многообещающие результаты в сегментации за счет адаптивности и способности к глобальному поиску оптимальных решений [11–13], но сталкиваются с такими ограничениями, как высокая вычислительная сложность и трудности в настройке параметров, что может снижать их эффективность в реальных условиях.

### Постановка задачи для сегментации изображений

Для достижения оптимальной сегментации необходимо определить параметры каждого сегмента таким образом, чтобы максимально точно соответствовать структуре объектов на изображении [11, 14, 15].

Пусть дано изображение  $I$ , которое необходимо разделить на сегменты  $S_1, S_2, \dots, S_N$ , представляющие собой объекты на изображении. Каждый сегмент  $S_i$  характеризуется набором параметров  $\Theta_i$ , которые определяют его положение, форму, размер и другие характеристики.

Целью задачи сегментации является нахождение оптимальных значений параметров  $\Theta_i$  для каждого сегмента  $S_i$  таким образом, чтобы достичь наилучшей сегментации, максимально соответствующей структуре объектов на изображении. Это может быть сформулировано как задача оптимизации следующего вида:

$$\min_{\Theta_1, \dots, \Theta_N} \varepsilon(I, S_1, \dots, S_N),$$

где  $\varepsilon$  – функция оценки качества сегментации, зависящая от изображения и параметров сегментов (чем меньше значение  $\varepsilon$ , тем лучше качество сегментации).

Для решения данной задачи с использованием алгоритма саранчи требуется определить способы представления параметров  $\Theta_i$ , операторы скрещивания и мутации, а также процедуру выбора наилучших решений (сегментов) на каждой итерации [16]. После настройки параметров алгоритма он может быть запущен для оптимизации сегментации изображения путем итеративного улучшения параметров сегментов на протяжении нескольких поколений.

### Алгоритм саранчи

Алгоритм представляет собой метод оптимизации, инспирированный поведением саранчи и базирующийся на концепции роя [10]. Он включает два уникальных механизма поиска, основанных на природных явлениях: операторы одиночной фазы и социальной фазы. Благодаря этим уникальным механизмам алгоритм саранчи преодолевает некоторые трудности, обычно связанные с другими алгоритмами, такими как преждевременная сходимость и ограниченное разнообразие решений.

В недавнем времени в сфере компьютерного зрения появилась потребность в разработке более точных моделей движения саранчи, чем те, которые могут быть созданы на основе простого наблюдения за ее поведением. Эксперименты, связанные с методами отслеживания насекомых, позволили создать более точные модели их движения. Важной характеристикой этих новых моделей является введение вероятностных элементов для имитации процесса принятия решений, используемого при борьбе с саранчой [17].

Алгоритм саранчи является перспективным методом оптимизации с имитацией природного поведения саранчи, и его развитие, основанное на более точных моделях движения, может привести к улучшению эффективности и точности оптимизации в различных областях применения.

В компьютерном моделировании процесс поиска саранчи часто включает использование алгоритма, основанного на наблюдениях за стадным поведением насекомых [12–14]. Это глобальный метод оптимизации, в котором поисковые агенты моделируются в виде индивидуальных особей, движущихся в многомерном пространстве решений.

Индивидуальные позиции саранчи определяются ограничениями пространства решений, где  $S_i$  представляет позицию индивида  $i$ , а  $d$  – размерность пространства. Позиции саранчи представляют потенциальные решения задачи оптимизации. Подобно другим методам роевой оптимизации алгоритм саранчи включает в себя итерационный процесс, в котором саранчи обновляют свои позиции в соответствии с поведенческими фазами – уединенной и социальной [15].

В *уединенной фазе* саранчи перемещаются индивидуально в поисках оптимальных решений, избегая сближения с другими особями. Это моделируется с помощью сил притяжения и отталкивания между индивидуальными саранчами. Общая социальная сила, действующая на индивида  $i$  в итерации  $t$ , определяется как сумма сил притяжения и отталкивания:

$$F_i^t = \sum_{j \neq i} f_{ij}^t,$$

где  $f_{ij}^t$  – сила взаимодействия между саранчами  $i$  и  $j$ , определенная как

$$f_{ij}^t = \begin{cases} 1 \\ d_{ij}^t \end{cases}, \text{ если } d_{ij}^t \neq 0,$$

где  $d_{ij}^t$  – евклидово расстояние между  $S_i$  и  $S_j$  на  $t$ -й итерации, а  $d_{\min}$  и  $d_{\max}$  – параметры, отражающие силу притяжения и масштаб длины.

*Социальная фаза* направлена на уточнение лучших решений, сгруппированных в подмножество  $P$  из общего множества позиций  $S$ . Для этого применяется оператор уединенной фазы, генерирующий случайные позиции внутри подпространства для каждой позиции из  $P$ .

Новые позиции  $S_{\text{new}}$  в подпространстве для каждой позиции  $S_{\text{best}}$  из  $P$  вычисляются следующим образом:

$$S_{\text{new}} = S_{\text{best}} + \alpha R,$$

где  $\alpha$  – коэффициент, а  $R$  – случайный вектор, элементы которого взяты из равномерного распределения. В итоге обновленное решение  $S_i^{\text{new}}$  назначается индивиду на следующей итерации. Алгоритм поиска саранчи объединяет уединенные и социальные фазы для оптимизации позиций саранчи в многомерном пространстве решений.

## Применение алгоритма саранчи для сегментации изображений

Данный алгоритм саранчи основан на эволюционной оптимизации, что дает возможность находить оптимальные параметры для сложных моделей сегментации [17]. Это особенно полезно для изображений с непредсказуемыми особенностями и изменчивыми условиями.

В отличие от методов, требующих больших объемов данных, алгоритм саранчи может работать с ограниченными данными, поскольку он способен итеративно улучшать параметры, находя оптимальные значения для задачи сегментации [7]. Кроме того, эволюционные методы, включая алгоритм саранчи, могут легко настраиваться для различных типов изображений, что делает их гибкими и адаптивными к разнообразным сценам. Использование данного алгоритма для сегментации изображений – перспективный подход, который может преодолеть ограничения других методов, обеспечивая более точные и адаптивные результаты даже в сложных условиях.

Представим модифицированный алгоритм поиска саранчи.

1. Инициализация популяции и определение параметров алгоритма. Алгоритм начинается с создания начальной популяции особей. Каждая особь представляет собой набор параметров сегментации  $\Theta_i$  для каждого сегмента  $S_i$ . Также определяются параметры алгоритма: размер популяции  $N$ , вероятности скрещивания  $\rho_n$  и мутации  $\rho_m$ , количество итераций  $T$  и другие.

2. Отбор особей с учетом поведения саранчи. Для оператора отбора особей используется аналогия с поведением саранчи, где более успешные особи оказывают большее влияние на выбор следующего поколения. Оценка качества особей может быть основана на значении функции оценки  $E$ , которая измеряет степень соответствия сегментации изображения структуре объектов. Пусть  $F(i)$  представляет оценку пригодности особи  $i$ , тогда вероятность выбора особи  $i$  пропорциональна ее пригодности:

$$P_i = \frac{F(i)}{\sum_{j=1}^N F(j)}.$$

3. Скрещивание с учетом локальной информации. Оператор скрещивания комбинирует параметры родительских особей для создания новых потомков. При этом локальная информация, извлекаемая из соседних сегмен-

тов на изображении, может быть использована для более адаптивного создания потомков. Например, параметры потомка  $\Theta_{i, \text{new}}$  могут быть вычислены как средневзвешенное параметров родителей  $\Theta_{i, \text{parent1}}$  и  $\Theta_{i, \text{parent2}}$  с учетом локальной информации:

$$\Theta_{i, \text{new}} = \alpha \Theta_{i, \text{parent1}} + (1 - \alpha) \Theta_{i, \text{parent2}},$$

где  $\alpha$  – коэффициент, учитывающий влияние каждого родителя.

4. Мутация и внесение случайности. Оператор мутации добавляет случайность в популяцию, что помогает избежать застревания в локальных оптимумах. Параметры сегментации немного изменяются, внося разнообразие в популяцию и расширяя область поиска. Для каждой особи  $i$  параметры  $\Theta_i$  могут быть изменены с использованием небольшого случайного приращения:

$$\Theta_{i, \text{new}} = \Theta_i + \Delta\Theta,$$

где  $\Delta\Theta$  – случайный вектор с небольшими компонентами.

Предложенный алгоритм генетического поиска саранчи объединяет эволюционную оптимизацию с принципами поведения саранчи, что делает его способным к настройке параметров сегментации изображений на основе образа поведения саранчи в природе [14]. Этот подход позволяет более адаптивно и эффективно решать задачу сегментации в условиях разнообразных сцен и сложных объектов на изображении.

## Программная реализация

Представим реализацию биоинспирированного алгоритма, адаптированного для задачи сегментации изображений с использованием концепции поведения саранчи в природе.

В приведенном далее коде представлены ключевые этапы алгоритма, включая инициализацию популяции особей, операторы отбора, скрещивания и мутации, а также основной цикл оптимизации:

```
import numpy as np
# image = load_image("image.jpg")
# segmentation_parameters = initialize_seg-
mentation_parameters()

# Параметры алгоритма
population_size = 50
crossover_probability = 0.7
mutation_probability = 0.1
num_iterations = 100

# Создание начальной популяции
population = []
for _ in range(population_size):
    individual = initialize_individual(seg-
mentation_parameters)
```

```
population.append(individual)

def fitness(individual):
    pass

def select_parents(population):
    fitness_values = [fitness(individual)
for individual in population]
    total_fitness = sum(fitness_values)
    probabilities = [fitness_value / to-
tal_fitness for fitness_value in fit-
ness_values]

    parents_indices = np.ran-
dom.choice(len(population), size=(2, popu-
lation_size), p=probabilities)
    parents = [(population[parent_indi-
ces[0]], population[parent_indices[1]]) for
parent_indices in parents_indices]

    return parents

def crossover(parent1, parent2):
    alpha = np.random.uniform(0, 1)
    child = {}
    for parameter_name in segmentation_pa-
rameters:
        parameter_parent1 = parent1[parame-
ter_name]
        parameter_parent2 = parent2[parame-
ter_name]
        child[parameter_name] = alpha * pa-
rameter_parent1 + (1 - alpha) * parame-
ter_parent2

    return child

def mutate(individual):
    mutation_rate = 0.1
    mutated_individual = {}

    for parameter_name in segmentation_pa-
rameters:
        parameter_value = individual[param-
eter_name]

        mutation = np.random.uniform(-muta-
tion_rate, mutation_rate)
        mutated_value = parameter_value +
mutation

        mutated_individual[parameter_name]
= mutated_value

    return mutated_individual

best_solution = None
best_fitness = float("-inf")

for iteration in range(num_iterations):
    parents = select_parents(population)
    offspring = []

    for i in range(population_size):
        parent1, parent2 = parents[i]
        child = crossover(parent1, parent2)

        if np.random.rand() < muta-
tion_probability:
            child = mutate(child)

        offspring.append(child)
```

```
population = offspring

for individual in population:
    current_fitness = fitness(individual)
    if current_fitness > best_fitness:
        best_fitness = current_fitness
        best_solution = individual

print("Лучшая сегментация:")
print(best_solution)
print("Оценка пригодности:", best_fitness)
```

Этот код может быть доработан и настроен под конкретные задачи сегментации и наборы данных.

Результатом работы алгоритма являются лучшая сегментация изображения и соответствующая оценка ее пригодности. Предлагаемый код служит в качестве отправной точки для исследования и разработки эффективных методов сегментации изображений с использованием биоинспирированных алгоритмов в естественных условиях.

В предложенной реализации биоинспирированного алгоритма, интегрированного с концепцией поведения саранчи, представлен метод для сегментации изображений, который позволяет находить оптимальные параметры для моделей сегментации с учетом особенностей естественных эволюционных процессов. Метод объединения генетического алгоритма и поведения саранчи открывает перспективы в области сегментации изображений, особенно в медицинской диагностике и компьютерном зрении.

### Экспериментальное исследование

В статье представлены результаты экспериментального исследования, направленного на сравнительный анализ эффективности трех различных биоинспирированных алгоритмов в контексте задачи сегментации изображений [15] – пчелиной колонии, светлячкового и алгоритма саранчи, сосредотачиваясь на их способности выделять объекты на изображениях и отделять их от фона.

Сегментация изображений является ключевой задачей в области компьютерного зрения, и ее эффективное решение имеет множество приложений, начиная от медицинской диагностики и автоматической обработки изображений, заканчивая робототехникой и анализом социальных данных. Для достижения целей был подготовлен набор из трех разнообразных изображений различной природы и разрешения. Подходы к сегментации должны были

справляться с этим разнообразием сцен и обеспечивать точное выделение объектов.

В ходе исследования были заданы одинаковые параметры для каждого из алгоритмов, чтобы обеспечить сравнение их производительности. После применения каждого алгоритма к набору изображений проведена оценка качества сегментации на основе выбранных метрик. Сравнение алгоритмов проводилось по множеству аспектов, включая точность выделения объектов, адаптивность к разнообразию сцен, скорость сходимости и другие характеристики.

Опишем более детально подготовку набора изображений и параметров алгоритмов.

Для целей эксперимента был составлен набор изображений, содержащих три разнообразные сцены. Он включал изображения различной природы и разрешения, созданные с целью охватить разнообразные сценарии, встречающиеся при сегментации изображений. Для обеспечения точного сравнения каждый из алгоритмов использовал одинаковые параметры [11]. Они включали в себя такие настройки, как размер популяции, количество итераций, коэффициенты мутации и другие параметры, влияющие на производительность алгоритма.

Были использованы предварительно размеченные области объектов для каждого изображения как эталонные для сравнения с результатами сегментации [14].

Визуальный пример процесса сегментации изображений с использованием трех различных биоинспирированных алгоритмов – алгоритма пчелиной колонии, светлячкового и алгоритма саранчи представлен на рисунке 1. Он включает исходное изображение, а также результаты сегментации, полученные при применении трех алгоритмов.

По результатам эксперимента можно сделать вывод, что в задаче сегментации изображений алгоритм саранчи демонстрирует наилучшие результаты среди представленных генетических алгоритмов. Его способность точно выделять объекты на изображении и создавать четкие границы позволяет достичь высокой степени точности. Этот алгоритм обладает хорошей способностью адаптации к различным типам изображений и сложным условиям, что делает его предпочтительным для решения подобных задач.

На рисунке 2 представлен график кривых сходимости методов сегментации изображений. График визуализирует процесс сходимости трех алгоритмов. Каждая кривая на графике отража-

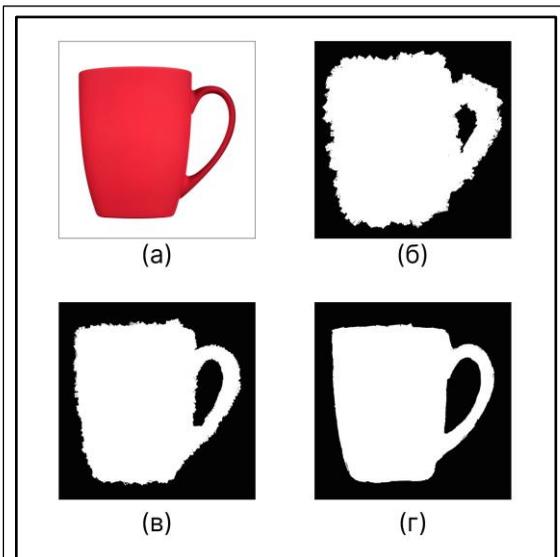


Рис. 1. Сегментация изображения:  
 а) исходное изображение; б) использование алгоритма пчелиной колонии,  
 в) алгоритма светлячка,  
 г) алгоритма саранчи

Fig. 1. Image segmentation: a) original image;  
 б) segmentation by a bee colony algorithm,  
 в) firefly algorithm, г) locust algorithm

ет изменение значения выбранной метрики в зависимости от числа итераций алгоритмов.

Цель этого графика – проиллюстрировать динамику процесса оптимизации и сравнить эффективность различных методов сегментации. Более явное и стремительное увеличение значения на кривой сходимости указывает на более быструю и точную сегментацию изображений при применении соответствующего метода. График наглядно показывает превосходство алгоритма саранчи в процессе сходимости и его способность обеспечивать более высокое качество сегментации по сравнению с другими алгоритмами [1].

В начальных итерациях наблюдается, что все три метода демонстрируют рост значений функции цели. Это может быть связано с процессом инициализации и с начальной фазой поиска оптимального решения. Однако по мере увеличения количества итераций кривые сходимости указывают на улучшение качества сегментации для всех трех методов. Кривая сходимости алгоритма саранчи демонстрирует наиболее стремительный рост значения функции цели на ранних итерациях, что свидетельствует о его высокой способности к быстрой адаптации к изображениям и к выделению объектов.

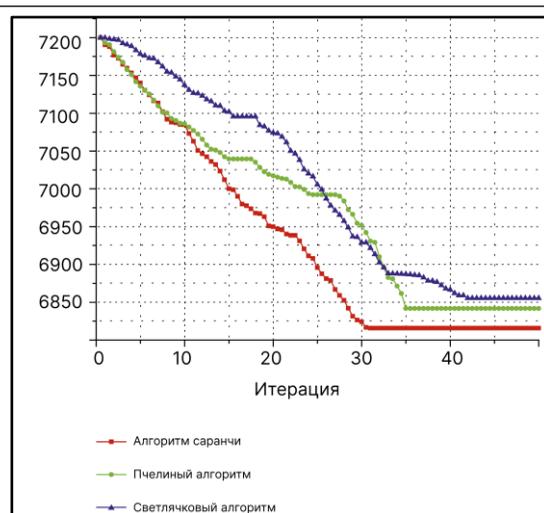


Рис. 2. Кривые сходимости методов сегментации

Fig. 2. Convergence curves of segmentation methods

### Заключение

Авторы статьи оценивают алгоритм саранчи как новаторский подход в области оптимизации и сегментации изображений. Данный метод демонстрирует значительные преимущества благодаря уникальному сочетанию механизмов одиночной и социальной фазы поиска. Он успешно преодолевает те сложности, которые часто возникают в других методах оптимизации.

Алгоритм саранчи показал высокую эффективность в задачах компьютерного зрения. Особенно важной является его способность к созданию более точных моделей движения саранчи и использованию вероятностных элементов для имитации процесса принятия решений, что делает алгоритм перспективным инструментом для решения сложных задач оптимизации и сегментации изображений, особенно в современных приложениях, где требуются высокая точность и эффективность.

Важным направлением исследований являются дальнейшее развитие и адаптация алгоритма саранчи. Особое внимание следует уделить улучшению его точности и скорости работы, а также адаптации к более широкому спектру задач обработки изображений.

Возможно, перспективным направлением станет исследование комбинации алгоритма саранчи с другими методами обработки изображений, то есть создание гибридных систем, способных повысить общую эффективность и точность процесса сегментации.

## Список литературы

1. Lei T., Jia X., Liu T., Liu S., Meng H., Nandi A.K. Adaptive morphological reconstruction for seeded image segmentation. *IEEE Transactions on Image Processing*, 2019, vol. 28, no. 11, pp. 5510–5523. doi: 10.1109/TIP.2019.2920514.
2. Zhang W., Bai F. Locust image segmentation by fusion of multiple colors. *Proc. 5th ICISCE*, 2018, pp. 295–300. doi: 10.1109/ICISCE.2018.00069.
3. Cuevas E., Fausto F., González A.A. Locust search algorithm applied to multi-threshold segmentation. In: *ISRL. New Advancements in Swarm Algorithms: Operators and Applications*, 2019, pp. 211–240. doi: 10.1007/978-3-030-16339-6\_8.
4. Camarena O., Cuevas E., Pérez-Cisneros M., Fausto F., González A., Valdivia A. Ls-II: An improved locust search algorithm for solving optimization problems. *Math. Problems in Eng.*, 2018, vol. 2018, art. 4148975. doi: 10.1155/2018/4148975.
5. Zaruba D., Zaporozhets D., Kureichik V. Artificial bee colony algorithm-A novel tool for VLSI placement. In: *AISC. Proc. ИПТ*, 2016, vol. 450, pp. 433–442. doi: 10.1007/978-3-319-33609-1\_39.
6. Debelee T.G., Schwenker F., Rahimeto S., Yohannes D. Evaluation of modified adaptive k-means segmentation algorithm. *Computational Visual Media*, 2019, vol. 5, no. 4, pp. 347–361. doi: 10.1007/s41095-019-0151-2.
7. Bhandari K. A novel beta differential evolution algorithm-based fast multilevel thresholding for color image segmentation. *Neural Computing and Applications*, 2020, vol. 32, pp. 4583–4613. doi: 10.1007/s00521-018-3771-z.
8. Кулиев Э.В., Шабатуков М.М., Кравченко Д.Ю. Основные методы классификации изображений // *ИС&ИТ-2021: тр. Междунар. науч.-технич. конгресса*. 2021. С. 241–245.
9. Курейчик В.В., Бова В.В., Курейчик В.В. Метод светлячковой оптимизации для решения комбинаторно-логических задач на графах // *ИТНОУ*. 2021. № 1. С. 20–24.
10. Кравченко Ю.А., Курсытис И.О. Биоинспирированный алгоритм приобретения новых знаний на основе классификации информационных ресурсов // *Изв. ЮФУ. Технич. науки*. 2019. № 2. С. 15–26.
11. Cuevas E., Gálvez J., Avalos O. Parameter estimation for chaotic fractional systems by using the locust search algorithm. *Computación y Sistemas*, 2017, vol. 21, no. 2, pp. 369–380. doi: 10.13053/cys-21-2-2741.
12. González A., Cuevas E., Fausto F., Valdivia A., Rojas R. A template matching approach based on the behavior of swarms of locust. *Applied Intelligence*, 2017, vol. 47, no. 4, pp. 1087–1098. doi: 10.1007/s10489-017-0937-9.
13. Luo H., Sun B., Zhou H., Cao W. Image segmentation with multi-feature fusion in compressed domain based on region-based graph. *IAJIT*, 2023, vol. 20, no. 2, pp. 159–169. doi: 10.34028/iajit/20/2/2.
14. Lahouaoui L., Abdelhak D. Markov random field model and expectation of maximization for images segmentation. *IJECS*, 2023, vol. 29, no. 2, pp. 772–779. doi: 10.11591/ijeecs.v29.i2.pp772-779.
15. Ding W., Zhao Y., Zhang R. An adaptive multi-threshold segmentation algorithm for complex images under unstable imaging environment. *IJCAT*, 2019, vol. 61, no. 4, pp. 265–272. doi: 10.1504/IJCAT.2019.103295.
16. Zhou X., Li Z., Xie H. et al. Leukocyte image segmentation based on adaptive histogram thresholding and contour detection. *Current Bioinformatics*, 2020, vol. 15, no. 3, pp. 187–195. doi: 10.2174/1574893614666190723115832.
17. Dhal K.G., Das A., Ray S., Gálvez J., Das S. Nature-inspired optimization algorithms and their application in multi-thresholding image segmentation. *Archives of Computational Methods in Engineering*, 2019, vol. 27, no. 3, pp. 855–888. doi: 10.1007/s11831-019-09334-y.

Software &amp; Systems

doi: 10.15827/0236-235X.142.238-246

2024, 37(2), pp. 238–246

## Study of the effectiveness of the search algorithm in image segmentation problems

Dmitry Yu. Zaporozhets<sup>1</sup>, Yury A. Kravchenko<sup>1</sup>, Nina V. Kulieva<sup>1</sup>,  
Marina M. Semenova<sup>1</sup>, Elmar V. Kuliev<sup>1</sup> ✉

<sup>1</sup> Southern Federal University, Taganrog, 347922, Russian Federation

## For citation

Zaporozhets, D.Yu., Kravchenko, Yu.A., Kulieva, N.V., Semenova, M.M., Kuliev, E.V. (2024) 'Study of the effectiveness of the search algorithm in image segmentation problems', *Software & Systems*, 37(2), pp. 238–246 (in Russ.). doi: 10.15827/0236-235X.142.238-246

## Article info

Received: 02.10.2023

After revision: 23.01.2024

Accepted: 07.02.2024

**Abstract.** Modern image processing techniques play a key role in a variety of applications, including medical diagnostics, robotics and computer vision. Image segmentation (selecting objects and areas in images) is an important task. The paper analyzes the application of three bioinspired algorithms: the bee colony algorithm, the firefly algorithm and the locust algorithm for the image segmentation problem. It presents a review of existing image segmentation methods, describes their limitations in terms of modern requirements for high accuracy and efficiency. Particular attention is paid to describing the mechanisms and operating principles of the bee colony and firefly colony algorithms, including their characteristics, advantages and disadvantages when solving segmentation problems. The authors of the paper have carried out an analysis of the locust algorithm regarding the image segmentation problem, including algorithm adaptation and experimental study using a variety of images. Algorithm adaptation involves introducing application-specific mechanisms to improve the accuracy and efficiency of finding an optimal solution that overcomes the limitations of existing methods. The paper presents experimental studies that confirm the locust algorithm superiority over other bioinspired approaches and demonstrates its effectiveness on a wide range of images. The advantages of the proposed locust algorithm are in image segmentation; it also opens new prospects when applied in various data processing fields. The paper focuses on such aspects of the locust

algorithm as its flexibility, scalability, and the ability to adapt to complex segmentation problems. It makes the algorithm particularly valuable for applying in various fields requiring high accuracy and efficiency of image processing.

**Keywords:** image segmentation, bioinspired algorithm, locust algorithm, bee algorithm, firefly algorithm, locust behavior modeling

**Acknowledgements.** The study was performed by the grant from the Russian Science Foundation no. 23–21–00089, <https://rscf.ru/project/23-21-00089/>, in the Southern Federal University

### References

1. Lei, T., Jia, X., Liu, T., Liu, S., Meng, H., Nandi, A.K. (2019) ‘Adaptive morphological reconstruction for seeded image segmentation’, *IEEE Transactions on Image Processing*, 28(11), pp. 5510–5523. doi: 10.1109/TIP.2019.2920514.
2. Zhang, W., Bai, F. (2018) ‘Locust image segmentation by fusion of multiple colors’, *Proc. 5th ICISCE*, pp. 295–300. doi: 10.1109/ICISCE.2018.00069.
3. Cuevas, E., Fausto, F., González, A.A. (2019) ‘Locust search algorithm applied to multi-threshold segmentation’, in *ISRL. New Advancements in Swarm Algorithms: Operators and Applications*, pp. 211–240. doi: 10.1007/978-3-030-16339-6\_8.
4. Camarena, O., Cuevas, E., Pérez-Cisneros, M., Fausto, F., González, A., Valdivia, A. (2018) ‘Ls-II: An improved locust search algorithm for solving optimization problems’, *Math. Problems in Eng.*, 2018, art. 4148975. doi: 10.1155/2018/4148975.
5. Zaruba, D., Zaporozhets, D., Kureichik, V. (2016) ‘Artificial bee colony algorithm-A novel tool for VLSI placement’, in *AISC. Proc. IITI*, 450, pp. 433–442. doi: 10.1007/978-3-319-33609-1\_39.
6. Debelee, T.G., Schwenker, F., Rahimeto, S., Yohannes, D. (2019) ‘Evaluation of modified adaptive k-means segmentation algorithm’, *Computational Visual Media*, 5(4), pp. 347–361. doi: 10.1007/s41095-019-0151-2.
7. Bhandari, K. (2020) ‘A novel beta differential evolution algorithm-based fast multilevel thresholding for color image segmentation’, *Neural Computing and Applications*, 32, pp. 4583–4613. doi: 10.1007/s00521-018-3771-z.
8. Kuliev, E.V., Shabatukov, M.M., Kravchenko, D.Yu. (2021) ‘Image classification method’, *Proc. IS&IT*, pp. 241–245 (in Russ.).
9. Kureichik, V., Bova, V., Kureichik Jr, V. (2021) ‘Firely optimization method for solving combinatorial logic problems on graphs’, *ITNOU*, (1), pp. 20–24 (in Russ.).
10. Kravchenko, Yu.A., Kursityis, I.O. (2019) ‘Bioinspired algorithm for acquiring new knowledge on the basis of the information resources classification’, *Izvestiya SFedU. Eng. Sci.*, (2), pp. 15–26 (in Russ.).
11. Cuevas, E., Gálvez, J., Avalos, O. (2017) ‘Parameter estimation for chaotic fractional systems by using the locust search algorithm’, *Computación y Sistemas*, 21(2), pp. 369–380. doi: 10.13053/cys-21-2-2741.
12. González, A., Cuevas, E., Fausto, F., Valdivia, A., Rojas, R. (2017) ‘A template matching approach based on the behavior of swarms of locust’, *Applied Intelligence*, 47(4), pp. 1087–1098. doi: 10.1007/s10489-017-0937-9.
13. Luo, H., Sun, B., Zhou, H., Cao, W. (2023) ‘Image segmentation with multi-feature fusion in compressed domain based on region-based graph’, *IJIT*, 20(2), pp. 159–169. doi: 10.34028/ijit/20/2/2.
14. Lahouaoui, L., Abdelhak, D. (2023) ‘Markov random field model and expectation of maximization for images segmentation’, *IJECS*, 29(2), pp. 772–779. doi: 10.11591/ijeecs.v29.i2.pp772-779.
15. Ding, W., Zhao, Y., Zhang, R. (2019) ‘An adaptive multi-threshold segmentation algorithm for complex images under unstable imaging environment’, *IJCAT*, 61(4), pp. 265–272. doi: 10.1504/IJCAT.2019.103295.
16. Zhou, X., Li, Z., Xie, H. et al. (2020) ‘Leukocyte image segmentation based on adaptive histogram thresholding and contour detection’, *Current Bioinformatics*, 15(3), pp. 187–195. doi: 10.2174/1574893614666190723115832.
17. Dhal, K.G., Das, A., Ray, S., Gálvez, J., Das, S. (2019) ‘Nature-inspired optimization algorithms and their application in multi-thresholding image segmentation’, *Archives of Computational Methods in Engineering*, 27(3), pp. 855–888. doi: 10.1007/s11831-019-09334-y.

### Авторы

**Запорожец Дмитрий Юрьевич**<sup>1</sup>,  
к.т.н., доцент,  
duzaporozhets@sfedu.ru

**Кравченко Юрий Алексеевич**<sup>1</sup>, д.т.н.,  
профессор, yakravchenko@sfedu.ru

**Кулиева Нина Владимировна**<sup>1</sup>,  
ассистент, holopova@sfedu.ru

**Семенова Марина Максимовна**<sup>1</sup>,  
аспирант, msemenova@sfedu.ru

**Кулиев Эльмар Валерьевич**<sup>1</sup>, к.т.н.,  
доцент, ekuliev@sfedu.ru

### Authors

**Dmitry Yu. Zaporozhets**<sup>1</sup>,  
Cand. of Sci. (Engineering), Associate Professor,  
duzaporozhets@sfedu.ru

**Yury A. Kravchenko**<sup>1</sup>, Dr.Sc. (Engineering),  
Professor, yakravchenko@sfedu.ru

**Nina V. Kulieva**<sup>1</sup>, Assistant,  
holopova@sfedu.ru

**Marina M. Semenova**<sup>1</sup>, Postgraduate Student,  
msemenova@sfedu.ru

**Elmar V. Kuliev**<sup>1</sup>, Cand. of Sci. (Engineering),  
Associate Professor, ekuliev@sfedu.ru

<sup>1</sup> Южный федеральный университет,  
г. Таганрог, 347922, Россия

<sup>1</sup> Southern Federal University,  
Taganrog, 347922, Russian Federation

## Разработка программного обеспечения технологического процесса электронно-лучевой сварки тонкостенных изделий

С.О. Курашкин <sup>1</sup>✉, В.В. Бухтояров <sup>2</sup>, В.С. Тынченко <sup>1</sup>, Е.В. Шуткина <sup>1</sup>

<sup>1</sup> Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева, г. Красноярск, 660037, Россия

<sup>2</sup> Институт нефти и газа Сибирского федерального университета, г. Красноярск, 660041, Россия

### Ссылка для цитирования

Курашкин С.О., Бухтояров В.В., Тынченко В.С., Шуткина Е.В. Разработка программного обеспечения технологического процесса электронно-лучевой сварки тонкостенных изделий // Программные продукты и системы. 2024. Т. 37. № 2. С. 247–253. doi: 10.15827/0236-235X.142.247-253

### Информация о статье

Группа специальностей ВАК: 2.3.3

Поступила в редакцию: 12.09.2023

После доработки: 17.01.2024

Принята к публикации: 06.02.2024

**Аннотация.** Целью данного исследования является повышение эффективности электронно-лучевой сварки (ЭЛС) тонкостенных конструкций. Для этого предлагается разработать и реализовать программные методы автоматизированного планирования и контроля процесса ЭЛС, включая статистическую и интеллектуальную обработку данных и математическую модель для расчета распределения температуры. На данный момент существует несколько решений для моделирования и управления процессом ЭЛС, но они имеют недостатки, такие как длительность моделирования в MATLAB при оптимизации параметров и коммерческое программное обеспечение, ориентированное на конкретное оборудование заказчика. Авторы предлагают автоматизированную систему управления ЭЛС (АСУ ЭЛС) тонкостенных деталей, разработанную на основе математической модели для расчета температурного распределения на поверхности свариваемых деталей. Эта система позволит моделировать процесс для разных сплавов с учетом геометрии и теплофизических параметров, планировать и прогнозировать, а также управлять им. Данные функции представлены в виде модулей, которые предоставляют настройку управляемых и регистрируемых параметров, целевых значений и допустимых диапазонов, что позволяет отслеживать и контролировать процесс и его отклонения. Система также может интегрироваться с различным электронно-лучевым оборудованием. Применение разработанной АСУ ЭЛС позволяет оптимизировать параметры, минимизировать дефекты сварных соединений и получать соединения требуемых размеров. Предметом исследования являются данные измерений параметров процесса ЭЛС, которые используются для контроля, оценки устойчивости и прогнозирования. АСУ ЭЛС может найти применение на предприятиях, занимающихся ЭЛС.

**Ключевые слова:** автоматизированное управление, моделирование, электронно-лучевая сварка, технологический процесс, аэрокосмические конструкции

**Введение.** Одним из наиболее высокотехнологичных и высокопроизводительных способов сварки является *электронно-лучевая сварка* (ЭЛС). Моделирование, сходное с электронно-лучевыми процессами сварки, показало, что динамическое позиционирование электронного пучка при создании нескольких источников обеспечивает возможность ввода тепла в несколько участков сварного соединения, что, в свою очередь, позволяет получить высококачественные соединения сплавов разных металлов [1–3]. Изучая процессы при ЭЛС, исследуют и импульсную модуляцию [4], которая приводит к изменению тепловложения. Технологические требования к качеству ЭЛС [5] напрямую влияют на результативность и правильность выполнения АСУ ЭЛС. Например, автор работы [6] представляет свою систему управления установкой ЭЛС, где как наиболее важную выделяет задачу позиционирования

свариваемых деталей, однако технологические характеристики не позволяют заранее знать время, необходимое для достижения скорости сварки. Система, представляющая собой программу для управления именно вакуумным оборудованием, представлена в работе [7]. Результаты практического применения моделей управления ЭЛС, а также основные принципы построения и моделирования систем ЭЛС приведены в статьях [8–10].

Обсудим различные методы ЭЛС. Автор работы [11] рассматривает манипуляторы электронно-лучевой пушки и стола с серводвигателем для повышения эффективности работы. При исследовании влияния ЭЛС на металлы и их сплавы очень важным является исследование формирования сварных соединений, выполненных электронным лучом, с использованием промежуточных вставок [12]. Определяются режимы сварки (ток луча, скорость сварки,



- обработка совокупности измерений параметров, полученных с испытательного стенда, статистическими методами для выявления вероятных некорректных значений и предварительной оценки трендов;

- визуализация данных измерений параметров, полученных с установки ЭЛС;

- сохранение значений параметров, полученных с установки ЭЛС, в БД для последующего использования.

В части планирования и обработки параметрической информации ПС СВАРКА обеспечивает выполнение следующих функций:

- создание и генерация плана реализации технологического процесса на установке ЭЛС с учетом интервалов варьирования определенных для него параметров;

- сопоставление параметрической информации и выполняемого на установке ЭЛС технологического процесса, сигнализация при отклонении параметров от требуемых более чем на установленную величину;

- статистическая обработка результатов выполнения технологического процесса на установке ЭЛС, включая оценку достоверности получаемых данных;

- корреляционный и регрессионный анализ полученных параметрических данных, выполняемый в автоматизированном режиме.

В части моделирования и прогнозирования развития технологического процесса ЭЛС ПС СВАРКА обеспечивает выполнение следующих функций:

- моделирование различных реализаций процесса ЭЛС, в том числе следующих методов построения моделей: построение статистических моделей, моделей методами интеллектуального анализа данных, моделей непараметрическими методами;

- оценка качества моделирования, основанная на вычислении статистических критериев оценки регрессионных моделей и воспроизводимости результатов;

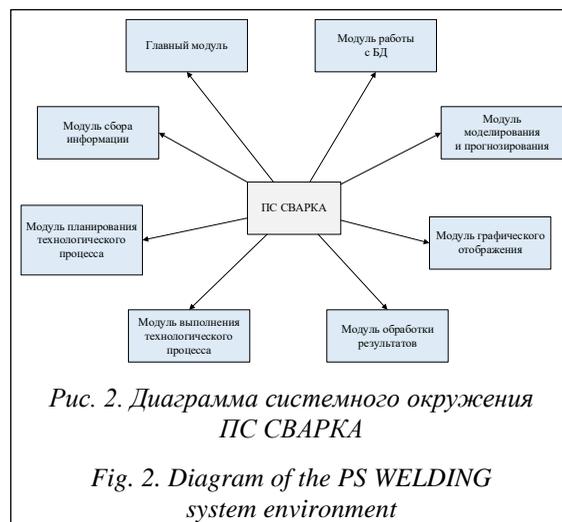
- прогнозирование параметров технологического процесса ЭЛС, включая штатные и нештатные ситуации, на основе построенных моделей;

- визуализация результатов моделирования и моделей прогнозирования;

- верификация моделей.

Диаграмма системного окружения разрабатываемой программы представлена на рисунке 2.

Программная система состоит из восьми взаимосвязанных логических модулей.



1. *Главный модуль программы* реализует связь между остальными семью модулями, а также функционал главной формы проекта.

2. *Модуль сбора информации с установки* предназначен для получения параметрической информации в процессе сварки на установке ЭЛС. Реализует взаимодействие с платами сбора данных и процесс мониторинга.

3. *Модуль планирования технологического процесса* предназначен для генерации планов технологических процессов на установке ЭЛС. Реализует функции по расчету значений варьируемых параметров в соответствии с заданными пользователем ограничениями и шагом дискретизации.

4. *Модуль выполнения технологического процесса* реализует четыре типа процесса, предоставляя информационное сопровождение технологического процесса и автоматизируя формирование протоколов.

5. *Модуль обработки результатов* автоматизирует расчет статистических критериев при обработке данных технологических процессов, в частности, реализует расчет критериев Кохрена и Фишера.

6. *Модуль графического отображения информации* структурирует предоставляемые пользователю данные в виде таблиц и графиков.

7. *Модуль моделирования и прогнозирования* реализует параметрические и непараметрические алгоритмы аппроксимации и позволяет провести моделирование согласно заданным настройкам выбранного алгоритма.

8. *Модуль работы с БД* является служебным и предназначен для взаимодействия ПС СВАРКА с БД, хранящейся на сервере MS SQL Server 2008.

### Результаты

Программа создавалась в среде разработки Embarcadero C++ Builder для операционной системы Windows и написана на языке программирования высокого уровня C++. Данные в ней хранятся в формате Microsoft SQL Server 2008.

Блок-схема работы системы представлена на рисунке 3.

После запуска программы есть возможность создать новый технологический процесс или запустить уже имеющийся в БД. Под технологическим процессом понимаются набор параметров самого процесса сварки и параметры материала; если меняется хотя бы один из параметров, меняется сам техпроцесс.

При разработке нового техпроцесса создается его план, сохраняющийся в БД, в котором инициализируют теплофизические, геометрические и технологические параметры. Данный этап выполняется модулем планирования технологического процесса. При создании плана можно выбрать варьируемые параметры из БД или ввести новые.

На основе сгенерированного плана технологического процесса существует возможность моделирования процесса сварки или прогнозирования его параметров. При моделировании ЭЛС можно выбрать один из трех алгоритмов: ИНС, непараметрическая регрессия, восстановление символьной зависимости. В результате моделирования будет получена информация о количестве прогонов алгоритма и средней ошибке моделирования.

В процессе моделирования программа позволяет провести прогноз на основе его результатов, кроме того, прогнозирование можно провести и без этапа модели-

рования, запустив сгенерированный план прогноза. Данный план получает данные для моделирования, на основе которых с помощью регрессионной модели рассчитывает изменения технологических параметров в зависимости от времени.

Если нет необходимости в этапах моделирования и прогнозирования, на основе плана технологического процесса просто проводится натуральный эксперимент.

Результаты проведенной работы сохраняются в БД программы, после чего отображаются на экране.

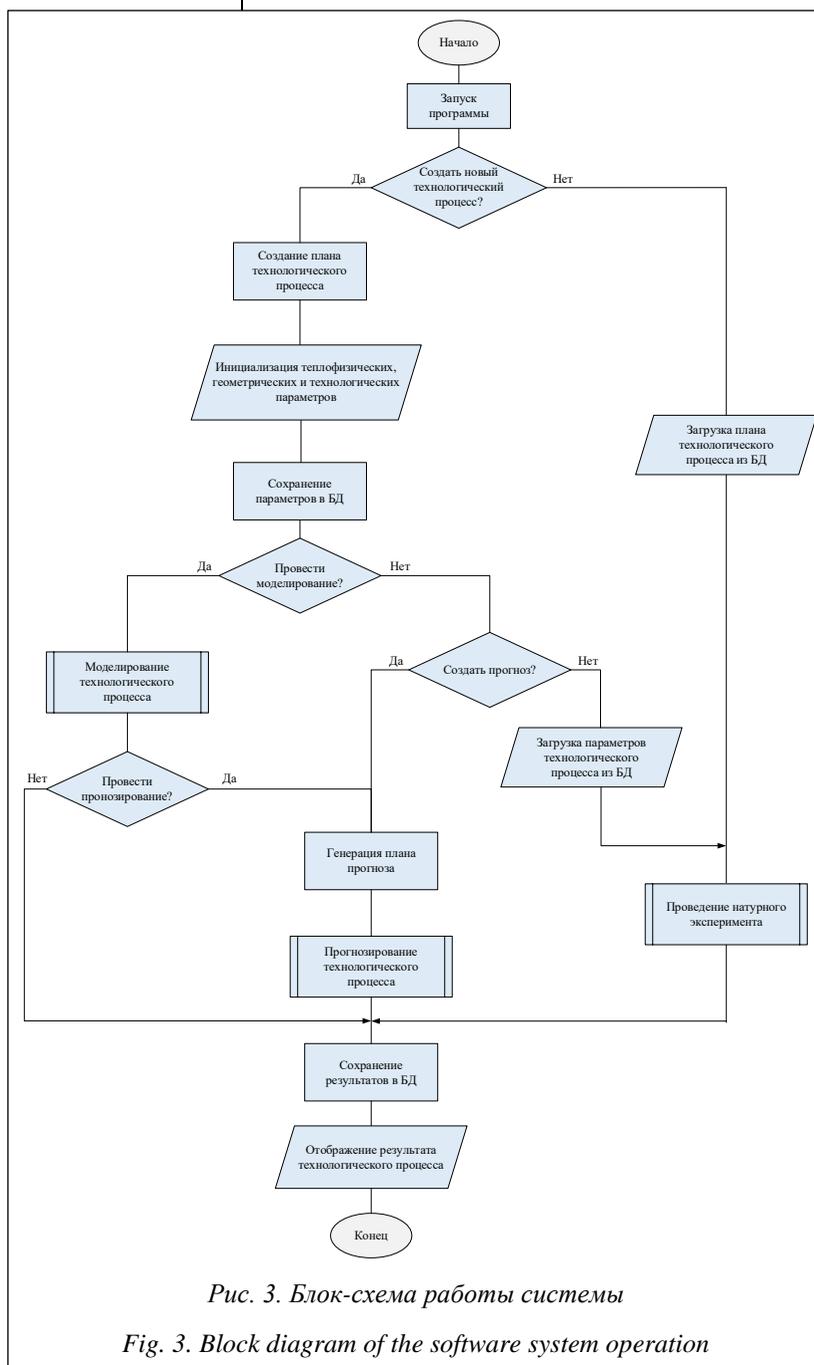


Рис. 3. Блок-схема работы системы

Fig. 3. Block diagram of the software system operation

Все операции с программой осуществляются через оконный интерфейс Windows-приложения (<http://www.swsys.ru/uploaded/image/2024-2/14.jpg>).

План процесса включает в себя его описание, дату создания, тип проводимой серии, варьируемые параметры, хранящиеся в БД, и этапы проведения.

Сам процесс представляет собой натуральный эксперимент, в ходе которого можно реализовать четыре типа проводимых стендовых опытов, предоставляя информационное сопровождение и автоматизируя формирование протоколов. В результате с испытательного стенда получаем информацию о технологических параметрах, таких как ускоряющее напряжение, ток фокусировки, ток сварки и скорость сварки, зависящих от заданных технологическими настройками технологического процесса.

Кроме того, существует возможность с главного окна программы запустить модули обработки результатов, моделирования, прогнозирования и модуль доступа к БД.

### Обсуждение

Для реализации прогнозирования и моделирования в разработанной программной системе используются общепринятые математические методы, которые хорошо зарекомендовали себя, а также дали приемлемый результат в ходе тестирования на реальных данных, полученных в результате сварки на реальном оборудовании. Проводимые эксперименты показали результаты прогнозирования технологического процесса, удовлетворяющие предъявленным требованиям. Однако существуют методы, демонстрирующие наилучшую степень предсказания построенной модели. Поэтому в дальнейшем планируется доработка представленной программы с использованием более современных методов машинного обучения.

В данном исследовании было проведено тестирование разработанной программной систе-

мы, показавшее ее работоспособность. Система позволяет выполнять сварку тонкостенных деталей и обеспечивает повторяемость технологического процесса за счет мониторинга процесса сварки и выбора оптимальных параметров технологического процесса.

### Заключение

В работе описано создание комплекса ориентированных на практику подходов к планированию, контролю и обработке результатов реализации технологического процесса ЭЛС. Полученные результаты реализованы в виде АСУ ЭЛС. В программе выполнены модули технологического процесса ЭЛС, моделирования и прогнозирования. Испытания АСУ ЭЛС проводились на электронно-лучевой установке А306.13. Замеры глубины и ширины сварного шва показали отклонение от требуемых значений в пределах 1 %. С применением разработанной АСУ абсолютное отклонение по глубине проплавления составило 0,000086 мм, по ширине 0,00006 мм. Без оптимального управления абсолютное отклонение по глубине проплавления составило 0,06 мм, по ширине 0,017 мм.

Таким образом, полученные результаты свидетельствуют о повышении эффективности управления процессом ЭЛС при использовании разработанной системы. Ее применение позволяет снизить количество дефектов в зоне сварного соединения и получить стыковые соединения заданных геометрических размеров.

Дальнейшее исследование предполагает разработку и внедрение специального диагностического контроля сварных швов неразрушающими методами (например, на основе видеонаблюдения и методов компьютерного зрения) с целью обеспечения надежности сварных конструкций. Предлагается также провести анализ других методов, используемых для прогнозирования на малых объемах данных, с целью улучшения точности предсказания.

### Список литературы

1. Ефимова И.Н. Особенности моделирования динамических режимов электронно-лучевой сварки // Образовательный, научный и инновационный процессы в нанотехнологиях: сб. докл. 2017. № 1. С. 70–75.
2. Chowdhury S., Yadaiah N., Khan S.M., Ozah R., Das B., Muralidhar M. A perspective review on experimental investigation and numerical modeling of electron beam welding process. *Materials Today: Proc.*, 2018, vol. 5, no. 2, pp. 4811–4817. doi: 10.1016/J.MATPR.2017.12.055.
3. Kaisheva D., Angelov V., Petrov P. Simulation of heat transfer at welding with oscillating electron beam. *Canadian J. of Phys.*, 2019, vol. 97, no. 10, pp. 1140–1146. doi: 10.1139/cjp-2018-0495.
4. Олшанская Т.В., Саломатова Е.С. Обзор современных способов управления электронным лучом при электронно-лучевой сварке // Вестн. ПНИПУ. Машиностроение, материаловедение. 2016. Т. 18. № 4. С. 169–187.
5. Драгунов В.К., Слива А.П., Жмурко И.Е. и др. Разработка технологии электронно-лучевой сварки сосудов высокого давления из стали 30ХГСА // Нефтяное хозяйство. 2016. № 12. С. 126–130.

6. Бурачевский Ю.А., Бакеев И.Ю., Миллер А.В., Новгородов Р.Г. Система управления и отображения информации электронно-лучевой установки // Докл. ТУСУР. 2017. Т. 20. № 2. С. 132–134. doi: 10.21293/1818-0442-2017-20-2-132-134.
7. Попенко Я.Н. Разработка программного обеспечения для управления вакуумным оборудованием установки электронно-лучевой сварки // Инструментальные методы и техника экспериментальной физики: матер. МНСК-2016. 2016. С. 35.
8. Симонов Д.К., Симонова М.Е. Автоматизация управления сварочным производством как фактор обеспечения надежности сварных соединений (на примере электронно-лучевой сварки) // Современные технологии в мировом научном пространстве: сб. ст. Междунар. науч. конф. 2019. № 1. С. 96–100.
9. Мадера А.Г., Кандалов П.И. Принципы построения программного комплекса для теплового проектирования электронных систем // Программные продукты и системы. 2018. Т. 31. № 3. С. 435–438. doi: 10.15827/0236-235X.123.435-438.
10. Паршуков Л.И., Ефремов Н.А. Исследование структуры и свойств сварного шва жаропрочного сплава ВЖ159-ИД // Тр. ВИАМ. 2019. № 3. С. 27–34. doi: 10.18577/2307-6046-2019-0-3-27-34.
11. Павлушин Н.В. Зонная электронно-лучевая сварка // Электронно-лучевая сварка и смежные технологии: матер. конф. 2020. С. 288–302.
12. Харитонов И.А., Драгунов В.К., Гончаров А.Л., Мартынов В.Н., Портнов М.А. Разработка способа электронно-лучевой сварки с использованием вставок // Сварка в России-2019: современное состояние и перспективы: матер. конф. 2019. С. 269.
13. Дармаев А.Н., Муравьев Э.К., Саблин В.М., Морев С.П. Особенности формирования электронного потока в электронно-оптических системах с автоэмиссионным катодом для установки электронно-лучевой сварки // Сварочное производство. 2021. № 6. С. 20–23.
14. Латышенко Г.И. Разработка автоматизированного комплекса электронно-лучевой аппаратуры и технологий для сварки тонкостенных узлов и деталей // Наука сегодня: проблемы и пути решения: матер. конф. 2018. С. 50–52.
15. Щербаков А.В., Гапонова Д.А., Родякина Р.В., Слива А.П. Электронно-лучевая технология аддитивного формообразования: вопросы управления и воспроизводимости технологических режимов // Технический оппонент. 2019. Т. 3. № 2. С. 30–37.

### Software for the technological process of electron beam welding of thin-walled products

Sergey O. Kurashkin <sup>1</sup>✉, Vladimir V. Bukhtoyarov <sup>2</sup>, Vadim S. Tynchenko <sup>1</sup>, Elizaveta V. Shutkina <sup>1</sup>

<sup>1</sup> Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, 660037, Russian Federation

<sup>2</sup> Institute of Oil and Gas of the Siberian Federal University, Krasnoyarsk, 660041, Russian Federation

#### For citation

Kurashkin, S.O., Bukhtoyarov, V.V., Tynchenko, V.S., Shutkina, E.V. (2024) 'Software for the technological process of electron beam welding of thin-walled products', *Software & Systems*, 37(2), pp. 247–253 (in Russ.). doi: 10.15827/0236-235X.142.247-253

#### Article info

Received: 12.09.2023

After revision: 17.01.2024

Accepted: 06.02.2024

**Abstract.** The purpose of the study is to increase the efficiency of electron-beam welding (EBW) of thin-walled structures. To this end, the authors of the paper propose developing and implementing software methods for EBW process automated planning and control, including statistical and data processing and a mathematical model for calculating temperature distribution. At the moment, there are several solutions for modeling and management of the EBW process, but they have drawbacks, such as the modeling time in MATLAB in parameter optimization and commercial software focused on specific customer hardware. The authors propose to develop an automated control system for EBW thin-walled parts (ACS EBW) based on a mathematical model for calculating the temperature distribution on the surface of welded parts. This system will allow to simulate the EBW process for different alloys taking into account their geometry and thermophysical parameters, as well as to control, plan and predict the EBW process. These functions are presented in the form of modules that provide the setting of managed and recorded parameters, target values and acceptable ranges, which allows you to monitor and

control the process and its deviations. The system can also be integrated with various electron beam equipment. The application of the EBW developed by the ACS makes it possible to optimize parameters, minimize defects in welded joints and obtain joints of the required dimensions. The subject of the study is data of measurements of parameters of the EBW process, which are used for monitoring, evaluation of stability and forecasting. ACS EBW can be used at enterprises engaged in electron-beam welding.

**Keywords:** automated control, modeling, electron beam welding, technological process, aerospace structures

### References

1. Efimova, I.N. (2017) 'Features of modeling dynamic modes of electron beam welding', *Proc. Educational, Sci. and Innovative Processes in Nanotechnology*, (1), pp. 70–75 (in Russ.).
2. Chowdhury, S., Yadaiah, N., Khan, S.M., Ozah, R., Das, B., Muralidhar, M. (2018) 'A perspective review on experimental investigation and numerical modeling of electron beam welding process', *Materials Today: Proc.*, 5(2), pp. 4811–4817. doi: 10.1016/J.MATPR.2017.12.055.
3. Kaisheva, D., Angelov, V., Petrov, P. (2019) 'Simulation of heat transfer at welding with oscillating electron beam', *Canadian J. of Phys.*, 97(10), pp. 1140–1146. doi: 10.1139/cjp-2018-0495.
4. Olshanskaya, T.V., Salomatova, E.S. (2016) 'Review of modern methods of management of the electron beam when the electron-beam welding', *Bull. PNRPU. Mechanical Eng., Materials Sci.*, 18(4), pp. 169–187 (in Russ.).
5. Dragunov, V.K., Sliva, A.P., Zhmurko, I.E. et al. (2016) '30KhGSA steel high-pressure vessels electron-beam welding technology', *Oil Industry J.*, (12), pp. 126–130 (in Russ.).
6. Burachevskiy, Yu.A., Bakeev, I.Yu., Miller, A.V., Novgorodov, R.G. (2017) 'System to control electron-beam setup and information display', *Proc. of TUSUR University*, (2), pp. 132–134 (in Russ.). doi: 10.21293/1818-0442-2017-20-2-132-134.
7. Popenko, Ya.N. (2016) 'Software development for vacuum equipment control of electron beam welding installation', *Proc. MNSC-2016 Instrumental Methods and Techniques of Experimental Phys.*, p. 35 (in Russ.).
8. Simonov, D.K., Simonova, M.E. (2019) 'Automation of welding production management as a factor in ensuring the reliability of welded joints (on the example of electron-beam welding)', *Proc. Int. Sci. Conf. Modern Tech. in the World Scientific Space*, (1), pp. 96–100 (in Russ.).
9. Madera, A.G., Kandalov, P.I. (2018) 'Principles of software construction for electronic system thermal design', *Software & Systems*, 31(3), pp. 435–438 (in Russ.). doi: 10.15827/0236-235X.123.435-438.
10. Parshukov, L.I., Efremov, N.A. (2019) 'Investigation structure and properties of seam weld from heat-resisting steel VZh159-ID', *Proc. of VIAM*, (3), pp. 27–34 (in Russ.).
11. Pavlushin, N.V. (2020) 'Zone electron-beam welding', *Proc. Electron-Beam Welding and Related Tech.*, pp. 288–302 (in Russ.).
12. Kharitonov, I.A., Dragunov, V.K., Goncharov, A.L., Martynov, V.N., Portnov, M.A. (2019) 'Development of a method of electron beam welding using inserts', *Proc. Welding in Russia-2019: Current State and Prospects*, p. 269 (in Russ.).
13. Darmaev, A.N., Muravyev, E.K., Sablin, V.M., Morev, S.P. (2020) 'Features of the formation of an electron flow in electron-optical systems with an auto-emission cathode for the installation of electron-beam welding', *Welding Production*, (6), pp. 20–23 (in Russ.).
14. Latyshenko, G.I. (2018) 'Development of an automated complex of electron beam equipment and technologies for welding thin-walled assemblies and parts', *Proc. Science Today: Problems and Solutions*, pp. 50–52 (in Russ.).
15. Shcherbakov, A.V., Gaponova, D.A., Rodyakina, R.V., Sliva, A.P. (2019) 'Electron-beam technology of additive shaping: Issues of control and reproducibility of technological modes', *Tech. Opponent*, 3(2), pp. 30–37 (in Russ.).

### Авторы

**Курашкин Сергей Олегович**<sup>1</sup>,

ассистент, scorpion\_ser@mail.ru

**Бухтояров Владимир Викторович**<sup>2</sup>, к.т.н.,

доцент, завкафедрой, vladber@list.ru

**Тынченко Вадим Сергеевич**<sup>1</sup>, к.т.н.,

доцент, vadimond@mail.ru

**Шуткина Елизавета Вячеславовна**<sup>1</sup>,

студент, elizaveta-shutkina@mail.ru

### Authors

**Sergey O. Kurashkin**<sup>1</sup>, Assistant,

scorpion\_ser@mail.ru

**Vladimir V. Bukhtoyarov**<sup>2</sup>, Cand. of Sci. (Engineering),

Associate Professor, Head of Chair, vladber@list.ru

**Vadim S. Tynchenko**<sup>1</sup>, Cand. of Sci. (Engineering),

Associate Professor, vadimond@mail.ru

**Elizaveta V. Shutkina**<sup>1</sup>, Student,

elizaveta-shutkina@mail.ru

<sup>1</sup> Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева, г. Красноярск, 660037, Россия

<sup>2</sup> Институт нефти и газа

Сибирского федерального университета,

г. Красноярск, 660041, Россия

<sup>1</sup> Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, 660037, Russian Federation

<sup>2</sup> Institute of Oil and Gas of the Siberian Federal University, Krasnoyarsk, 660041, Russian Federation

## Построение модели предиктивной аналитики неисправностей промышленного оборудования

А.В. Чернухин <sup>1</sup>✉, Е.А. Богданова <sup>2</sup>, Т.В. Савицкая <sup>1</sup>

<sup>1</sup> Российский химико-технологический университет имени Д.И. Менделеева,  
г. Москва, 125047, Россия

<sup>2</sup> Московский государственный университет имени М.В. Ломоносова,  
г. Москва, 119991, Россия

### Ссылка для цитирования

Чернухин А.В., Богданова Е.А., Савицкая Т.В. Построение модели предиктивной аналитики неисправностей промышленного оборудования // Программные продукты и системы. 2024. Т. 37. № 2. С. 254–261. doi: 10.15827/0236-235X.142.254-261

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 31.10.2023

После доработки: 11.11.2023

Принята к публикации: 26.12.2023

**Аннотация.** В статье описаны новые подходы к построению модели предиктивной аналитики неисправностей промышленного оборудования. Актуальность исследования обусловлена необходимостью заблаговременно идентифицировать неисправности, приводящие к снижению эффективности работы или простоям крупных производственных линий промышленного предприятия. В рамках исследования были изучены современные подходы, использующие машинное обучение для решения схожих задач, подробно описаны их достоинства и недостатки. Данные для обучения моделей включают сведения о показаниях датчиков, измеряющих токи, температуры, давления и вибрации, для шести эксгаустеров за три года. Для обучения и тестирования были отобраны только техмesta с наличием неисправностей за анализируемый период. В связи со сложной взаимосвязью между показаниями датчиков и состоянием эксгаустера было принято решение о создании предсказательной модели, основанной на мягком голосовании между тремя алгоритмами, имеющими различные подходы к классификации: сверточная нейронная сеть, логистическая регрессия и метод опорных векторов. Предложенный подход реализован с помощью высокоуровневого языка программирования Python с использованием дистрибутива Anaconda. Описан порядок выполнения исследования, представлены блок-схемы разработанных алгоритмов и их подробное описание. Проведено всестороннее тестирование разработанного программного обеспечения, реализующего модель прогнозирования неисправностей. Результаты исследования подтверждают работоспособность предлагаемой модели прогнозирования неисправности технических узлов оборудования, которая может быть адаптирована для предиктивной аналитики неисправностей большого числа типов промышленного оборудования крупных предприятий металлургической, химической и других отраслей промышленности.

**Ключевые слова:** математическое моделирование, интеллектуальный анализ данных, предиктивная аналитика, машинное обучение, нейронные сети, промышленное оборудование, прогнозирование неисправностей

**Введение.** Проблема возникновения неисправностей промышленного оборудования, приводящих к снижению эффективности работы или к простоям крупных производственных линий, а следовательно, к убыткам, актуальна во всем мире [1, 2]. Вместе с тем активно продолжается автоматизация всех отраслей промышленного производства, которая обуславливает развитие информационно-технических систем предприятия [3], а значит, систем сбора, хранения и мониторинга состояния оборудования в режиме реального времени. При этом эксперты в области промышленных информационных систем сходятся во мнении, что аккумулированные большие объемы данных могут использоваться для прогнозирования неисправностей задолго до их возникновения [4, 5]. Для этого необходимо построение отдельной подсистемы предиктивной аналитики, исполь-

зуемой для гибкого планирования технического обслуживания и ремонта [6].

Методы прогнозирования на основе классического машинного обучения широко используются в области диагностики и прогнозирования неисправностей оборудования благодаря хорошей интерпретируемости результатов обучения, относительно невысоким требованиям к вычислительным ресурсам и широкому выбору подходов, которые обеспечивают разносторонний анализ данных [7, 8]. Хотя классические алгоритмы машинного обучения хорошо показывают себя в диагностике и прогнозировании неисправностей, они уступают нейросетевым алгоритмам в идентификации сложных взаимосвязей между показаниями большого числа датчиков и состоянием оборудования [9, 10]. Было проведено большое количество исследований, направленных на разработку алгорит-

мов на основе нейронных сетей [11], но в подавляющем большинстве случаев диагностика в них осуществляется только для одного технического места (подшипник, ротор и др.), что говорит о сложности или невозможности масштабирования таких подходов [12, 13]. В подобных работах для обучения и тестирования используются наборы данных с большим количеством примеров неисправностей, за счет чего достигается высокая точность предсказания (ассигасу = 0.8–0.9). Однако возникновение неисправности – достаточно редкое явление в работе промышленного оборудования, в связи с чем часто отсутствует возможность накопления обширной библиотеки отказов [14]. Это является основной трудностью в диагностике неисправной работы в условиях реального производства с помощью нейронных сетей, так как они в таких случаях проявляют склонность к переобучению [15, 16].

Соответственно, в системе предиктивной аналитики, работающей в режиме реального времени, необходимо использовать как методы классического машинного обучения, так и нейросетевые алгоритмы, за счет чего будут нивелированы их недостатки, связанные с переобучением в условиях небольшого количества данных о неисправностях [17, 18]. Для повышения точности предсказания между моделями в ансамбле будет проводиться так называемое мягкое голосование. Модель необходимо обучать на данных с долей неисправной работы менее 15 % по каждому техническому месту, а раннее прогнозирование возникновения неисправностей для всех ключевых технических мест следует осуществлять одновременно.

Целью данного исследования является разработка комплексной модели, позволяющей прогнозировать неисправности промышленного оборудования с учетом специфики реальных промышленных данных. Ее отличительной особенностью является использование ансам-

блевого подхода к предсказанию неисправностей, основанного на проведении операции голосования между методами классического машинного обучения и нейросетевым алгоритмом, что позволяет максимально разносторонне рассмотреть признаковое пространство и выявить наиболее важные свойства.

Разработанное ПО позволит на основе текущих данных о состоянии оборудования, получаемых с датчиков, заблаговременно идентифицировать зарождающиеся неисправности.

Для обучения моделей использованы данные, полученные от ПАО «Северсталь», в качестве оборудования выступают эксгаустеры агломерационной машины – центробежные нагнетатели, осуществляющие просос воздуха через слой шихты.

### Предобработка данных

В наборе обучающих данных присутствуют сведения о показаниях датчиков для шести эксгаустеров за три года (с 01.01.2019 по 31.12.2021) в десятисекундных интервалах. Формат данных представлен в таблице 1.

Датчики измеряют такие показания состояния оборудования, как токи (амперы), температуры (градусы Цельсия), давления (КПа), вибрации (мм/с). Всего проводился анализ показаний 16 датчиков.

Информация о неисправностях технических мест эксгаустеров (Т.М.  $i$ ) была предоставлена в виде таблицы с бинарным кодированием (0 – штатное состояние технического места, 1 – неисправная работа) (табл. 2).

Количество и набор технических мест отличались для каждого из эксгаустеров (от 20 до 35). Для обучения и тестирования были отобраны техместа с наличием неисправностей за анализируемый период – маслоохладитель, подшипник опорный, вибропреобразователь, электроаппаратура, электродвигатель, улита, корпус, ротор.

Таблица 1

Формат данных о показаниях датчиков за анализируемый период

Table 1

Format of data on sensor readings over the analyzed period

Дата	Датчик 1	Датчик 2	...	Датчик 15	Датчик 16
01.01.2019 00:00:00	123	0	...	345	17
01.01.2019 00:00:10	135	0.57	...	320	17
01.01.2019 00:00:20	127	0.93	...	387	19
...	...	...	...	...	...
31.12.2021 23:59:40	69	2.4	...	341	0
31.12.2021 23:59:50	63	1.2	...	390	3

Таблица 2

**Формат данных, содержащих метки неисправностей за анализируемый период**

Table 2

**Format of data containing fault labels over the analyzed period**

Дата	Т.М. 1	Т.М. 2	...	Т.М. 20	Т.М. 21
01.01.2019 00:00:00	0	0	...	1	0
01.01.2019 00:00:10	0	0	...	1	0
01.01.2019 00:00:20	0	0	...	1	0
...	...	...	...	...	...
31.12.2021 23:59:40	1	0	...	0	0
31.12.2021 23:59:50	1	0	...	0	0

Предобработка показаний датчиков включала в себя следующие этапы.

- *Удаление выбросов.* Производились обнуление отрицательных показаний датчиков и уменьшение амплитуды значений, превышающих среднее более чем в сто раз (так как подобные значения не являются нормальными показаниями датчиков и негативно сказываются на обучении).

- *Приведение данных к часовым интервалам из десятисекундных.* Показания с шагом в 10 секунд являются избыточными и неоптимальными для обучения предсказательного алгоритма в связи с зашумленностью данных и большим объемом занимаемой памяти (необходимы хранение и работа с таблицей, содержащей 9 460 800 строк для каждого эксгаустера). Для дальнейшей работы для каждого из часовых интервалов (360 строк) рассчитывались статистические характеристики, такие как среднее, медиана, максимальные и минимальные значения. В дальнейшем именно эти значения использованы в качестве признаков при обучении модели.

- *Стандартизация и замена пропущенных значений.* В связи с различной природой показаний датчиков производилась их стандартизация. Для этого рассчитывались обобщенное для всех эксгаустеров среднее значение и стандартное отклонение для каждого датчика, которые и использовались для стандартизации. Оставшиеся пропущенные значения заполнялись путем интерполяции.

- *Добавление новых признаков.* Для получения информации об отклонении показаний датчиков за текущий час от показаний за предыдущие сутки рассчитывалась разность между средними значениями за эти периоды. Полученные изменения в дальнейшем также использовались в качестве признаков, поскольку позволяют не только учитывать информацию о локальных изменениях состояния

системы, но и выявлять отклонения от состояния за более долгий промежуток времени.

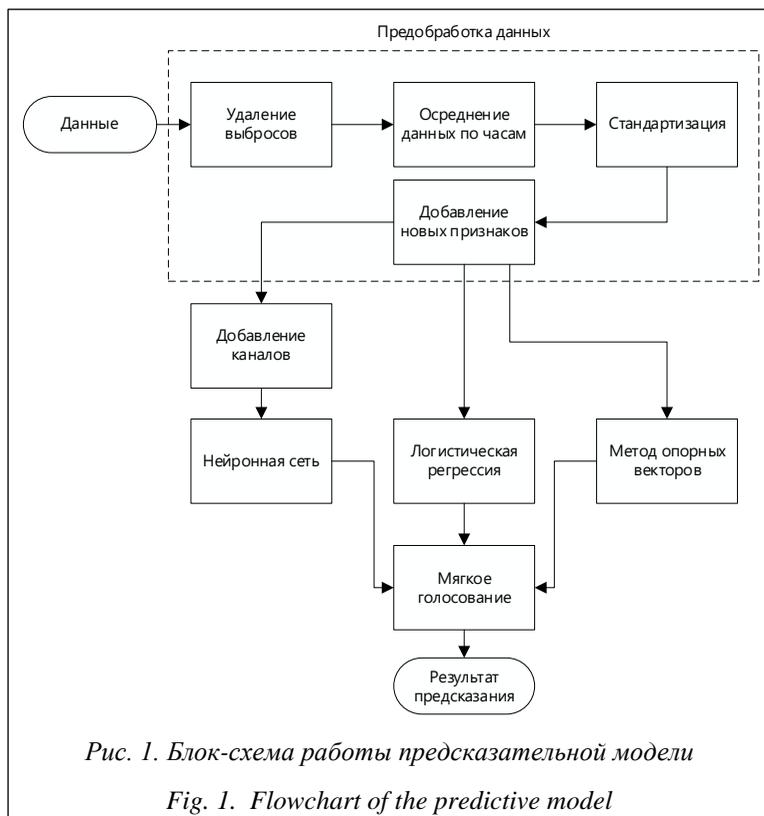
Весь набор данных состоит из 26 280 векторов, характеризующих часовые интервалы для шести эксгаустеров. Для тренировочного датасета отбирались интервалы для каждого эксгаустера, содержащие неисправности, при их отсутствии – фрагмент со штатным состоянием. Для валидации отбирались интервалы, следующие после отобранных для обучения и содержащие хотя бы одну продолжительную (более 3 суток) неисправность из всех эксгаустеров. В качестве тестового набора отбирались самые поздние интервалы, содержащие неисправности. В связи с тем, что периоды неисправного состояния оборудования для разных технических мест не совпадали, для каждого из них разделение происходило индивидуально, с учетом причин, повлекших за собой неисправность.

### Архитектура предсказательной модели

В связи со сложной взаимосвязью между показаниями датчиков и состоянием эксгаустера было принято решение о создании предсказательной модели, основанной на мягком голосовании между тремя алгоритмами, имеющими различные подходы к классификации (рис. 1). Суть данного ансамблевого метода заключается в подборе к каждому из предсказательных алгоритмов весов, отражающих значимость прогноза. Подбор весов осуществлялся с использованием метода градиентного бустинга. По итогу для каждого объекта рассчитывалась вероятность принадлежности к классу по следующей формуле:

$$P_1 = w_a p_{a1} + w_b p_{b1} + w_c p_{c1},$$

где  $P_1$  – итоговая вероятность принадлежности объекта к классу 1 (неисправность);  $p_{a1}$ ,  $p_{b1}$ ,  $p_{c1}$  – вероятности принадлежности объекта к классу 1, предсказанные тремя отдельными алгорит-



мами;  $w_a$ ,  $w_b$ ,  $w_c$  – веса, присваиваемые прогнозам отдельных алгоритмов.

Для голосования использованы сверточная нейронная сеть, логистическая регрессия и метод опорных векторов. Выбор обусловлен низкой корреляцией их предсказаний друг с другом, составляющей менее 0.6, с относительно высоким качеством предсказания для большей части анализируемых технических мест (значение индекса Жаккара выше 0.5).

Для логистической регрессии выбран базовый алгоритм оптимизации (LBFGS), максимальное количество итераций – 2 000 (что обусловлено большим количеством признаков). Метод опорных векторов обучался с использованием ядра rbf (Radial Basis Function) в связи с нелинейной разделимостью объектов в анализируемом наборе данных. Данные для моделей подавались на вход в виде векторов размера 96 (по 6 признаков для каждого из 16 датчиков), ответом являлась метка 0 или 1 в случае неисправности в течение часа, описываемого вектором признаков.

Было опробовано несколько архитектур с разным количеством сверточных и полносвязных слоев. Итоговая нейронная сеть (рис. 2) содержит базовые слои, в которых происходит подбор весов.

- *Сверточный слой (Conv1D)*. Выполняет функции обобщения признаков в каналах и

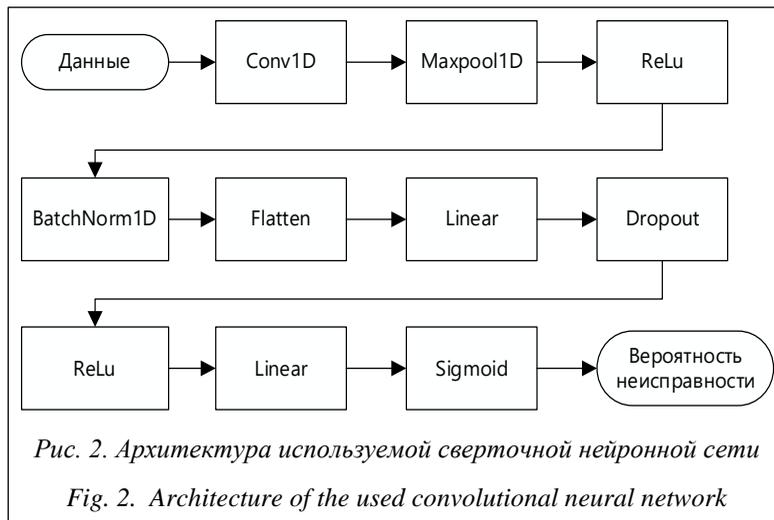
выявления значимых паттернов с увеличением числа каналов с 6 до 32. В качестве каналов используются статистические характеристики, рассчитанные для показателей каждого из 16 датчиков, в частности, среднее, медиана, максимальное и минимальное значения за час, а также разность между средним и медианой за час и за месяц. В связи с размером входного тензора выбран размер ядра свертки, равный 3. После сверточного слоя располагается слой подвыборки по максимальному значению (MaxPooling), который уменьшает в два раза размер тензора. В качестве функции активации использовалась выпрямленная линейная функция активации (ReLU).

- *Два полносвязных слоя (Linear)* с уменьшением количества каналов (100, 1). Для первого слоя использовалась функция активации ReLu. В последнем слое непосредственно происходит предсказание наличия неисправности и используется сигмоидная функция активации (Sigmoid), которая выводит вероятность принадлежности объекта к классу 1 (наличие неисправности).

Таким образом, на вход алгоритм принимает 4D-тензор размера  $6 \times 16$ , на выходе получается одно число, отражающее вероятность наличия неисправности за час, который описывается входным тензором.

В целях нормализации и регуляризации после сверточных слоев использовалась пакетная нормализация (Batch Normalization), а после первых двух полносвязных слоев применялся метод прореживания (Dropout) со значением отсева 0.3 (из-за наличия пакетной нормализации значение выбрано меньше стандартного 0.5).

В качестве оптимизатора был выбран оптимизатор Адама. Выбор обусловлен его эффективностью и возможностью добавления L2 регуляризации (параметр `weight_decay`), которая необходима для ограничения значения весов для признаков, что препятствует переобучению и способствует поиску общих закономерностей. В результате подбора гиперпараметров (настраиваемых вручную параметров обучения) использована скорость обучения, равная 0.0001, и `weight_decay` = 0.001. Для расчета



ошибки применялась функция потерь BCELoss, оптимальная для решения задачи бинарной классификации. При оптимизации также подбирался размер пакета (batch size), оказывающий влияние на эффективность Batch Normalization и на продуктивность обучения. Исходя из размера обучающей выборки выбран batch size = 16. Обучение проводилось в течение 30 эпох, сохранялась модель с наименьшим значением функции потерь на валидационном наборе данных.

Метрикой качества выступает индекс Жаккара, рассчитанный по формуле

$$J = \frac{TP}{TP + FP + FN},$$

где  $J$  – индекс Жаккара;  $TP$  – истинно положительные значения из матрицы путаницы;  $FP$  – ложно положительные;  $FN$  – ложно отрицательные. Выбор данной метрики обусловлен более объективной работой с наборами данных, содержащими низкую долю одного из классов.

Программное обеспечение разработано на высокоуровневом языке программирования Python с использованием дистрибутива Anaconda, включающего набор свободных библиотек, используемых при работе с данными и в машинном обучении.

### Описание и сравнение полученных результатов

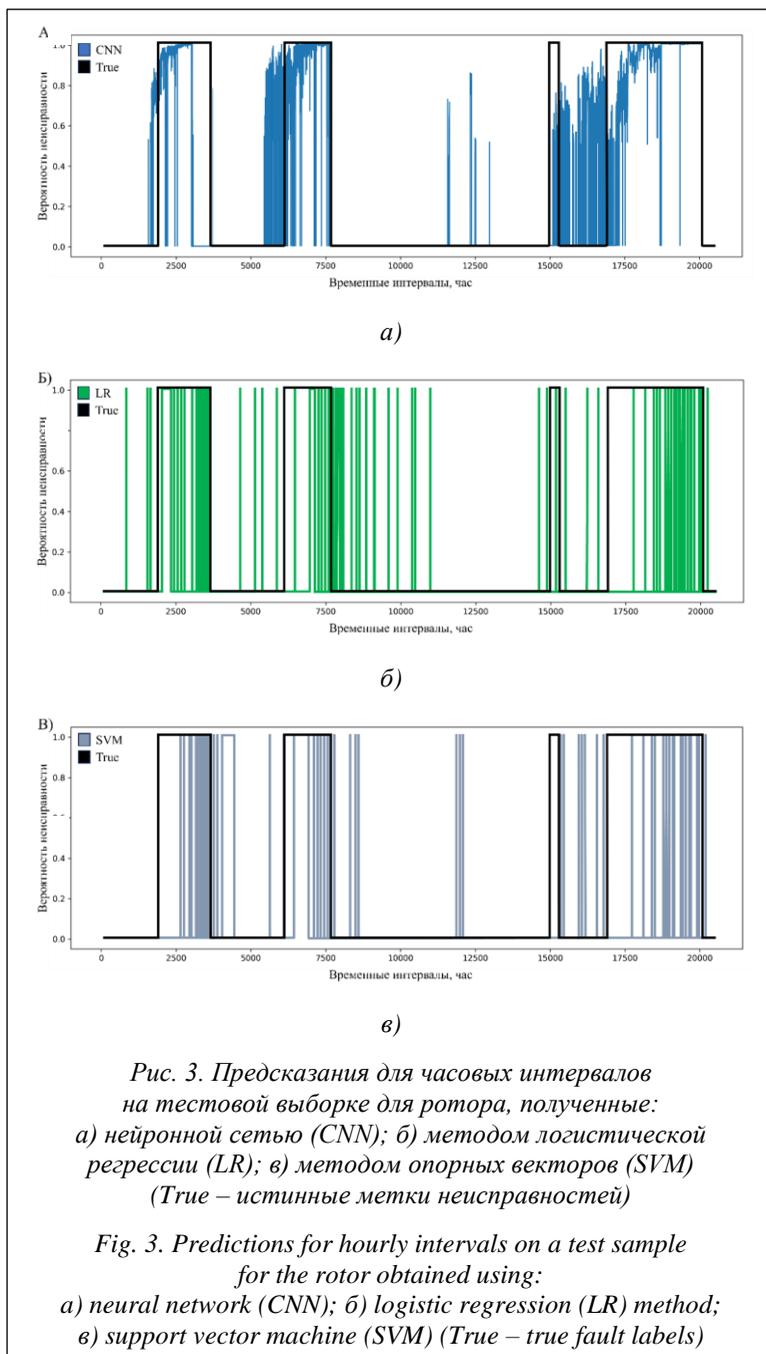
На первом этапе работы были обучены предсказательные модели отдельно для каждого технического места. В результате обучения нейронной сети получены вероятности неисправности в каждый момент времени. Для идентификации неисправностей в разных

техместах индивидуально подбирались пороги принятия решения. На рисунке 3а представлено распределение вероятностей наличия неисправности на тестовой выборке для ротора (в качестве порога взято значение 0.5), при этом индекс Жаккара равен 0.6 (<http://www.swsys.ru/uploaded/image/2024-2/5.jpg>). В тестовых временных интервалах для трех эксгаустеров, объединенных в общий тестовый набор, наблюдаются три продолжительные неисправности и одна более короткая (рис. 3а). Можно заметить,

что в трех случаях модель начинала присваивать высокую вероятность развития неисправности заранее, до официального начала, обозначенного оператором. Это, с одной стороны, отрицательно повлияло на значение метрики качества, а с другой, является положительным моментом, позволяя диагностировать начало развития неисправности до того, как это явным образом скажется на работе эксгаустера и будет замечено оператором. Однако одна из неисправностей не была диагностирована, что может быть связано с причиной, слабо представленной в обучающем наборе данных.

Можно заметить, что результаты, полученные другими моделями, достаточно сильно отличаются и относят к неисправностям периоды их активного развития (рис. 3б, 3в), а также не идентифицируют непродолжительную неисправность, что подтверждает упомянутое предположение. При этом значение метрики качества для логистической регрессии равно 0.51, для метода опорных векторов – 0.4 (<http://www.swsys.ru/uploaded/image/2024-2/5.jpg>). Полученные данные об особенностях прогнозирования разных этапов развития неисправности ротора свидетельствуют о возможности улучшения качества предсказания при использовании ансамблевого подхода, объединяющего ранее названные модели. Метрики качества для используемых подходов, включая ансамблевые, отображены на рисунке <http://www.swsys.ru/uploaded/image/2024-2/5.jpg>.

Благодаря использованию ансамблевого подхода, включающего операцию мягкого голосования между моделями, для ряда технических мест удалось значительно улучшить качество предсказания на тестовой выборке. Наилучшие показатели наблюдаются для кор-



пуса, улита, электроаппаратуры и вибропреобразователя. Наиболее высокое качество предсказания получили технические места, неис-

оборудования, решение может успешно масштабироваться в комплексную систему предиктивной аналитики предприятия.

**Список литературы**

1. James A.T., Gandhi O., Deshmukh S. Fault diagnosis of automobile systems using fault tree based on digraph modeling. *Int. J. Syst. Assur. Eng. Manag.*, 2018, vol. 9, pp. 494–508. doi: 10.1007/s13198-017-0693-6.
2. Bhakta K., Sikder N., Nahid A.A., Islam M.M.M. Fault diagnosis of induction motor bearing using Cepstrum-based preprocessing and ensemble learning algorithm. *Proc. Int. Conf. ECCE*, 2019, pp. 1–6. doi: 10.1109/ECACE.2019.8679223.
3. Трофимов В.В. Алгоритмизация и программирование. М.: Юрайт, 2017. 137 с.
4. Оклей П.И. Прогнозирование остаточного ресурса и вероятности наступления отказа оборудования – основа проектирования производственной программы ремонтных работ тепловой электростанции // *Контроллинг*. 2017. № 65. С. 54–63.

правная работа которых имеет быстрое развитие (электроаппаратура, вибропреобразователь, электродвигатель), а также с наиболее заметными и легко идентифицируемыми проявлениями неисправной работы (улита, корпус, вибропреобразователь). Следовательно, разметка данных для этих точек может производиться в более оперативном режиме, что важно для обучения моделей с учителем. Однако для остальных техузлов, несмотря на более низкие значения метрик, обученные модели предсказывают зарождение неисправности до момента идентификации ее сотрудниками, что позволит заранее подготовиться к развитию аномалии в работе оборудования и принять необходимые превентивные меры.

**Заключение**

Таким образом, с использованием разработанных моделей можно проводить исследование работы эксгаустеров в режиме реального времени, а их результаты применять для планирования работ по оптимальному техническому обслуживанию.

В текущем виде разработанное ПО может использоваться в пилотном проекте на одном из промышленных предприятий и покажет свою эффективность при прогнозировании неисправностей техузлов эксгаустера. В дальнейшем, в случае получения данных по другим типам

5. Nan X., Zhang B., Liu C., Gui Z., Yin X. Multi-modal learning-based equipment fault prediction in the internet of things. *Sensors*, 2022, vol. 22, no. 18, art. 6722. doi: 10.3390/s22186722.
6. Lv H., Chen J., Pan T., Zhang T., Feng Y., Liu Sh. Attention mechanism in intelligent fault diagnosis of machinery: A review of technique and application. *Measurement*, 2022, vol. 199, art. 111594. doi: 10.1016/j.measurement.2022.111594.
7. Боровский А.С., Шумилина Н.А. Нечеткая ситуационная сеть для оценки проектного риска отказа оборудования // Тр. ИСА РАН. 2018. Т. 68. № 1. С. 87–93.
8. Xu Z., Mei X., Wang X., Yue M., Jin J., Yang Y., Li Ch. Fault diagnosis of wind turbine bearing using a multi-scale convolutional neural network with bidirectional long short term memory and weighted majority voting for multi-sensors. *Renewable Energy*, 2022, vol. 182, pp. 615–626. doi: 10.1016/J.RENENE.2021.10.024.
9. Антонов А.В. Методика статистического анализа данных об отказах оборудования АЭС в условиях неоднородного потока событий // Изв. вузов. Ядерная энергетика. 2016. № 3. С. 20–29. doi: 10.26583/npe.2016.3.03.
10. Cheng Y., Yuan H., Liu H., Lu C. Fault diagnosis for rolling bearing based on SIFT-KPCA and SVM. *Eng. Computations*, 2017, vol. 34, no. 1, pp. 53–65. doi: 10.1108/EC-01-2016-0005.
11. Mao W., Feng W., Liu Y., Zhang D., Liang X. A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. *MSSP*, 2021, vol. 150, art. 107233. doi: 10.1016/j.ymsp.2020.107233.
12. Иванов В.К., Виноградова Н.В., Палюх Б.В., Сотников А.Н. Современные направления развития и области приложения теории Демпстера–Шафера // ИИИПР. 2018. № 4. С. 32–42. doi: 10.14357/20718594180403.
13. Javed K., Gouriveau R., Zerhouni N. State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels. *MSSP*, 2017, vol. 94, no. 9, pp. 214–236. doi: 10.1016/j.ymsp.2017.01.050.
14. Hossain M., Abu-Siada A., Muyeen S. Methods for advanced wind turbine condition monitoring and early diagnosis: A literature review. *Energies*, 2018, vol. 11, no. 5, art. 1309. doi: 10.3390/en11051309.
15. Xiao C., Liu Z., Zhang T., Zhang X. Deep learning method for fault detection of wind turbine converter. *Applied Sci.*, 2021, vol. 11, no. 3, art. 1280. doi: 10.3390/app11031280.
16. Пальчевский Е.В., Антонов В.В., Еникеев Р.Р. Прогнозирование на основе искусственной нейронной сети второго поколения для поддержки принятия решений в особо значимых ситуациях // Программные продукты и системы. 2022. Т. 35. № 3. С. 384–395. doi: 10.15827/0236-235X.139.384-395.
17. Kumar A., Kumar R. Time-frequency analysis and support vector machine in automatic detection of defect from vibration signal of centrifugal pump. *Measurement*, 2017, vol. 108, pp. 119–133. doi: 10.1016/j.measurement.2017.04.041.
18. Abbasi A.R., Mahmoudi M.R., Avazzadeh Z. Diagnosis and clustering of power transformer winding fault types by cross-correlation and clustering analysis of FRA results. *IET Generation, Transmission & Distribution*, 2018, vol. 12, no. 19, pp. 4301–4309. doi: 10.1049/iet-gtd.2018.5812.

## Developing predictive analytics models for industrial equipment failures

Artem V. Chernukhin <sup>1</sup>✉, Elizaveta A. Bogdanova <sup>2</sup>, Tatiana V. Savitskaya <sup>1</sup>

<sup>1</sup> Mendeleev University of Chemical Technology of Russian, Moscow, 125047, Russian Federation

<sup>2</sup> Lomonosov Moscow State University, Moscow, 119991, Russian Federation

### For citation

Chernukhin, A.V., Bogdanova, E.A., Savitskaya, T.V. (2024) 'Development of predictive analytics models for industrial equipment failures', *Software & Systems*, 37(2), pp. 254–261 (in Russ.). doi: 10.15827/0236-235X.142.254-261

### Article info

Received: 31.10.2023

After revision: 11.11.2023

Accepted: 26.12.2023

**Abstract.** The paper describes new approaches to building a model for predictive analytics of industrial equipment failures. The relevance of the study is due to the need to identify failures that lead to a decrease in operating efficiency or downtime of large production lines at an industrial enterprise in advance. The study examines modern approaches, which use machine learning to solve similar problems, and describes their advantages and disadvantages in detail. Data for training models was obtained from PJSC Severstal and include information on sensor readings measuring currents, temperatures, pressure and vibrations for six exhausters over three years. Training and testing involved only technical locations with faults during the analyzed period. Due to the complex relationship between sensor readings and exhauster condition, it was decided to create a predictive model based on “soft voting” between three algorithms with different classification approaches: a convolutional neural network, logistic regression and a support vector machine. The proposed approach is based on the high-level Python programming language using the Anaconda distribution. The paper describes the procedure for performing the research, presents block diagrams of the developed algorithms and their detailed description. The authors carried out comprehensive testing of the developed software that implements the fault prediction model. The study results confirm the performance of the proposed model for predicting failures of equipment technical components, which can be adapted for predictive analytics of failures of a large number of types of industrial equipment at large enterprises in metallurgical, chemical and other industries.

**Keywords:** mathematical modeling, data mining, predictive analytics, machine learning, neural networks, industrial equipment, fault prediction

### References

1. James, A.T., Gandhi, O., Deshmukh, S. (2018) 'Fault diagnosis of automobile systems using fault tree based on digraph modeling', *Int. J. Syst. Assur.*, 9, pp. 494–508. doi: 10.1007/s13198-017-0693-6.
2. Bhakta, K., Sikder, N., Nahid, A.A., Islam, M.M.M. (2019) 'Fault diagnosis of induction motor bearing using Cepstrum-based preprocessing and ensemble learning algorithm', *Proc. Int. Conf. ECCE*, pp. 1–6. doi: 10.1109/ECACE.2019.8679223.
3. Trofimov, V.V. (2017) *Algorithmization and Programming*. Moscow, 137 p. (in Russ.).
4. Okley, P.I. (2017) 'Predicting the residual service life and probability of failure of equipment – base of design of production program of repair works of thermal power plant', *Controlling*, (65), pp. 54–56 (in Russ.).
5. Nan, X., Zhang, B., Liu, C., Gui, Z., Yin, X. (2022) 'Multi-modal learning-based equipment fault prediction in the internet of things', *Sensors*, 22(18), art. 6722. doi: 10.3390/s22186722.
6. Lv, H., Chen, J., Pan, T., Zhang, T., Feng, Y., Liu, Sh. (2022) 'Attention mechanism in intelligent fault diagnosis of machinery: A review of technique and application', *Measurement*, 199, art. 111594. doi: 10.1016/j.measurement.2022.111594.
7. Borovsky, A.S., Shumilina, N.A. (2018) 'Unclear situational network for assessment of the design risk of equipment failure', *Proc. ISA RAS*, 68, pp. 87–93 (in Russ.).
8. Xu, Z., Mei, X., Wang, X., Yue, M., Jin, J., Yang, Y., Li, Ch. (2022) 'Fault diagnosis of wind turbine bearing using a multi-scale convolutional neural network with bidirectional long short term memory and weighted majority voting for multi-sensors', *Renewable Energy*, 182, pp. 615–626. doi: 10.1016/J.RENENE.2021.10.024.
9. Antonov, A.V. (2016) 'Statistical analysis of the nuclear power plant equipment failure data in nonhomogeneous failure flow', *Izv. Vuzov. Yadernaya Energetika*, 3, pp. 20–29. doi: 10.26583/npe.2016.3.03 (in Russ.).
10. Cheng, Y., Yuan, H., Liu, H., Lu, C. (2017) 'Fault diagnosis for rolling bearing based on SIFT-KPCA and SVM', *Eng. Computations*, 34(1), pp. 53–65. doi: 10.1108/EC-01-2016-0005.
11. Mao, W., Feng, W., Liu, Y., Zhang, D., Liang, X. (2021) 'A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis', *MSSP*, 150, art. 107233. doi: 10.1016/j.ymsp.2020.107233.
12. Ivanov, V.K., Vinogradova, N.V., Palyukh, B.V., Sotnikov, A.N. (2018) 'Modern directions of development and areas of application of the Dempster-Shafer theory', *Artificial Intelligence and Decision Making*, (4), pp. 32–42 (in Russ.). doi: 10.14357/20718594180403.
13. Javed, K., Gouriveau, R., Zerhouni, N. (2017) 'State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels', *MSSP*, 94(9), pp. 214–236. doi: 10.1016/j.ymsp.2017.01.050.
14. Hossain, M., Abu-Siada, A., Muyeen, S. (2018) 'Methods for advanced wind turbine condition monitoring and early diagnosis: A literature review', *Energies*, 11(5), art. 1309. doi: 10.3390/en11051309.
15. Xiao, C., Liu, Z., Zhang, T., Zhang, X. (2021) 'Deep learning method for fault detection of wind turbine converter', *Applied Sci.*, 11(3), art. 1280. doi: 10.3390/app11031280.
16. Palchevsky, E.V., Antonov, V.V., Enikeev, R.R. (2022) 'Forecasting based on a second-generation artificial neural network to support decision-making in critical situations', *Software & Systems*, 35(3), pp. 384–395 (in Russ.). doi: 10.15827/0236-235X.139.384-395.
17. Kumar, A., Kumar, R. (2017) 'Time-frequency analysis and support vector machine in automatic detection of defect from vibration signal of centrifugal pump', *Measurement*, 108, pp. 119–133. doi: 10.1016/j.measurement.2017.04.041.
18. Abbasi, A.R., Mahmoudi, M.R., Avazzadeh, Z. (2018) 'Diagnosis and clustering of power transformer winding fault types by cross-correlation and clustering analysis of FRA results', *IET Generation, Transmission & Distribution*, 12(19), pp. 4301–4309. doi: 10.1049/iet-gtd.2018.5812.

### Авторы

**Чернухин Артем Валерьевич**<sup>1</sup>,  
аспирант, chernukhin.a.v@muctr.ru  
**Богданова Елизавета Александровна**<sup>1</sup>,  
аспирант, eabogdanova.bioinf@gmail.com  
**Савицкая Татьяна Вадимовна**<sup>1</sup>,  
д.т.н., профессор, savitskaia.t.v@muctr.ru

### Authors

**Artem V. Chernukhin**<sup>1</sup>, Postgraduate Student,  
chernukhin.a.v@muctr.ru  
**Elizaveta A. Bogdanova**<sup>2</sup>, Postgraduate Student,  
eabogdanova.bioinf@gmail.com  
**Tatiana V. Savitskaya**<sup>1</sup>, Dr.Sc. (Engineering),  
Professor, savitskaia.t.v@muctr.ru

<sup>1</sup> Российский химико-технологический университет имени Д.И. Менделеева, г. Москва, 125047, Россия

<sup>2</sup> Московский государственный университет имени М.В. Ломоносова, г. Москва, 119991, Россия

<sup>1</sup> Mendeleev University of Chemical Technology of Russian, Moscow, 125047, Russian Federation  
<sup>2</sup> Lomonosov Moscow State University, Moscow, 119991, Russian Federation

## Распознавание траекторий струй огнетушащего вещества из пожарного ствола на основе цифровых изображений

И.Н. Пожаркова<sup>1, 2</sup>✉

<sup>1</sup> Сибирская пожарно-спасательная академия ГПС МЧС России, г. Железнодорожск, 662972, Россия

<sup>2</sup> Сибирский федеральный университет, г. Красноярск, 660041, Россия

### Ссылка для цитирования

Пожаркова И.Н. Распознавание траекторий струй огнетушащего вещества из пожарного ствола на основе цифровых изображений // Программные продукты и системы. 2024. Т. 37. № 2. С. 262–269. doi: 10.15827/0236-235X.142.262-269

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 28.11.2023

После доработки: 11.03.2024

Принята к публикации: 18.03.2024

**Аннотация.** Статья посвящена разработке технологии высокоточного распознавания траекторий струй огнетушащего вещества из пожарного ствола на основе их цифровых изображений, полученных по результатам натурального эксперимента. Целью исследования является повышение эффективности сбора описывающих эти траектории эмпирических данных. Основу предлагаемой методики составляет выделение границ потока огнетушащего вещества с геометрической привязкой соответствующих точек. В работе представлены основные операции созданной технологии: коррекция геометрических искажений (в том числе перспективных), предварительная обработка изображений с выделением потока огнетушащего вещества, распознавание границ струи с учетом геометрической привязки, постобработка результатов распознавания. Кроме того, описан метод выделения первичного контура траектории на основе сечений в полярной системе координат. Для демонстрации эффективности предлагаемой технологии в статье приведены результаты использования ее программной реализации на базе разработанных модулей. Распознавались верхняя и нижняя границы струи воды из пожарного лафетного ствола на изображении, полученном в сложных с точки зрения выделения траектории условиях. Практическая значимость представленной методики заключается в сборе эмпирических данных в объеме, необходимом для решения различных задач, связанных, в частности, с построением и отладкой алгоритмов программного управления пожарными роботами. При этом технология позволяет использовать полученные цифровые изображения, которые могут иметь геометрические искажения и зашумление внешними объектами, характерными для струй значительных линейных размеров на открытых и закрытых пространствах. Научная новизна исследования заключается в разработке метода первичного выделения границ струи на основе сечений в полярных координатах, позволяющего повысить пространственное разрешение распознавания, а также технологии формирования существенного скачка яркостей на границах струи на основе адаптации для решения соответствующей задачи и последовательного применения существующих общих методик обработки цифровых изображений.

**Ключевые слова:** распознавание образов, цифровое изображение, сечение в полярных координатах, машинное обучение, траектория струи, пожарные стволы

**Введение.** Определение различных параметров свободных струй является основой решения множества задач [1]. При этом оценка соответствующих характеристик, полученных в результате наблюдений или натуральных испытаний, в достаточно широком диапазоне значений факторов, оказывающих на них влияние, может составлять довольно сложную проблему. Так, при исследовании траекторий движения огнетушащего вещества из пожарного ствола масштаб наблюдаемых явлений измеряется десятками метров (в частности, дальность струи из лафетного ствола ЛС-С60У превышает 70 метров [2]). Это значительно затрудняет определение исследуемых параметров пожарной ствольной техники с приемлемым относительно решаемых на их основе задач пространственным разрешением. Однако получение соответствующих эмпирических данных

в достаточном объеме является необходимым условием для настройки, валидации (экспериментальной оценки адекватности) моделей прогнозирования движения огнетушащего вещества, построения и отладки алгоритмов наведения пожарных роботов, систем технического зрения, формирования обучающей и тестовой выборки, полнота которых во многом влияет на качество машинного обучения разрабатываемых моделей и алгоритмов [3, 4]. Одним из способов решения указанной проблемы является распознавание [5, 6], то есть определение целевых характеристик исследуемых потоков жидкости на основе цифровых изображений (фото и видео), получаемых в результате наблюдения или эксперимента, с учетом геометрических искажений, зашумления внешними объектами, распыления огнетушащего вещества при движении, характерном для струй

значительных линейных размеров на открытых и закрытых пространствах.

Существенная часть исследований в данной области посвящена преимущественно распознаванию струй из стволов пожарных роботов в соответствующих системах технического зрения. Так, в [7] представлен метод идентификации траектории движения огнетушащего вещества на коротком начальном участке, при этом не учитываются геометрические искажения, а также раздробление и распад потока на значительном удалении от камеры. Алгоритм распознавания струй значительных линейных размеров, описанный в [8], ориентирован в первую очередь на оценку скорости отдельных участков струи на основе анализа соответствующих видеозаписей. При этом при предварительной обработке цифровых изображений не учитываются возможное зашумление внешними объектами, а также отклонение фрагментов огнетушащего вещества, распыленного в результате движения, от основного потока. Указанные недостатки рассмотренных методов не позволяют использовать их для оценки с вы-

соким пространственным разрешением параметров свободных струй, имеющих большую (несколько десятков метров) дальность. Данная статья посвящена разработке технологии высокоточного распознавания траекторий огнетушащего вещества на основе их цифровых изображений с учетом негативного влияния факторов, характерных для соответствующих натуральных испытаний: геометрические искажения, наложение на изображение струи внешних объектов, распад струи в результате движения и другие.

### Описание метода

Функциональная диаграмма предлагаемой в данной статье технологии представлена на рисунке 1.

В качестве траектории струи огнетушащего вещества в рамках данного исследования используется множество лежащих на ее верхней (крайние капли) и нижней границах точек пространства в установившемся режиме, когда значения скоростей в этих точках не испытывают значительных изменений. Основу предла-

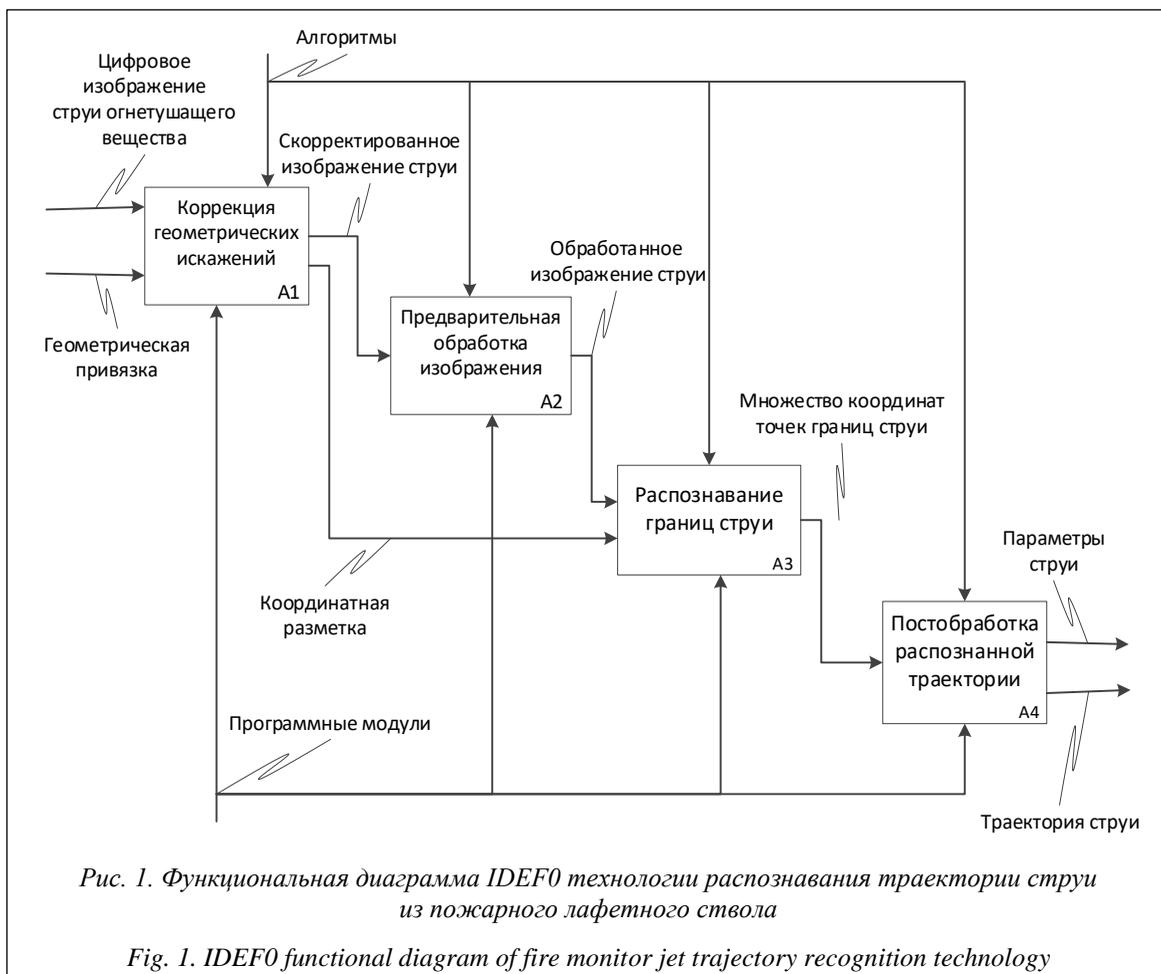


Рис. 1. Функциональная диаграмма IDEF0 технологии распознавания траектории струи из пожарного лафетного ствола

Fig. 1. IDEF0 functional diagram of fire monitor jet trajectory recognition technology

гаемой технологии составляют четыре этапа: коррекция геометрических искажений [9], предварительная обработка изображения струи, распознавание границ струи, постобработка распознанной траектории струи.

Коррекция геометрических искажений состоит из трех последовательно применяемых преобразований изображения.

– Устранение дисторсии (искривления прямых линий оптической системой объектива фотокамеры). Осуществляется на основе тестового снимка равномерной прямоугольной сетки, в соответствии с которым производится коррекция изображения [9]. Если объектив камеры не имеет выраженной дисторсии, данная операция пропускается.

– Поворот изображения в случае, когда положение матрицы в момент съемки отклоняется от горизонтали. Осуществляется путем вращения пикселей фотоснимка на соответствующий угол относительно заданной точки. Если камера выставлена по уровню, данная операция пропускается.

– Коррекция перспективных искажений, которые проявляются изменением размеров фрагментов изображения, в данном случае участков струи, в зависимости от расстояния до них. Если матрица фотокамеры и струя лежат в параллельных плоскостях, а съемка осуществляется с достаточно большого расстояния, то перспективные искажения отсутствуют или незначительны и, соответственно, данная операция пропускается.

Предварительная обработка изображения струи состоит из трех последовательно применяемых преобразований:

– первичная коррекция яркостей [10], заключающаяся в затемнении областей изображения, максимально близких к абсолютно белым – (255, 255, 255), в цветовой модели RGB для предотвращения потери деталей в них при последующих преобразованиях;

– коррекция локального контраста [11], заключающаяся в его повышении в отдельных диапазонах яркостей: высоких, средних и при необходимости низких;

– коррекция глобального контраста [11], заключающаяся в значительном снижении яркости (до абсолютно черного цвета – (0, 0, 0) в цветовой модели RGB) темных областей изображения и в повышении (до абсолютно белого цвета) светлых областей.

Основная цель данных преобразований – выделение контуров струи за счет формирования существенного скачка яркостей на ее гра-

ницах. При проведении натуральных испытаний в темное время суток можно использовать прожекторы [12] для обеспечения соответствующего перепада между неосвещенным окружающим пространством и рассеиваемой огнетушащим веществом подсветкой. В сложных с точки зрения выделения струи случаях возможно итеративное повторение перечисленных выше операций. Стоит отметить, что используемые при этом преобразования [9–11] являются общими методами обработки цифровых изображений, не связанными напрямую с задачами распознавания. Значения каждого из настроечных параметров этих операций для коррекции снимков огнетушащего вещества в рамках данного исследования определяются в соответствии с условиями, в которых они были получены при проведении экспериментов: освещение, модель камеры и объектива, степень распыления струи, зашумление внешними объектами и т.д. При этом, если в рамках серии натуральных испытаний указанные факторы существенно не изменяются, адаптация алгоритмов производится однократно и используется для обработки соответствующих результатов в пакетном режиме. Для автоматической подстройки параметров каждого из этапов коррекции под особенности обрабатываемых изображений используется машинное обучение [13].

Распознавание верхней и нижней границ струи в рамках предлагаемой технологии осуществляется в три этапа:

– выделение изображения огнетушащего вещества в виде множества точек на основе соответствующего диапазона цветов, а также перепада яркостей на границе потока жидкости;

– масштабирование координат полученного множества точек с учетом геометрической привязки исходного изображения; коэффициент масштабирования определяется на основе известного линейного размера ориентира (например, расстояние от пожарного ствола до определенной точки в плоскости движения струи), который присутствует на соответствующих цифровых снимках;

– выделение первичных контуров границы струи; производится по крайним точкам в различных сечениях потока огнетушащего вещества, при этом выбор метода формирования сечений, полностью перекрывающих траекторию движения жидкости, во многом определяет качество полученного результата.

Так, наиболее простым способом выделения первичных контуров верхней и нижней

границ на основе множества точек струи потока огнетушащего вещества, двигающегося в одной плоскости, является перебор по прямоугольной координатной сетке: ось  $OX$  – горизонтальная, ось  $OY$  – вертикальная. Для каждого значения абсциссы в диапазоне дальности струи определяется ее точка с максимальной и минимальной ординатами. Множества точек, полученных перебором таких сечений вдоль горизонтальной оси, представляют собой контуры верхней и нижней границ. Однако данный способ имеет существенный недостаток: на излете поток огнетушащего вещества имеет низкую горизонтальную скорость, за счет чего траектория его движения в этой области, особенно при больших углах возвышения выходного патрубка лафетного ствола, близка к вертикальной. В данном случае пространственное разрешение метода выделения контуров границ с использованием прямоугольной сетки будет невысоким. Кроме того, этот подход чувствителен к существенным отклонениям фрагментов огнетушащего вещества, распыленного в результате движения, от основного потока, формирующего траекторию струи. Поэтому в рамках предлагаемой технологии был разработан метод формирования сечений в полярной системе координат (рис. 2).

За начало координат в полярной системе принимается центральная точка проекции струи на горизонтальную ось:

$$X_0 = (X_{max} + X_{min})/2,$$

$$Y_0 = Y(X_{min}),$$

где  $X_0, Y_0$  – координаты центральной точки проекции струи на горизонтальную ось в пря-

моугольной системе, м;  $X_{max}, X_{min}$  – максимальная и минимальная абсциссы множества точек струи, м;  $Y(X_{min})$  – ордината точки струи с минимальной абсциссой, м.

Сечения струи в полярной системе координат формируются поворотом луча из точки  $(X_0, Y_0)$  на угол  $\varphi$  относительно горизонтального положения (рис. 2). Для каждого такого сечения определяются точки струи, находящиеся на максимальном и минимальном расстояниях ( $r_{max}$  и  $r_{min}$  соответственно) от  $(X_0, Y_0)$ . Координаты этих точек в декартовой системе определяются по следующим формулам:

$$X_{pmax} = X_0 + r_{max}\cos(\varphi),$$

$$Y_{pmax} = Y_0 + r_{max}\sin(\varphi),$$

$$X_{pmin} = X_0 + r_{min}\cos(\varphi),$$

$$Y_{pmin} = Y_0 + r_{min}\sin(\varphi),$$

где  $r_{max}, r_{min}$  – расстояния от точки  $(X_0, Y_0)$  до точек струи, находящихся от нее, соответственно, на максимальном и минимальном расстояниях и лежащих на секущем луче, м;  $\varphi$  – угол поворота секущего луча, рад.

Угол  $\varphi$  при этом меняется от  $\varphi_1$  до  $\varphi_2$  (по часовой стрелке):

$$\varphi_1 = \arccos\left(\frac{X_{min} - X_0}{\sqrt{(X_{min} - X_0)^2 + (Y(X_{min}) - Y_0)^2}}\right) =$$

$$= \{Y(X_{min}) = Y_0\} = \arccos(-1) = \pi,$$

$$\varphi_2 = \arccos\left(\frac{X_{max} - X_0}{\sqrt{(X_{max} - X_0)^2 + (Y(X_{max}) - Y_0)^2}}\right),$$

где  $Y(X_{max})$  – ордината точки струи с максимальной абсциссой, м.

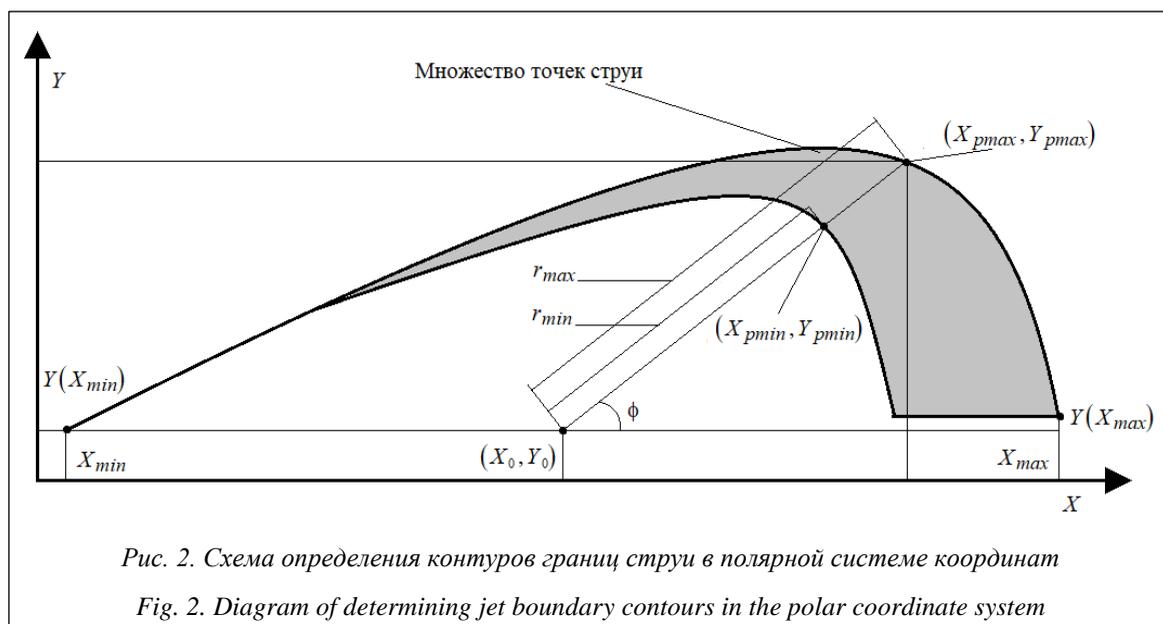


Рис. 2. Схема определения контуров границ струи в полярной системе координат

Fig. 2. Diagram of determining jet boundary contours in the polar coordinate system

Для снижения чувствительности метода к существенным отклонениям фрагментов огнетушащего вещества, распыленного в результате движения, от основного потока, формирующего траекторию струи, накладываются ограничения на разницу между значениями  $r_{\max}$ ,  $r_{\min}$  соседних сечений.

Постобработка распознанной траектории струи осуществляется в два этапа:

- коррекция первичных контуров верхней и нижней границ: удаление точек, находящихся на значительном расстоянии от соседних, заполнение существенных разрывов на основе интерполяции [14] и т.д.;

- сглаживание скорректированных контуров границ; в силу особенностей формирования и движения потока огнетушащего вещества, а также зашумления изображения струи внешними объектами или фрагментами ее распыленной части соответствующая траектория имеет изломанный вид, затрудняющий восприятие, хранение и анализ полученных результатов, поэтому для решения данных проблем в предлагаемой технологии применяется методика сглаживания на основе кубических сплайнов [14].

Для реализации основных операций предлагаемой технологии были разработаны следующие программные модули:

- модуль коррекции изображений, основные функциональные возможности которого заключаются в устранении геометрических искажений и предварительной обработке изображений (рис. 1); разработан в рамках данного исследования с использованием библиотеки `opencv` [15], реализующей основные алгоритмы по работе с цифровыми изображениями в системах технического зрения;

- модуль распознавания границ струи, основные функциональные возможности которого заключаются в формировании множества точек огнетушащего вещества, масштабировании координат с учетом геометрической привязки и выделения первичных контуров огибающих кривых; разработан в рамках данного исследования без использования дополнительных специализированных библиотек;

- модуль постобработки, основные функциональные возможности которого заключаются в коррекции и сглаживании выделенных контуров верхней и нижней границ огнетушащего вещества; разработан в рамках данного исследования с использованием библиотеки `scipy` [16], реализующей основные алгоритмы научных и инженерных расчетов.

## Результаты

Представим результаты использования разработанного метода для распознавания траектории на основе цифрового изображения струи воды из лафетного ствола. Соответствующие натурные испытания производились на экспериментальном полигоне Инженерного центра пожарной робототехники «ЭФЭР» [17] в безветренный день, а потому основной поток огнетушащего вещества двигался в одной плоскости без существенных отклонений от нее. Цифровая камера с объективом, имеющим незначительную дисторсию, была установлена строго по горизонтальному уровню на большом удалении от пожарного ствола так, что ее матрица параллельна плоскости струи. Обеспечение таких условий проведения эксперимента позволило минимизировать геометрические искажения, а потому их коррекция не осуществлялась.

В условиях эксперимента (существенное изменение цвета неба вдоль горизонтальной оси, наличие шумовых внешних объектов, а также фрагментов распыленной воды, в том числе сливающихся с фоном) выделение границ струи было сложным. Последовательное выполнение основных операций этапа предварительной обработки изображения (<http://www.swsys.ru/uploaded/image/2024-2/16.jpg>) позволило обеспечить требуемый перепад яркостей почти на всей траектории движения потока огнетушащего вещества, за исключением небольшого участка в ее конце.

Представленные результаты выделения множества точек струи воды на основе предварительно обработанного изображения (синие точки) и распознавания ее верхней и нижней границ с использованием прямоугольной (зеленые точки) и полярной (красные точки) систем координат (<http://www.swsys.ru/uploaded/image/2024-2/17.jpg>) показывают, что на участках компактной и раздробленной струй оба метода дают схожие результаты.

Однако в распыленной области точность и пространственное разрешение выделенных контуров верхней и нижней границ струи с использованием полярной системы координат выше. На рисунке 3 представлены результаты постобработки распознанной траектории струи (красные точки): скорректированные (зеленые точки) и сглаженные (синие линии) контуры границ, а на рисунке 4 – их совмещение с анализируемым изображением после предварительной обработки.

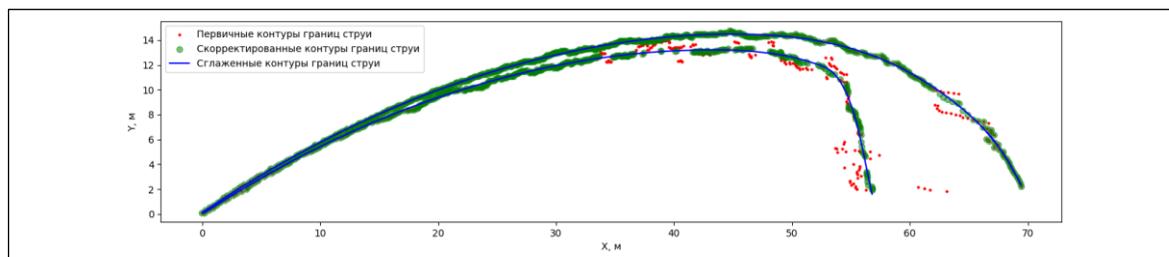


Рис. 3. Результаты постобработки распознанной струи

Fig. 3. Results of post-processing of the recognized jet



Рис. 4. Совмещенное изображение струи и ее границ, построенных на основе распознавания

Fig. 4. A combined image of the jet and its boundaries based on recognition

Распознанная траектория с достаточно высокой точностью совпадает с границами цифрового изображения огнетушащего вещества. На основе полученных кривых также были определены дальность и высота струи воды из пожарного ствола: 69,4 м и 14,6 м соответственно. Аналогичные величины, измеренные вручную с использованием наложенной на фотографию координатной сетки, имеют значения 70 м и 14,7 м – относительные погрешности не превышают 1 %.

### Заключение

Представленная технология предназначена для получения эмпирических данных с высокой точностью и пространственным разрешением, что дает возможность применять их для определения тактико-технических характеристик ствольной техники, валидации различных моделей прогнозирования траектории струи, а также формирования выборок для решения соответствующих задач машинного обучения. Существенное снижение трудоемкости обработки за счет автоматизации позволяет анализировать большие объемы результатов натурных испытаний. Также имеется возможность адаптации технологии для распознавания траекторий струй в трехмерном пространстве, например, при смещении огнетушащего вещества под воздействием бокового ветра, с использованием синхронизированных по времени изображений

в различных плоскостях и соответствующей коррекции перспективных искажений.

Технология коррекции на основе существующих методов обработки цифровых изображений, последовательно применяемых и настраиваемых с учетом условий проведения соответствующих натурных испытаний, позволяет упростить распознавание траектории струй при их визуальном зашумлении внешними объектами или распыленными фрагментами огнетушащего вещества за счет создания существенного перепада яркостей на его границах. Использование метода сечений в полярных координатах для первичного выделения исследуемых кривых, учитывающего их особенности, повышает разрешение получаемых на его основе результатов по сравнению с применением декартовой системы. Разработка указанных методик составляет научную новизну данного исследования.

Основным недостатком рассмотренной технологии является значительное (несколько секунд) время распознавания струи (предварительная обработка изображения, выделение контура, постобработка результатов), что не позволяет использовать ее в режиме реального времени, в частности, в системах технического зрения пожарных роботов. Однако увеличить скорость выполнения основных операций можно за счет, во-первых, снижения пространственного разрешения траектории, высокое значение которого не требуется для таких задач

управления пожарными роботами, как наведение струи на цель, а во-вторых, применения нейросетевых моделей, позволяющих распознавать различные объекты в режиме реаль-

ного времени, например, YOLO, для обучения которых можно использовать выборки, сформированные на основе представленной в данной статье технологии.

### Список литературы

1. Zhu J., Li W., Lin D., Zhao G. Study on water jet trajectory model of fire monitor based on simulation and experiment. *Fire Technology*, 2019, vol. 55, pp. 773–787. doi: 10.1007/s10694-018-0804-1.
2. Горбань Ю.И., Немчинов С.Г., Цариченко С.Г., Туровский А.А. Роботизированные установки пожаротушения. Параметры эффективного тушения // *Пожаровзрывобезопасность*. 2023. Т. 32. № 4. С. 58–67. doi: 10.22227/0869-7493.2023.32.04.58-67.
3. Пожаркова И.Н. Применение методов машинного обучения для прогнозирования траектории струи огнетушащего вещества из лафетного ствола // *Актуальные проблемы пожарной безопасности: матер. XXXV междунар. науч.-практич. конф.* 2023. С. 334–341.
4. Vabalas A., Gowen E., Poliakoff E., Casson A.J. Machine learning algorithm validation with a limited sample size. *PLoS One*, 2019, vol. 14, no. 11, art. e0224365. doi: 10.1371/journal.pone.0224365.
5. Patil A., Rane M. Convolutional neural networks: an overview and its applications in pattern recognition. In: *SIST. Proc. ICTIS*, 2021, vol. 195, pp. 21–30. doi: 10.1007/978-981-15-7078-0\_3.
6. Zhang X.Y., Liu C.L., Suen C.Y. Towards robust pattern recognition: A review. *Proc. of the IEEE*, 2020, vol. 108, no. 6, pp. 894–922. doi: 10.1109/jproc.2020.2989782.
7. Zhu J., Pan L., Zhao G. An improved near-field computer vision for jet trajectory falling position prediction of intelligent fire robot. *Sensors*, 2020, vol. 20, no. 24, art. 7029. doi: 10.3390/s20247029.
8. Zhou X., Xin Y., Dorofeev S. Evaluation of automatic water cannon for fire protection of low-piled storage in high-clearance facilities. *Fire Safety J.*, 2021, vol. 120, art. 103080. doi: 10.1016/j.firesaf.2020.103080.
9. Liu Y., Cheng D., Wang Q. et al. Optical distortion correction considering radial and tangential distortion rates defined by optical design. *Results in Optics*, 2021, vol. 3, art. 100072. doi: 10.1016/j.rio.2021.100072.
10. Reinhard E., Stark M., Shirley P., Ferwerda J. Photographic tone reproduction for digital images. In: *Seminal Graphics Papers: Pushing the Boundaries*, 2023, vol. 2, art. 69, pp. 661–670. doi: 10.1145/3596711.3596781.
11. Pierson C., Cauwerts C., Bodart M., Wienold J. Tutorial: Luminance maps for daylighting studies from high dynamic range photography. *Leukos*, 2021, vol. 17, no. 2, pp. 140–169. doi: 10.1080/15502724.2019.1684319.
12. Пожаркова И.Н., Цариченко С.Г., Немчинов С.Г. Моделирование траектории струи огнетушащего средства из пожарного лафетного ствола при возмущающих воздействиях // *Безопасность труда в промышленности*. 2022. № 11. С. 7–13. doi: 10.24000/0409-2961-2022-11-7-13.
13. Paolanti M., Frontoni E. Multidisciplinary pattern recognition applications: A review. *Comput. Sci. Review*, 2020, vol. 37, art. 100276. doi: 10.1016/j.cosrev.2020.100276.
14. Bertolazzi E., Frego M., Biral F. Point data reconstruction and smoothing using cubic splines and clusterization. *Math. and Comput. in Simulation*, 2020, vol. 176, pp. 36–56. doi: 10.1016/j.matcom.2020.04.002.
15. Gollapudi S. OpenCV with Python. In: *Learn Computer Vision Using OpenCV*, 2019, pp. 31–50. doi: 10.1007/978-1-4842-4261-2\_2.
16. Virtanen P., Gommers R., Oliphant T.E. et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 2020, vol. 17, no. 3, pp. 261–272. doi: 10.1038/s41592-019-0686-2.
17. Горбань Ю.И., Немчинов С.Г. Пожарные роботы в пожарной автоматике: научно-технические исследования, алгоритмы поведения и дизайн // *Пожаровзрывобезопасность*. 2020. Т. 29. № 5. С. 82–88.

## Recognizing jet trajectories of fire monitor extinguishing agent based on digital images

Irina N. Pozharkova <sup>1,2</sup>✉

<sup>1</sup> Siberian Fire and Rescue Academy of SFS EMERCOM of Russia, Zheleznogorsk, 662972, Russian Federation

<sup>2</sup> Siberian Federal University, Krasnoyarsk, 660041, Russian Federation

### For citation

Pozharkova, I.N. (2024) 'Recognizing jet trajectories of fire monitor extinguishing agent based on digital images', *Software & Systems*, 37(2), pp. 262–269 (in Russ.). doi: 10.15827/0236-235X.142.262-269

### Article info

Received: 28.11.2023

After revision: 11.03.2024

Accepted: 18.03.2024

**Abstract.** The paper focuses on developing a technology for high-precision recognition of fire monitor jet trajectories based on their digital images obtained from a full-scale experiment. The purpose of the study is to increase the efficiency of collecting empirical data describing jet trajectories of a fire extinguishing agent from fire monitors based on the recognizing digital images. The proposed method is based on identification of boundaries of a fire-extinguishing agent jet with geometrical reference of corresponding points. The work presents main operations of the developed procedure: correction of geo-

metric distortions including promising ones, preliminary image processing with detecting a fire extinguishing agent jet, recognition of jet boundaries taking into account geometric binding, post-processing of recognition results. In addition, the paper describes a method for selecting a primary trajectory contour based on sections in a polar coordinate system. In order to demonstrate the effectiveness of the proposed technology, the paper presents the results of using its software implementation based on the developed modules. The upper and lower boundaries of the fire monitor water jets were recognized in the image obtained in difficult conditions in terms of identifying a trajectory. The practical significance of the presented technique consists in collecting empirical data in the amount necessary for solving various problems related, in particular, to constructing and debugging software control algorithms for fire robots. At the same time, the developed technology allows using digital images obtained as a result of observation or experiment, which can have geometric distortion and noise due to external objects characteristic of large-scale jets in open and closed spaces. The scientific novelty of the research is in developing a primary jet boundary allocation method based on sections in polar coordinates, which increases the spatial resolution of recognition, and a technology of forming significant brightness jump on jet boundaries based on adaptation for solving the corresponding task and for consecutive application of the existing common techniques for digital images processing.

**Keywords:** pattern recognition, digital image, section in polar coordinates, machine learning, jet trajectory, fire monitor

### References

- Zhu, J., Li, W., Lin, D., Zhao, G. (2019) 'Study on water jet trajectory model of fire monitor based on simulation and experiment', *Fire Technology*, 55, pp. 773–787. doi: 10.1007/s10694-018-0804-1.
- Gorban, Yu.I., Nemchinov, S.G. (2023) 'Robotic fire extinguishing systems. Effective extinguishing parameters', *Fire and Explosion Safety*, 32(4), pp. 58–67 (in Russ.). doi: 10.22227/0869-7493.2023.32.04.58-67.
- Pozharkova, I.N. (2023) 'Use machine learning methods to predict the fire extinguishing agent jet trajectory from the monitor in case of disturbances', *Proc. XXXV Int. Sci. and Pract. Conf. Actual Fire Safety Problems*, pp. 334–341 (in Russ.).
- Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J. (2019) 'Machine learning algorithm validation with a limited sample size', *PloS One*, 14(11), art. e0224365. doi: 10.1371/journal.pone.0224365.
- Patil, A., Rane, M. (2021) 'Convolutional neural networks: an overview and its applications in pattern recognition', in *SIST. Proc. ICTIS*, 195, pp. 21–30. doi: 10.1007/978-981-15-7078-0\_3.
- Zhang, X.Y., Liu, C.L., Suen, C.Y. (2020) 'Towards robust pattern recognition: A review', *Proc. of the IEEE*, 108(6), pp. 894–922. doi: 10.1109/jproc.2020.2989782.
- Zhu, J., Pan, L., Zhao, G. (2020) 'An improved near-field computer vision for jet trajectory falling position prediction of intelligent fire robot', *Sensors*, 20(24), art. 7029. doi: 10.3390/s20247029.
- Zhou, X., Xin, Y., Dorofeev, S. (2021) 'Evaluation of automatic water cannon for fire protection of low-piled storage in high-clearance facilities', *Fire Safety J.*, 120, art. 103080. doi: 10.1016/j.firesaf.2020.103080.
- Liu, Y., Cheng, D., Wang, Q. et al. (2021) 'Optical distortion correction considering radial and tangential distortion rates defined by optical design', *Results in Optics*, 3, art. 100072. doi: 10.1016/j.rio.2021.100072.
- Reinhard, E., Stark, M., Shirley, P., Ferwerda, J. (2023) 'Photographic tone reproduction for digital images', in *Seminal Graphics Papers: Pushing the Boundaries*, 2, art. 69, pp. 661–670. doi: 10.1145/3596711.3596781.
- Pierson, C., Cauwerts, C., Bodart, M., Wienold, J. (2021) 'Tutorial: Luminance maps for daylighting studies from high dynamic range photography', *Leukos*, 17(2), pp. 140–169. doi: 10.1080/15502724.2019.1684319.
- Pozharkova, I.N., Tsarichenko, S.G., Nemchinov, S.G. (2022) 'Modeling the trajectory of a fire extinguishing agent jet from a fire monitor under disturbing influences', *Occupational Safety in Industry*, (11), pp. 7–13 (in Russ.). doi: 10.24000/0409-2961-2022-11-7-13.
- Paolanti, M., Frontoni, E. (2020) 'Multidisciplinary pattern recognition applications: A review', *Comput. Sci. Review*, 37, art. 100276. doi: 10.1016/j.cosrev.2020.100276.
- Bertolazzi, E., Frego, M., Biral, F. (2020) 'Point data reconstruction and smoothing using cubic splines and clusterization', *Math. and Comput. in Simulation*, 176, pp. 36–56. doi: 10.1016/j.matcom.2020.04.002.
- Gollapudi, S. (2019) 'OpenCV with Python', in *Learn Computer Vision Using OpenCV*, pp. 31–50. doi: 10.1007/978-1-4842-4261-2\_2.
- Virtanen, P., Gommers, R., Oliphant, T.E. et al. (2020) 'SciPy 1.0: Fundamental algorithms for scientific computing in Python', *Nature Methods*, 17(3), pp. 261–272. doi: 10.1038/s41592-019-0686-2.
- Gorban, Yu.I., Nemchinov, S.G. (2020) 'Fire-fighting robots are in the fire automation: science and technology studies, the behavior of algorithms and design', *Fire and Explosion Safety*, 29(5), pp. 82–88 (in Russ.).

### Авторы

**Пожаркова Ирина Николаевна**<sup>1, 2</sup>,  
к.т.н., профессор, доцент,  
pozharikova@mail.ru

### Authors

**Irina N. Pozharkova**<sup>1, 2</sup>,  
Cand. of Sci. (Engineering), Professor,  
Associate Professor, pozharikova@mail.ru

<sup>1</sup> Сибирская пожарно-спасательная академия  
ГПС МЧС России,

г. Железногорск, 662972, Россия

<sup>2</sup> Сибирский федеральный университет,  
г. Красноярск, 660041, Россия

<sup>1</sup> Siberian Fire and Rescue Academy  
of SFS EMERCOM of Russia,  
Zheleznogorsk, 662972, Russian Federation  
<sup>2</sup> Siberian Federal University,  
Krasnoyarsk, 660041, Russian Federation

## Компьютерное моделирование химического состава стали для оптимизации ее производства за счет комбинирования плавков

А.Г. Карачунов<sup>1</sup>✉, Е.В. Ершов<sup>1</sup>

<sup>1</sup> Череповецкий государственный университет,  
г. Череповец, 162600, Россия

### Ссылка для цитирования

Карачунов А.Г., Ершов Е.В. Компьютерное моделирование химического состава стали для оптимизации ее производства за счет комбинирования плавков // Программные продукты и системы. 2024. Т. 37. № 2. С. 270–281. doi: 10.15827/0236-235X.142.270-281

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 23.10.2023

После доработки: 21.02.2024

Принята к публикации: 26.02.2024

**Аннотация.** Целью работы является программная реализация поиска оптимального химического состава стали на основе объединения нескольких плавков в рамках оперативного производственного плана. В настоящее время в распоряжении предприятий черной металлургии применяется ряд программных инструментов для составления производственного расписания, в том числе позволяющих произвести расчет объема ферросплавов, необходимых для реализации заданного химического состава стали, а также выборочной проверки химических составов на комбинированность между собой. Недостаток указанных решений заключается в отсутствии возможности подбора для выполнения конкретной производственной задачи в рамках оперативного планирования химического состава, удовлетворяющего критериям минимального расхода ферросплавов и обеспечивающего минимальный объем перепроизводства полупродукта при фиксированном объеме плавки. В работе описываются программное обеспечение модели увеличения комбинированности химических составов стали при корректировке содержания химических элементов «Калькулятор комбинированности химических составов стали» и лежащие в его основе алгоритмы. Разработка основана на свойстве унификации (комбинирования) нескольких химических составов стали для различных заказов таким образом, чтобы сформированный на их основе химический состав удовлетворял требованиям стандартов на производство соответствующих заказов марок стали и имел возможность реализации в производстве. Внедрение в производство описанного решения позволит оптимизировать производственный план в части излишков, обеспечить снижение количества отказных заказов, одиночных плавков, объемов нереализованной продукции, а также повысить серийность разливаемой стали за счет внедрения новых химических составов.

**Ключевые слова:** комбинирование, оптимизация, модель, алгоритмы, программное обеспечение, химический состав стали

**Введение.** При оценке эффективности работы предприятия сталеплавильного производства одним из важнейших параметров является объем произведенных излишков. Непосредственное влияние на этот экономический показатель оказывает оптимальность распределения загрузки агрегатов. Например, на этапе внепечной обработки стали, где осуществляется добавление в плавку легирующих присадок, как правило, только установка типа пещковш принимает полупродукт в объеме более 300 тонн. Требования к химическому составу стали описываются в стандартах на ее производство, причем одна марка стали может производиться по нескольким стандартам, а один стандарт может содержать требования к нескольким маркам стали (в соответствии с ГОСТ Р 54384-2011). При грамотном подборе химического состава, который будет соответствовать требованиям стандартов для изготовления нескольких марок стали, можно обеспечить их выплавку за один производственный

цикл, тем самым обеспечив выполнение нескольких заказов.

На современных металлургических и других типах предприятий широко применяются АСУ производством, позволяющие использовать различные модели управления производственным процессом [1–3].

При использовании подобных систем наибольшую сложность представляет процесс моделирования, в рамках которого предполагается построение производственно-логистической цепи. Блок моделирования представляет собой несколько оптимизационных моделей, в составе которых должны быть модели управления производственными мощностями предприятий черной металлургии и модели оптимизации перечня заказов по различным критериям, одним из основных при этом является максимум маржинального дохода. Яркий пример таких задач – распределение работы агрегатов в рамках оперативного планирования с учетом технологических особенностей, что, в свою оче-

редь, отражает потребность предприятия в разработке и применении дополнительных цифровых моделей и методов оптимизации технологических процессов [4–6].

В работах отечественных [7–9] и иностранных [10–12] авторов приведены примеры использования различных цифровых инструментов для повышения эффективности планирования сталеплавильного производства. Их применение обеспечивает снижение себестоимости продукции, однако не учитывается подход к оптимизации химического состава стали за счет его унификации для решения нескольких производственных задач.

В работе [13] для подготовки производственного плана предложена модель увеличения комбинируемости химических составов стали при корректировке содержания химических элементов, используемая на этапе выбора химического состава для выполнения плавки, включающей в себя один или несколько заказов. Данная модель отвечает требованиям концепции бережливого производства. В ее основе лежит поэтапное формирование в рамках требований стандарта на производство заданной марки стали множества диапазонов массовых долей содержания химических элементов, а основным исходным параметром для модели является один из уже используемых на предприятии химических составов. В результате такого поиска подбирается химический состав стали, имеющий наименьшие изменения относительно исходного и обеспечивающий наибольшую унификацию (выполнение правил комбинируемости) с другими химическими составами, используемыми в производстве.

Нахождение химического состава, требуемого для выполнения стандарта на производство стали с учетом наложенных ограничений, является классической задачей для применения математических методов оптимизации, обеспечивающих решение многофакторных задач [14]. В рассматриваемом случае по причине независимости параметров массовых долей содержания для разных химических элементов необходимо применение поиска (перебора) по множеству вариантов их диапазонов с ограничениями.

Использование данной модели при составлении производственных планов позволяет добиться сокращения излишков при каждой плавке и в то же время предотвратить чрезмерный расход легирующих добавок (ферросплавов).

Таким образом, в данной статье предлагается рассмотреть алгоритмическое обеспече-

ние указанной модели и ее программную реализацию.

### Алгоритмы моделирования поиска оптимального химического состава

Для работы модели используются БД «Химические составы», содержащая сведения об используемых на предприятии химических составах с указанием соответствующих им марок стали и стандартов на их производство, и «Стандарты на производство» с описанием требований стандартов на производство марок стали.

Алгоритм предварительной обработки исходных данных (рис. 1) заключается в извлечении необходимых параметров из БД, их подготовке и в формировании структур данных, требуемых для работы модели увеличения комбинируемости химических составов стали.

Входные данные представляют собой набор следующих параметров:

- химический состав; это набор записей о  $\omega_{\min}^i$  минимальных и  $\omega_{\max}^i$  максимальных допустимых значениях массовых долей содержания химических элементов  $E_i$ , где  $i$  – номер химического элемента; в одном из полей данных о химическом составе хранится список комбинируемых с ним химических составов, используемых на предприятии, аналогичный список формируется для диапазона  $\omega^i$  по одному из химических элементов;

- марка стали; выбирается из перечня реализуемых выбранным химическим составом;

- стандарт на производство выбранной марки стали; выбирается из перечня стандартов на производство, структура данных аналогична химическому составу.

Представим этапы алгоритма обработки исходных данных об используемых в производстве химических составах стали и стандартах на их производство.

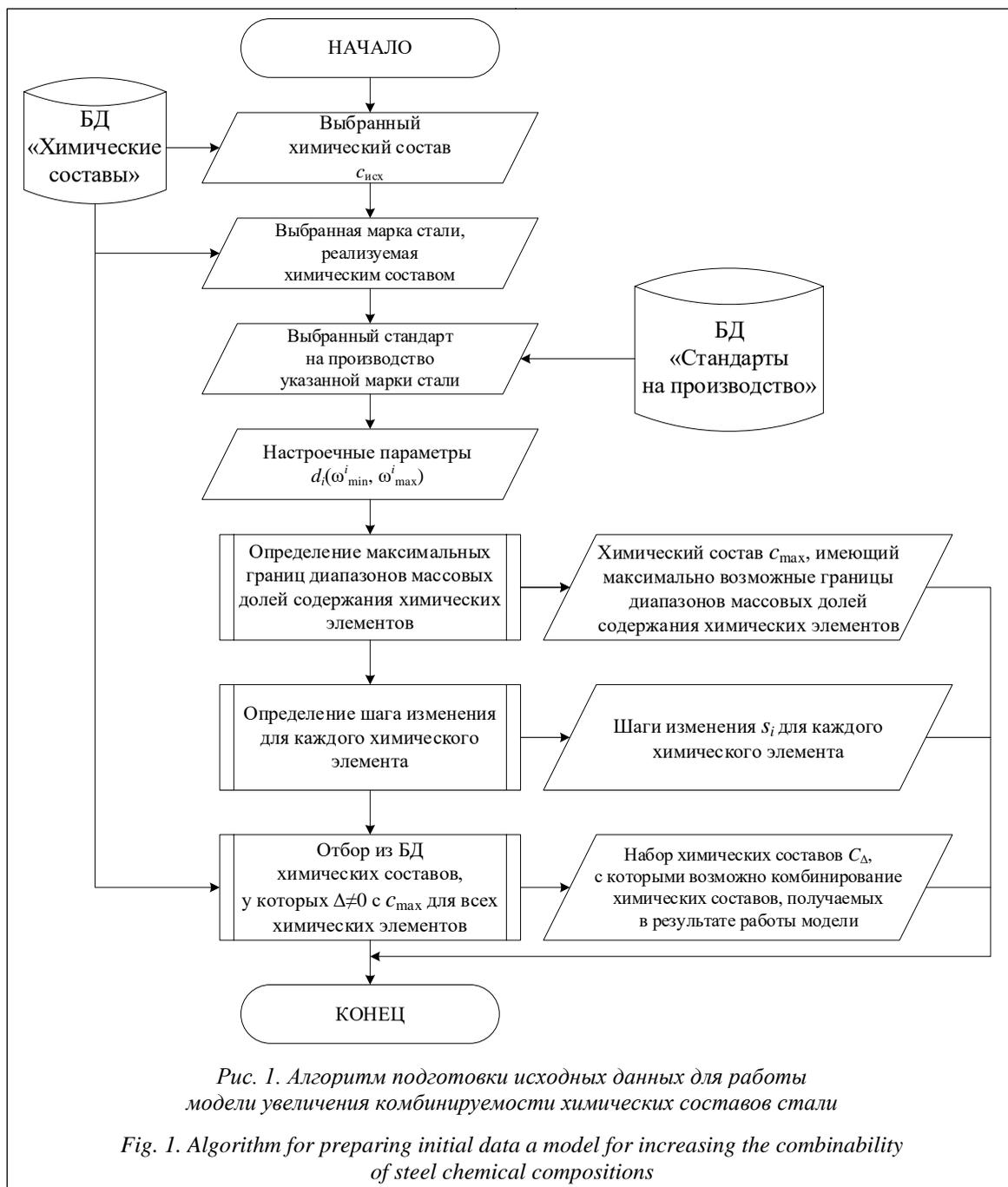
*Этап 1.* Выбор химического состава  $c_{\text{исх}}$  из БД «Химические составы».

*Этап 2.* Выбор из БД «Химические составы» марки стали по ключу идентификатора химического состава.

*Этап 3.* Выбор из БД «Стандарты на производство» стандарта на производство указанной на втором этапе марки стали.

*Этап 4.* Задание настроечных параметров  $d_i(\omega_{\min}^i, \omega_{\max}^i)$  для каждого химического элемента.

Значения данных параметров определяются технологией производства, для каждого хими-



ческого элемента может быть установлено несколько значений в зависимости от того, в каких пределах находятся установленные описанием химического состава границы массовых долей содержания элемента.

*Этап 5.* Определение максимально широких границ диапазонов массовых долей содержания для измененного химического состава.

С учетом требований выбранного стандарта на производство указанной марки стали, а также производственной задачи определяется химический состав  $C_{max}$ , имеющий максимальные

границы диапазонов массовых долей содержания химических элементов.

*Этап 6.* Определение шага изменения  $S_i$  для каждого химического элемента, содержащегося в выбранном химическом составе.

Размерность шагов зависит от производственных требований и возможностей по реализации отдачи материала в плавку с заданной точностью.

*Этап 7.* Выборка  $C_{\Delta}$  из БД «Химические составы» тех химических составов, которые имеют ненулевые пересечения диапазонов мас-

совых долей содержания для каждого химического элемента, что означает их потенциальную возможность для комбинирования.

*Этап 8.* Передача сформированных данных.

Сформированные в результате работы описанного алгоритма структуры данных передаются для применения модели увеличения комбинированности химических составов стали.

Алгоритм проверки химических составов на комбинированность между собой лежит в основе модели увеличения комбинированности химических составов стали и заключается в последовательном проведении соответствующих проверок диапазонов массовых долей содержания каждого химического элемента.

Входными данными для алгоритма являются пара выбранных для проверки на комби-

нируемость химических составов  $c_1$  и  $c_2$ , а также параметры  $d_i(\omega_{\min}^i, \omega_{\max}^i)$ .

Алгоритм предусматривает выполнение следующих шагов.

*Шаг 1.* Выбор для проверки пары химических составов.

*Шаг 2.* Выполнение для  $i$  от 1 до  $n$ , где  $n$  – количество химических элементов, содержащихся в описании химических составов предприятия, цикла проверок массовых долей содержания  $i$ -го химического элемента на комбинированность между собой.

Алгоритм указанной проверки представляет собой выполнение следующих условий (рис. 2):

- в качестве входных данных берутся значения минимальных и максимальных границ

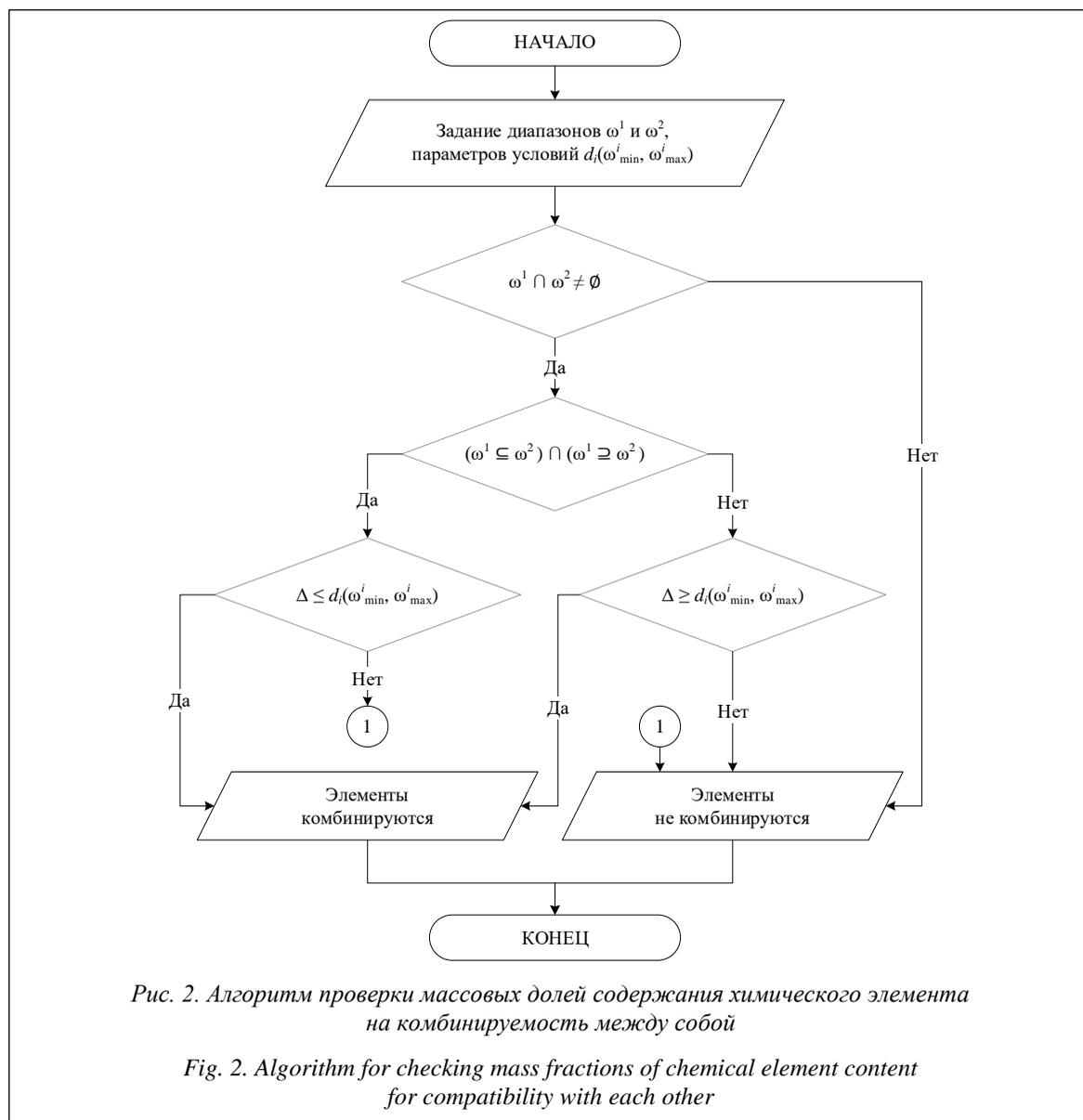


Рис. 2. Алгоритм проверки массовых долей содержания химического элемента на комбинированность между собой

Fig. 2. Algorithm for checking mass fractions of chemical element content for compatibility with each other

диапазонов массовых долей содержания химического элемента  $\omega_{\min}^1, \omega_{\max}^1, \omega_{\min}^2, \omega_{\max}^2$ , а также параметры  $d_i(\omega_{\min}^i, \omega_{\max}^i)$ ;

- вычисляется ширина пересечения заданных диапазонов; если она равна нулю, то делается вывод о том, что по данному химическому элементу комбинирование невозможно;

- в противном случае производится проверка на вхождение одного диапазона в другой;

- для случая вхождения одного диапазона в другой выполняется проверка удовлетворения условия: разница  $\Delta = |\omega_{\min}^1, \omega_{\max}^2|$  максимальных границ должна быть меньше либо равна  $d_i(\omega_{\min}^i, \omega_{\max}^i)$ ;

- для случая только пересечения диапазонов без полного вхождения одного в другой выполняется проверка удовлетворения следующего условия: разница  $\Delta = \min(\omega_{\max}^1, \omega_{\max}^2) - \max(\omega_{\min}^1, \omega_{\min}^2)$  должна быть больше либо равна  $d_i(\omega_{\min}^i, \omega_{\max}^i)$ ;

- если условие, описанное в предыдущих двух шагах, выполняется, то комбинирование по химическому элементу возможно, в противном случае – нет.

Если по какому-либо химическому элементу комбинирование невозможно, производится выход из цикла с передачей вывода о невозможности комбинирования проверяемых химических составов.

*Шаг 3.* Передача вывода о возможности комбинирования пары химических составов.

Алгоритм поиска оптимальных химических составов стали, имеющих улучшенные характеристики комбинируемости (рис. 3), применяется для формирования множества возможных химических составов с измененными относительно химического состава  $c_{\text{исх}}$  границами массовых долей содержания химических элементов и последующего выбора из него имеющих оптимальные характеристики (минимальный вес  $W(c)$ ).

В качестве входных данных алгоритма используются полученные на этапе предварительной обработки данных исходный химический состав  $c_{\text{исх}}$ , химический состав, имеющий максимальные границы диапазонов массовых долей содержания химических элементов  $c_{\text{max}}$ , рассчитанные для каждого химического элемента шаги изменения  $S_i$ , набор химических составов  $C_{\Delta}$ .

Алгоритм представляет собой выполнение следующих этапов.



*Этап 1.* Получение входных данных.

*Этап 2.* Формирование множества диапазонов массовых долей содержания с измененными границами для каждого химического элемента. Данный этап выполняется для каждого химического элемента и заключается в следующих действиях (рис. 4):

- задание начальных значений границ массовых долей содержания  $\omega_{\max}^{\text{исх}}$  и  $\omega_{\min}^{\text{исх}}$  химического состава  $c_{\text{исх}}$  соответствующим переменным  $\omega_{\min}^i$  и  $\omega_{\max}^i$ ;

- последовательное выполнение для всех химических элементов двух вложенных циклов изменения переменных  $\omega_{\min}^i$  и  $\omega_{\max}^i$  с шагом  $S_i$ ;

- на каждой итерации циклов, описанных на предыдущем шаге, перед изменением переменных формирование диапазона массовой

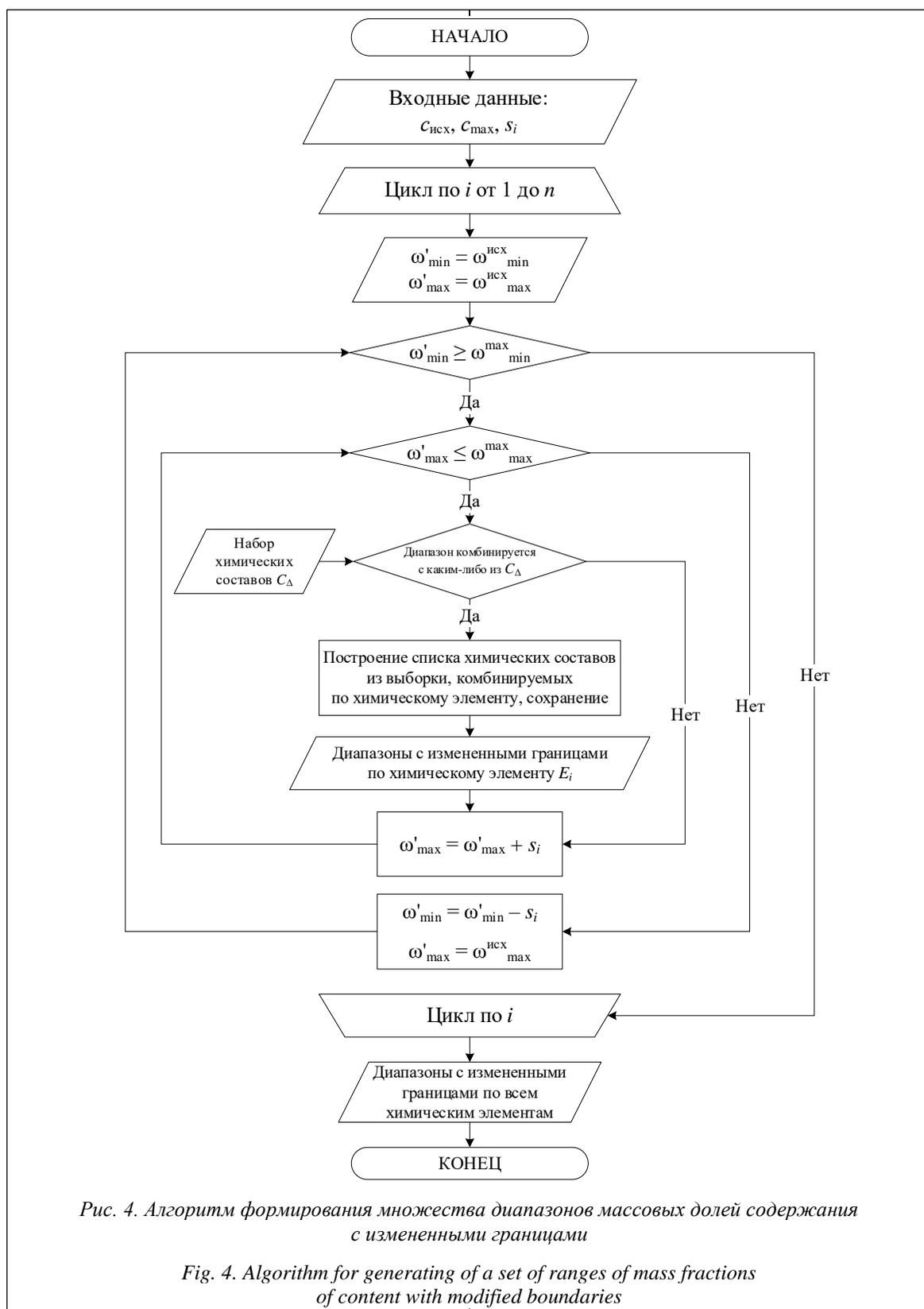


Рис. 4. Алгоритм формирования множества диапазонов массовых долей содержания с измененными границами

Fig. 4. Algorithm for generating of a set of ranges of mass fractions of content with modified boundaries

доли содержания химического элемента с границами  $(\omega'_{\min}, \omega'_{\max})$ , а затем проверка его на возможность комбинирования по химическому

элементу с соответствующими диапазонами химических составов, содержащихся в выборке  $C_{\Delta}$ ;

– если для какого-либо химического состава из выборки  $C_{\Delta}$  комбинирование возможно, то для построенного диапазона  $\omega'$  составление списка химических составов, с которыми возможно комбинирование, после чего сохранение диапазона  $\omega'$  в выходную структуру данных, в противном случае переход к следующей итерации цикла.

Выходными данными этапа является множество диапазонов массовых долей содержа-

ния химических элементов с измененными границами относительно  $C_{исх}$ .

*Этап 3.* Формирование химических составов из построенных диапазонов массовых долей содержания с измененными границами и перебор всех возможных вариаций химических составов с измененными границами.

Порядок формирования, отбора и сохранения полученных химических составов в выходную структуру данных описан на блок-схеме, представленной на рисунке 5.

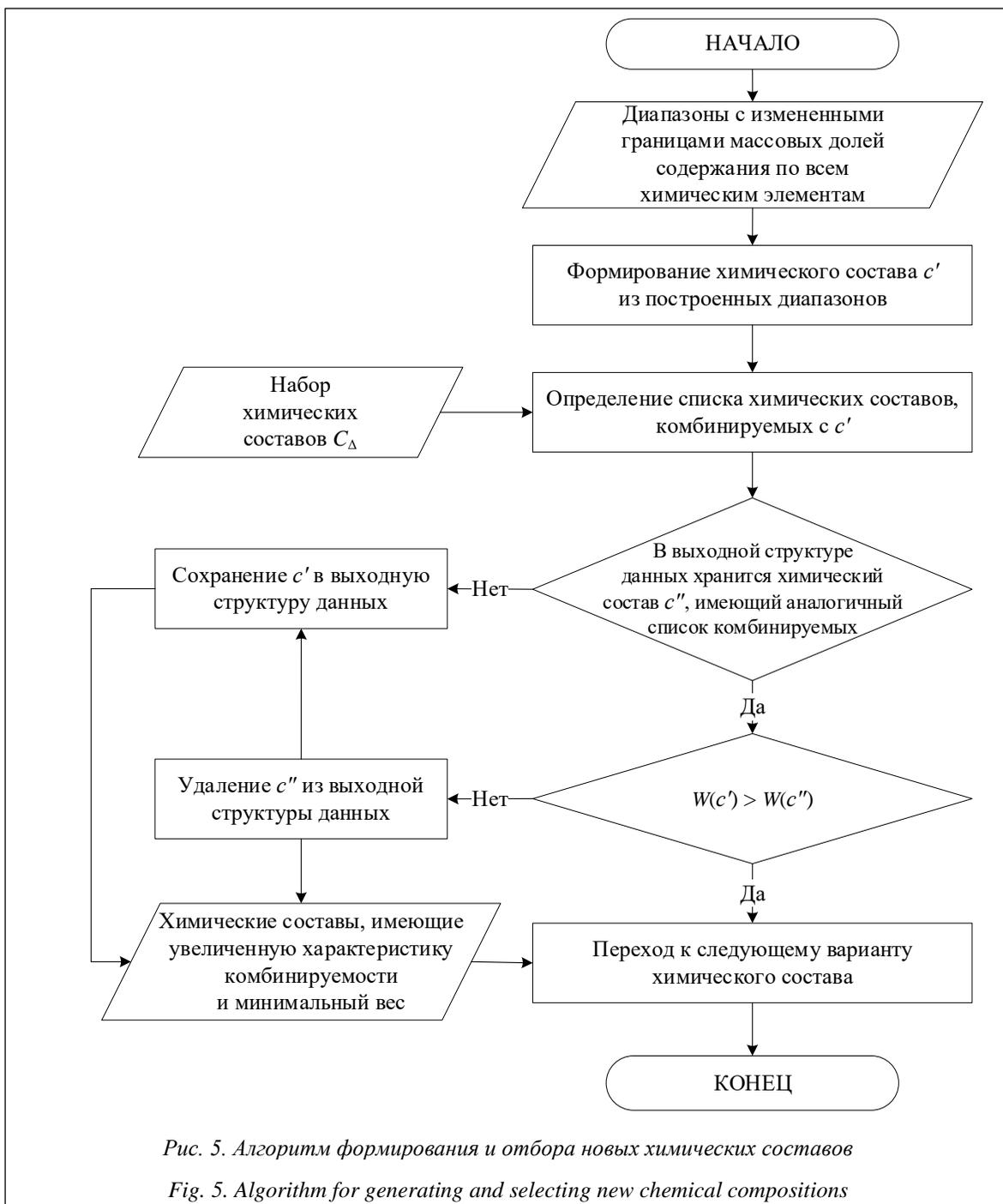


Рис. 5. Алгоритм формирования и отбора новых химических составов

Fig. 5. Algorithm for generating and selecting new chemical compositions

**Этап 4.** Передача сформированных химических составов, имеющих улучшенную характеристику комбинируемости и минимальные изменения относительно  $S_{исх}$ .

На завершающем этапе в выходной структуре данных для каждого химического состава производится расчет оценки стоимости реализации (на основе вводимых параметров стоимости химических элементов).

### Описание разработанного ПО

Представленное ПО модели увеличения комбинируемости химических составов стали «Калькулятор комбинируемости химических составов стали» [15] реализовано на языке программирования C# с использованием стандартных библиотек. Для работы созданной программы необходимо наличие установленной операционной системы семейства Windows версии 7 или выше.

Программа выполнена в составе нескольких модулей (рис. 6):

- модуль интерфейса реализует ввод исходных данных путем загрузки .xls таблиц с БД, отображение их на форму и взаимодействие с пользователем;
- модуль подготовки данных производит извлечение параметров из файлов БД и формирование необходимых структур данных для дальнейшей обработки;

- модуль обработки данных отвечает за одновременное выполнение для нескольких наборов входных данных («Химический состав», «Марка стали», «Стандарт на производство») алгоритмов формирования химических составов, имеющих улучшенную характеристику комбинируемости, и отбора оптимальных из них с последующим сохранением в выходную структуру данных;

- модуль вывода данных осуществляет расчет оценочной стоимости производственной реализации сформированных химических составов и сохранение результатов работы программы в выходном файле типа .xls.

Интерфейс программы «Калькулятор комбинируемости химических составов стали» представляет собой меню выбора пути к файлам с исходными и выходными данными, после загрузки которых происходит переход к основной форме работы с химическими составами (<http://www.swsys.ru/uploaded/image/2024-2/7.jpg>). Для удобства работы предусмотрен переход к темной версии оконного интерфейса.

На форме находятся несколько выпадающих списков, имеющих функцию контекстного поиска по вводимым данным, в которых производится выбор модифицируемого химического состава, соответствующих ему марки стали и стандарта на ее производство.

После выбора входных параметров на форме отображаются массовые доли содержа-

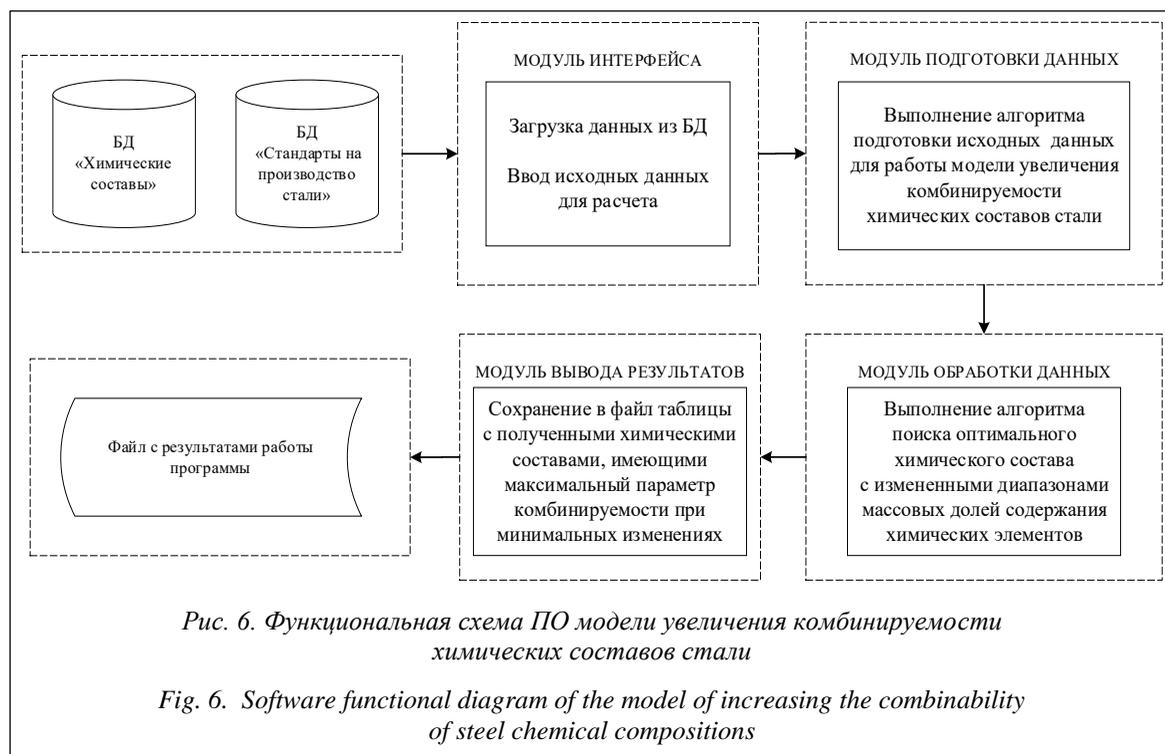


Рис. 6. Функциональная схема ПО модели увеличения комбинируемости химических составов стали

Fig. 6. Software functional diagram of the model of increasing the combinability of steel chemical compositions

ния химических элементов выбранного в качестве исходного химического состава стали, а также стандарта на производство выбранной марки стали (если стандартом не определены массовые доли содержания некоторых химических элементов, применяются табличные значения), количество химических составов, комбинируемых с выбранным, и примерное время получения результата для введенных данных. Далее на форму выводится химический состав, имеющий максимально допустимые в рамках модели границы массовых долей содержания химических элементов, а также шаги изменения исходного химического состава для каждого элемента. В целях задания точности расчетов, а также снижения количества итераций и времени работы программы для получения результатов пользователю дана возможность отредактировать исходный химический состав и шаги его изменения.

После запуска исходных параметров для обработки в правой части формы отображается статус выполнения каждой расчетной задачи с указанием времени, прошедшего с начала расчета, и примерного времени, требуемого для его завершения. Для каждой задачи предусмотрено выполнение с использованием отдельного ядра (физического или виртуального) центрального процессора рабочей станции. Также реализована возможность отмены выполнения для каждой запущенной задачи.

Результаты работы программы выгружаются в файлы электронных таблиц типа .xls в виде нескольких вкладок (<http://www.swsys.ru/uploaded/image/2024-2/8.jpg>), на которых содержится описание исходного и полученных в результате расчетов химических составов, выбранного стандарта на производство стали, а также перечня комбинируемых химических составов для каждого из вновь полученных. Измененные относительно исходного химического состава параметры выделяются цветом.

Результаты моделирования для марки стали Ст3сп и стандарта ГОСТ 380-2005, полученные для БД сталеплавильного производства, содержащей 1 250 химических составов, отражены на рисунке 7.

В данном примере продемонстрированы результаты поиска в пространстве химических составов: при изменении границ массовых долей содержания химических элементов C, S, Al химического состава номер 717 получены производные химические составы 717<sup>1</sup>, 717<sup>2</sup> и 717<sup>3</sup>. Аналогичные изменения по остальным химическим элементам в данном случае влияния на комбинируемость химических составов не оказали. Шаг изменения массовой доли содержания указанных химических элементов, равный 0,01 %, обеспечивает технологическую возможность учета такого изменения при реальной отдаче материала. Черным цветом выделены номера химических составов, с кото-

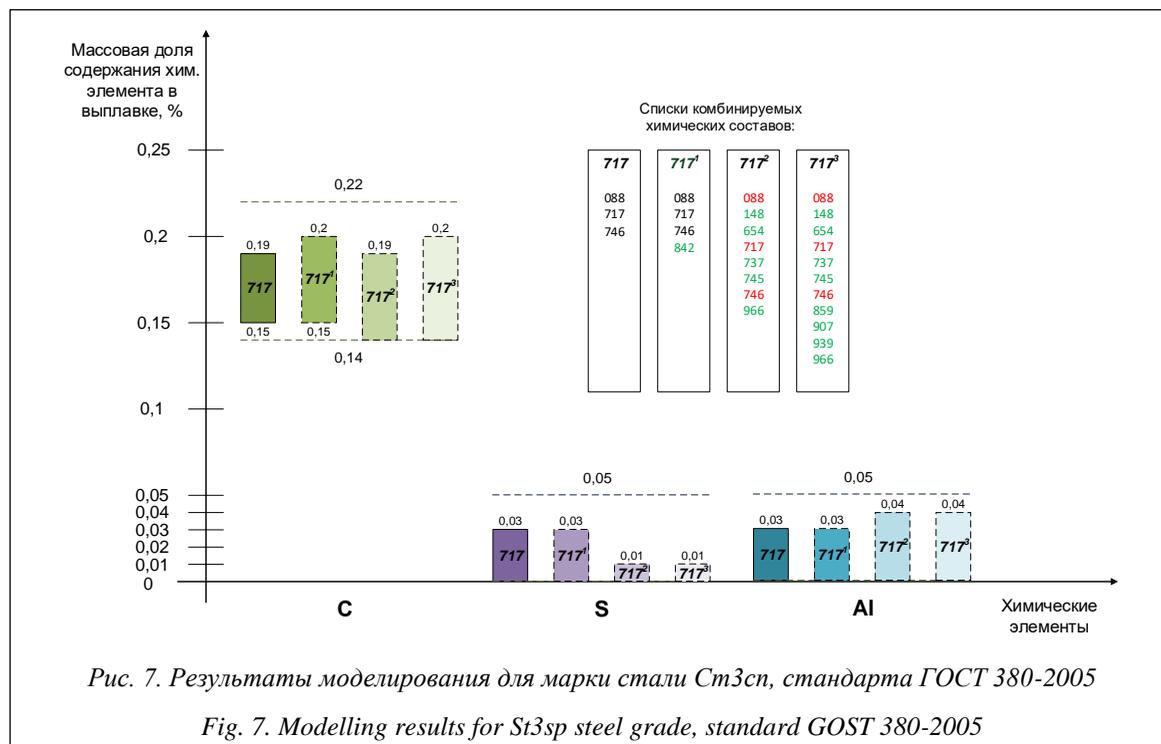


Рис. 7. Результаты моделирования для марки стали Ст3сп, стандарта ГОСТ 380-2005

Fig. 7. Modelling results for St3sp steel grade, standard GOST 380-2005

рыми комбинируется исходный, красным – те, для которых свойство комбинируемости по сравнению с исходным химическим составом утрачено, зеленым – свойство комбинируемости относительно исходного получено. Для исходного химического состава список комбинируемых химических составов, помимо его самого, содержит еще два, тогда как для вновь полученных химических составов их количество будет равно трем, пяти и восьми соответственно. В результате при минимальных изменениях исходного химического состава увеличено число возможных комбинаций выплавленных марок стали за один производственный цикл.

### Заключение

В результате проведенного исследования авторами было разработано ПО для поиска оптимальных химических составов с точки зрения комбинирования плавок стали.

Программа «Калькулятор комбинируемости химических составов стали» применяется в системе планирования сталеплавильного производства и обеспечивает возможность подбора химических составов для построения оптимального производственного расписания плавок. Кроме того, разработанный программный продукт может быть использован для синтеза новых перспективных химических составов и оценки возможности их комбинирования с уже используемыми в производстве стали.

Результаты планирования, которые получены с помощью описанного ПО, могут быть использованы в целях прогнозирования затрат на производство стали в ходе оперативного планирования, а также для поддержки принятия решения о включении заказов малой серии в производственный план. Применение программы на статистических данных сталеплавильного производства показало возможность снижения излишков выплавленного металла на 4,95 % в годовом выражении.

### Список литературы

1. Аристова Н.И. Эволюция промышленной автоматизации: 2003–2023 гг. // Автоматизация в промышленности. 2023. № 1. С. 3–10.
2. Фролов Е.Б., Нестеров П.А., Косьяненко А.В. Цифровое производство и системы производственного планирования // Автоматизация и ИТ в энергетике. 2020. № 9. С. 5–13.
3. Скорякова Е.А., Сулаберидзе В.Ш. Проблемы автоматизации процесса производственного планирования // Научно-технические проблемы в космических исследованиях Земли. 2019. Т. 11. № 1. С. 78–85.
4. Боева Л.М., Коврижных О.А. Оперативная корректировка производственных планов с использованием технологий и алгоритмов гибких производственных систем // Экономика. Информатика. 2021. Т. 48. № 4. С. 802–809. doi: 10.52575/2687-0932-2021-48-4-802-809.
5. Чистякова Т.Б., Шашихина О.Е., Новожилова И.В. Интеллектуальная информационная система оптимизации процесса производственного планирования // Проектирование и обеспечение качества информационных процессов и систем: сб. докладов. 2022. С. 124–127.
6. Данилов А.Д., Ломакин В.А. Интеллектуальная система планирования гибкого автоматизированного производства // Вестн. ВГТУ. 2019. Т. 15. № 1. С. 7–11.
7. Цуканов М.А., Божкова О.А. Повышение эффективности процессов оперативного управления и технологической координации на основе оценки производственного расписания // Вестн. ВГТУ. 2018. Т. 14. № 5. С. 27–32.
8. Коврижных О.А., Боева Л.М. Оперативное планирование сталеплавильного производства на основе прецедентов // Современные проблемы горно-металлургического комплекса. Наука и производство: матер. научн.-практ. конф. 2023. С. 404–410.
9. Зайцев А.А., Мурашев П.А., Ласкин А.А., Коряковцев А.В., Цымай Д.В. Имитационное моделирование кокалюдоменного и сталеплавильного производств ЧерМК ПАО «Северсталь» // Черная металлургия. Бюлл. науч.-технич. и экономич. информации. 2022. Т. 78. № 8. С. 657–661. doi: 10.32339/0135-5910-2022-8-657-661.
10. Sun L., Li S., Qi Y., Ma T. An improved heuristic algorithm for the order planning of steelmaking production. In: LNEE. Proc. CIAC, 2018, vol. 458, pp. 131–141. doi: 10.1007/978-981-10-6445-6\_15.
11. Xu Zj., Zheng Z., Gao, Xq. Operation optimization of the steel manufacturing process: A brief review. Int. J. Miner. Metall. Mater., 2021, vol. 28, pp. 1274–1287. doi: 10.1007/s12613-021-2273-7.
12. Wang G., Wu J., Yang Y., Su L. Robust optimization for a steel production planning problem with uncertain demand and product substitution. Computers & Operations Research, 2024, vol. 165, art. 106569. doi: 10.1016/j.cor.2024.106569.
13. Карачунов А.Г., Ершов Е.В., Виноградова П.А., Иванов А.А. Метод повышения эффективности планирования сталеплавильного производства на основе унификации химических составов стали // Вестн. ЧГУ. 2022. № 6. С. 42–52. doi: 10.23859/1994-0637-2022-6-111-3.
14. Колобов А.В., Варфоломеев И.А. Повышение эффективности бизнес-системы предприятия на основе применения цифровых инструментов в металлургии // Сталь. 2020. № 10. С. 69–73.

15. Карачунов А.Г., Ершов Е.В., Дмитриев М.В. Калькулятор комбинируемости химических составов стали: Свид. о регистр. ПрЭВМ № 2022613350. Рос. Федерация, 2022.

Software &amp; Systems

doi: 10.15827/0236-235X.142.270-281

2024, 37(2), pp. 270–281

### Computer modeling of steel chemical composition to optimize steel production by combining heats

Aleksandr G. Karachunov , Evgeny V. Ershov <sup>1</sup><sup>1</sup>Cherepovets State University, Cherepovets, 162600, Russian Federation

#### For citation

Karachunov, A.G., Ershov, E.V. (2024) 'Computer modeling of steel chemical composition to optimize steel production by combining heats', *Software & Systems*, 37(2), pp. 270–281 (in Russ.). doi: 10.15827/0236-235X.142.270-281

#### Article info

Received: 23.10.2023

After revision: 21.02.2024

Accepted: 26.02.2024

**Abstract.** The purpose of the work is software implementation of the algorithm for searching an optimal steel chemical composition based on combining several melts within the framework of an operational production plan. Nowadays, ferrous metallurgy enterprises use a number of software tools for drawing up a production schedule, including those for calculating the volume of ferroalloys required to implement a given steel chemical composition, as well as selective testing of chemical compositions for compatibility with each other. The disadvantage of these solutions is the lack of a possibility of selecting a chemical composition for solving a specific production problem in terms of operational planning that satisfies the criteria of ferroalloy minimum consumption and ensures a minimum volume of semi-product overproduction at a fixed volume of melting. The paper describes a software model for increasing the combinability of steel chemical compositions when adjusting the content of chemical elements "Calculator of combinability of steel chemical compositions" and the underlying algorithms. The development is based on the property of unifying (combining) several chemical compositions of steel for various orders in such a way that the chemical composition formed on their basis meets the requirements of standards for the production of steel grades corresponding to orders and can be implemented in production. The implementation of the described solution into production will allow optimizing a production plan in terms of surplus, ensuring a reduction in the number of rejected orders, single melts, unsold product volumes, and also increasing the serial production of cast steel due to introducing new chemical compositions.

**Keywords:** combination, optimization, model, algorithms, software, steel chemical composition

#### References

1. Aristova, N.I. (2023) 'Evolution of industrial automation in the recent 20 years', *Automation in Industry*, (1), pp. 3–10 (in Russ.).
2. Frolov, E.B., Nesterov, P.A., Kasyanenko, A.V. (2020) 'Digital production and systems production planning', *Automation and IT in the Energy Sector*, (9), pp. 5–13 (in Russ.).
3. Skorniakova, E.A., Sulaberidze, V.Sh. (2019) 'Problems of production planning process automation', *H&ES Research*, 11(1), pp. 78–85 (in Russ.).
4. Boeva, L.M., Kovrizhnykh, O.A. (2021) 'Operational rescheduling of production plans using technologies and algorithms of flexible production systems', *Economics. Inform. Tech.*, 48(4), pp. 802–809 (in Russ.). doi: 10.52575/2687-0932-2021-48-4-802-809.
5. Chistyakova, T.B., Shashikhina, O.E., Novozhilova, I.V. (2022) 'Intelligent information system for optimizing the production planning process', *Proc. IPSDQA*, pp. 124–127 (in Russ.).
6. Danilov, A.D., Lomakin, V.A. (2019) 'Intelligent system of flexible automated manufacturing scheduling', *Bull. of VSTU*, 15(1), pp. 7–11 (in Russ.).
7. Tsukanov, M.A., Bozhkova, O.A. (2018) 'Software implementation of pathfinding for multiple objects with areas of different permeability', *Bull. of VSTU*, 14(5), pp. 27–32 (in Russ.).
8. Kovrizhnykh, O.A., Boeva, L.M. (2023) 'Operational planning of steelmaking based on precedents', *Proc. Modern Problems of Mining and Metallurgical Complex. Science and Production*, pp. 404–410 (in Russ.).
9. Zaitsev, A.A., Murashev, P.A., Laskin, A.A., Koryakovtsev, A.V., Tsymai, D.V. (2022) 'Simulation modeling of ironmaking and steelmaking processes at Cherepovets steel mill of PAO Severstal', *Ferrous Metallurgy. Bull. of Sci., Tech. and Economic Information*, 78(8), pp. 657–661 (in Russ.). doi: 10.32339/0135-5910-2022-8-657-661.
10. Sun, L., Li, S., Qi, Y., Ma, T. (2018) 'An improved heuristic algorithm for the order planning of steelmaking production', in *LNEE. Proc. CIAC*, 458, pp. 131–141. doi: 10.1007/978-981-10-6445-6\_15.
11. Xu, Zj., Zheng, Z., Gao, Xq. (2021) 'Operation optimization of the steel manufacturing process: A brief review', *J. Miner. Metall. Mater.*, 28, pp. 1274–1287. doi: 10.1007/s12613-021-2273-7.
12. Wang, G., Wu, J., Yang, Y., Su, L. (2024) 'Robust optimization for a steel production planning problem with uncertain demand and product substitution', *Computers & Operations Research*, 165, art. 106569. doi: 10.1007/978-981-10-6445-6\_15.

13. Karachunov, A.G., Ershov, E.V., Vinogradova, P.A., Ivanov, A.A. (2022) 'Method to increase the efficiency of steelmaking production planning based on the standardization of chemical compositions of steel', *ChSU Bull.*, (6), pp. 42–52 (in Russ.). doi: 10.23859/1994-0637-2022-6-111-3.

14. Kolobov, A.V., Varfolomeev, I.A. (2020) 'Increasing the business system efficiency of an enterprise based on the application of digital instruments in metallurgy', *Steel in Translation*, 50(10), pp. 740–744 (in Russ.).

15. Karachunov, A.G., Ershov, E.V., Dmitriev, M.V. (2022) *Calculator of Combinability of Chemical Compositions of Steel*, Pat. RF, № 2022613350.

#### Авторы

**Карачунов Александр Геннадьевич**<sup>1</sup>,  
аспирант, pmkarachunov@mail.ru

**Ершов Евгений Валентинович**<sup>1</sup>, д.т.н.,  
профессор, директор Института информационных  
технологий Череповецкого государственного  
университета, evershov@chsu.ru

<sup>1</sup> Череповецкий государственный университет,  
г. Череповец, 162600, Россия

#### Authors

**Aleksandr G. Karachunov**<sup>1</sup>,  
Postgraduate Student, pmkarachunov@mail.ru

**Evgeny V. Ershov**<sup>1</sup>, Dr.Sc. (Engineering),  
Professor, Director of the Institute  
of Information Technologies  
Cherepovets State University, evershov@chsu.ru

<sup>1</sup> Cherepovets State University, Cherepovets,  
162600, Russian Federation

## Повышение уровня автоматизации и объективности принятия решений в процессе транспортировки жидкого чугуна на производстве

В.А. Емельянов<sup>1</sup>, С.Г. Черный<sup>2,3</sup>✉, Н.Ю. Емельянова<sup>1</sup>

<sup>1</sup> Финансовый университет при Правительстве Российской Федерации, г. Москва, 125167, Россия

<sup>2</sup> Керченский государственный морской технологический университет, г. Керчь, 298300, Россия

<sup>3</sup> Херсонский технический университет, г. Геническ, 275502, Россия

### Ссылка для цитирования

Емельянов В.А., Черный С.Г., Емельянова Н.Ю. Повышение уровня автоматизации и объективности принятия решений в процессе транспортировки жидкого чугуна на производстве // Программные продукты и системы. 2024. Т. 37. № 2. С. 282–288. doi: 10.15827/0236-235X.142.282-288

### Информация о статье

Группа специальностей ВАК: 2.3.3

Поступила в редакцию: 29.09.2023

После доработки: 04.12.2023

Принята к публикации: 23.01.2024

**Аннотация.** Статья посвящена прикладной задаче разработки системы поддержки принятия решений для процесса транспортировки жидкого чугуна. Ее актуальность обусловлена необходимостью повышения уровня автоматизации и объективности принятия решений в данном процессе на производстве. При проектировании функциональной архитектуры системы использован язык моделирования ArchiMate. Для моделирования вариантов ее использования применены объектно-ориентированный подход (включая объектно-ориентированный анализ, объектно-ориентированное проектирование и программирование) и унифицированный язык моделирования UML. В ходе работы построена модель функциональной архитектуры системы поддержки принятия решений для процесса транспортировки жидкого чугуна. Выполнен анализ функционала и проведено объектно-ориентированное проектирование статической структуры системы. Разработано программное обеспечение, позволяющее вычислять вес чугуна на основе первичных данных с датчиков и осуществляющее поддержку принятия решений относительно ремонта и эксплуатации ковшей (чугуновозов). Предложенная модель функциональной архитектуры системы позволила разработать специализированное программное обеспечение для автоматизации процесса принятия решений относительно ремонта и эксплуатации ковшей. Разработанное программное обеспечение протестировано на данных об эксплуатации ковшей на металлургическом комбинате.

**Ключевые слова:** система поддержки принятия решений, программное обеспечение, архитектура, транспортировка чугуна, ковш

**Введение.** Одним из важных процессов при производстве металла является транспортировка жидкого чугуна из доменного цеха в конвертерный, которая осуществляется с помощью чугуновозных ковшей и чугуновозов. Его особенность заключается в необходимости отслеживания параметров процесса и объектов транспортировки (ковшей, жидкого чугуна) в режиме реального времени из-за термического воздействия жидкого чугуна и высоких механических нагрузок. В процессе транспортировки критически важными являются контроль и измерение температуры и веса жидкого чугуна, времени его транспортирования, так как задержки при перевозке могут напрямую повлиять на работу сталеразливочных машин. Процесс транспортировки предусматривает постоянный контроль за состоянием ковшей, включая диагностику их футеровки и определение наростов.

### Анализ источников

Для контроля и измерения параметров чугуна и ковшей (чугуновозов) существуют и

применяются разные автоматизированные системы [1–3]. Однако анализ источников показал, что контроль и измерение параметров жидкого чугуна и чугуновозов в таких системах зачастую реализуется средствами, обладающими высоким уровнем погрешности измерения [4–6]. В работе [7] описывается система оценки состояния футеровки ковшей на основе инфракрасной термометрии. В [8] предложена система контроля состояния ковша на основе LabVIEW, позволяющая отслеживать техническое состояние внутренней поверхности ковша. Обобщение результатов этих исследований позволяет сделать вывод о том, что все они направлены на повышение уровня автоматизации контроля состояния ковшей в процессе транспортировки жидкого чугуна. В работах [9, 10] отмечается, что, несмотря на существующую тенденцию к организации контроля состояния чугуна и ковшей с помощью автоматизированных систем, на практике такой контроль осуществляется технологом визуально, с ручным вводом данных о перевозках, что вносит субъективный фактор в данный процесс

и приводит к проблемным и аварийным ситуациям при транспортировке чугуна [11, 12]. Таким образом, научные исследования по автоматизации контроля и мониторинга процесса транспортировки жидкого чугуна по-прежнему актуальны.

### Проектирование системы поддержки принятия решений

Для автоматизации процесса транспортировки жидкого чугуна и поддержки принятия решений в процессе эксплуатации ковшей (чугуновозов) авторы предлагают новую систему, которая позволяет отслеживать параметры чугуна и ковша в режиме реального времени и осуществлять поддержку принятия решений о допустимости использования ковшей.

Выполнено проектирование архитектуры системы поддержки принятия решений (СППР) для процесса транспортировки жидкого чугуна. Проектируемая СППР программно реализует методы определения веса жидкого чугуна и диагностики футеровки критического футерованного оборудования, представленные в работах [13, 14]. Их программная реализация позволяет СППР осуществлять поддержку принятия решений о допустимости использования ковшей для транспортировки жидкого чугуна. Отличительной особенностью разработанного метода является определение веса жидкого чугуна с помощью многослойной нейронной сети, которая оценивает входные данные о весе чугуна, полученные тремя разными способами, что позволило снизить погрешность измерения веса жидкого чугуна [14].

В соответствии с разработанным методом для определения веса жидкого чугуна используется многослойная нейронная сеть, на вход которой поступают следующие параметры:

- значение начального веса миксера (ковша);
- вес передвижного миксера с жидким чугуном, определяемый с помощью тензодатчиков;
- вес порожнего миксера;
- интенсивность заливки чугуна в передвижной миксер;
- время начала заливки чугуна в передвижной миксер;
- время окончания заливки чугуна в передвижной миксер;
- плановый вес чугуна во время заливки;
- расчетный вес чугуна во время заливки;
- марка чугуна.

В скрытом слое нейронной сети происходит анализ этих входных данных о весе чугуна, полученных следующими способами:

– определение веса чугуна на основе данных тензодатчиков;

– определение веса на основе учета времени заливки жидкого чугуна в миксер;

– определение веса в зависимости от интенсивности заливки жидкого чугуна в миксер.

В выходном слое нейронной сети дается итоговая оценка значения веса жидкого чугуна.

Для исследования нейронной сети применялась среда MATLAB. В качестве тестовой выборки использована информация о 120 операциях взвешивания. Обучение нейронной сети осуществлялось по алгоритму обратного распространения ошибки с сигмоидальной функцией активации. Измерение качества распознавания выполнялось путем расчета среднеквадратической ошибки [14]. Для предотвращения процесса переобучения на основе графиков изменения ошибок было определено оптимальное значение числа эпох обучения, которое составило 380 эпох.

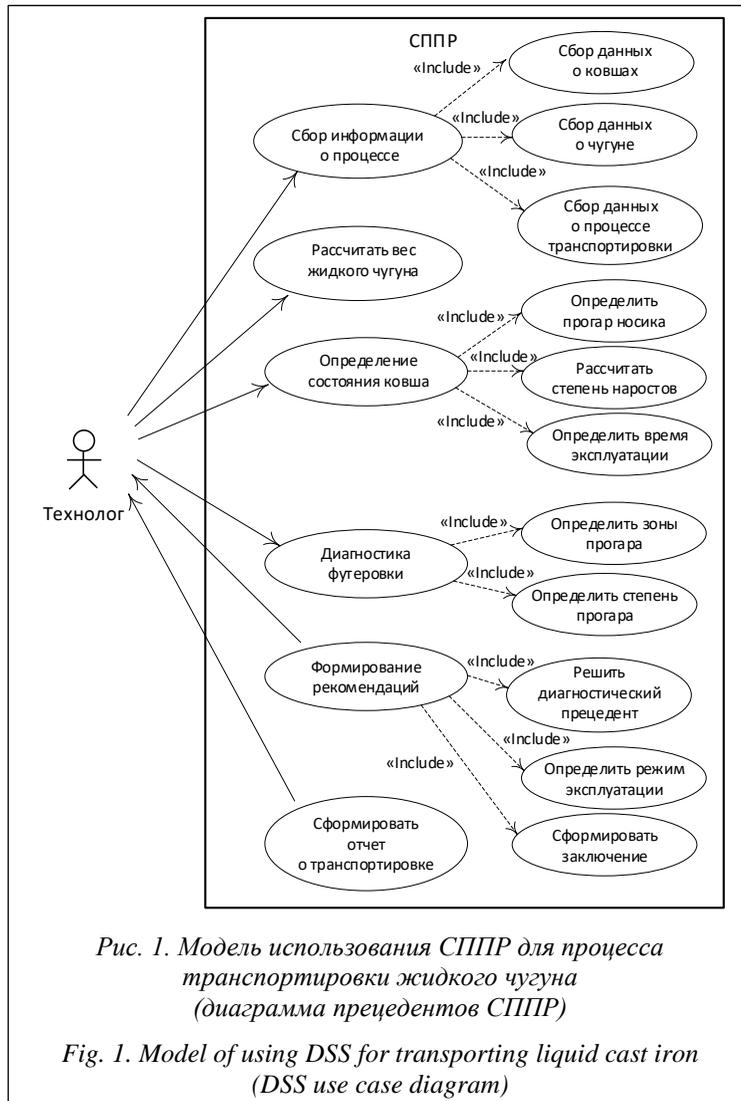
Проектирование функционала СППР для процесса транспортировки жидкого чугуна выполнено с помощью унифицированного языка моделирования. Для отражения обобщенной модели использования СППР спроектирована диаграмма вариантов (рис. 1).

Модель использования СППР для процесса транспортировки жидкого чугуна содержит шесть базовых вариантов.

1. Прецедент «Сбор информации о процессе». Эта функциональность СППР подразумевает сбор данных с датчиков, платформ и БД о ковшах, чугуне (температура, время заливки чугуна, вес чугуна и др.) и о самом процессе транспортировки (план-задание на транспортировку, эталонные значения параметров ковшей, чугуна, времени транспортировки и др.).

2. Прецедент «Рассчитать вес жидкого чугуна». В рамках данного варианта предполагается реализация разработанного авторами нейросетевого метода для вычисления массы транспортируемого жидкого чугуна. Вычисляется точное значение веса чугуна на основе обработки нейронной сетью данных о весе чугуна, полученных разными способами измерения (данные с тензодатчиков/весоизмерительной платформы, данные о весе чугуна, сформированные по времени и интенсивности его заливки в ковш) [14].

3. Прецедент «Определение состояния ковша». В ходе реализации данной функциональности выполняются сканирование ковша, вычисление степени наростов и степени прогара носика, визуализация результатов контроля со-



стояния ковша на рабочем месте технолога (диспетчера).

4. Прецедент «Диагностика футеровки». Предполагается реализация авторского метода автоматизированной диагностики футеровки на основе нейронной сети для выявления зон прогара футеровки. С этой целью устанавливается тепловизор для формирования термограмм ковша, распознавание которых осуществляется нейронной сетью. Распознавание термограмм и выявление зон прогара футеровки является основой для дальнейшей поддержки принятия решений о ремонте или эксплуатации ковшей.

5. Прецедент «Формирование рекомендаций». Данный вариант использования системы предполагает формирование рекомендаций по режиму эксплуатации ковшей на основе данных о выявленных зонах прогара футеровки, о наростах и весе перевезенного чугуна.

6. Прецедент «Сформировать отчет о транспортировке». Данная функциональность СППР позволяет формировать отчеты по процессу в соответствии с его технологической документацией.

На рисунке 2 представлена верхнеуровневая модель СППР для процесса транспортировки жидкого чугуна, построенная с помощью ПО Archi на языке моделирования архитектуры ArchiMate и наглядно отображающая базовые компоненты системы.

При проектировании архитектуры СППР для процесса транспортировки жидкого чугуна выделены категории метаданных, необходимых для ее функционирования. Основными являются следующие:

- документ-задание на транспортировку жидкого чугуна;
- тип и спецификация ковша;
- регламент (техническая документация) процесса транспортировки жидкого чугуна;
- данные о жидком чугуне (температура, вес, время нахождения в ковше и др.);
- параметры тепловизоров для формирования термограмм ковшей, необходимых для обнаружения зон прогара футеровки;
- первичные данные о футеровке (материал, состав и др.).

### Разработка ПО системы

На базе предложенной функциональной модели СППР для процесса транспортировки жидкого чугуна создано ПО, реализующее функции системы, представленные на модели использования.

Ни рисунке 3 приведено окно для вычисления веса чугуна модуля определения веса жидкого чугуна. Функции данного модуля:

- сбор первичных данных о транспортируемом чугуне (данные с тензодатчиков/весоизмерительной платформы, данные о весе чугуна, сформированные по времени и интенсивности его заливки в ковш);
- вычисление веса жидкого чугуна на основе собранных первичных данных;

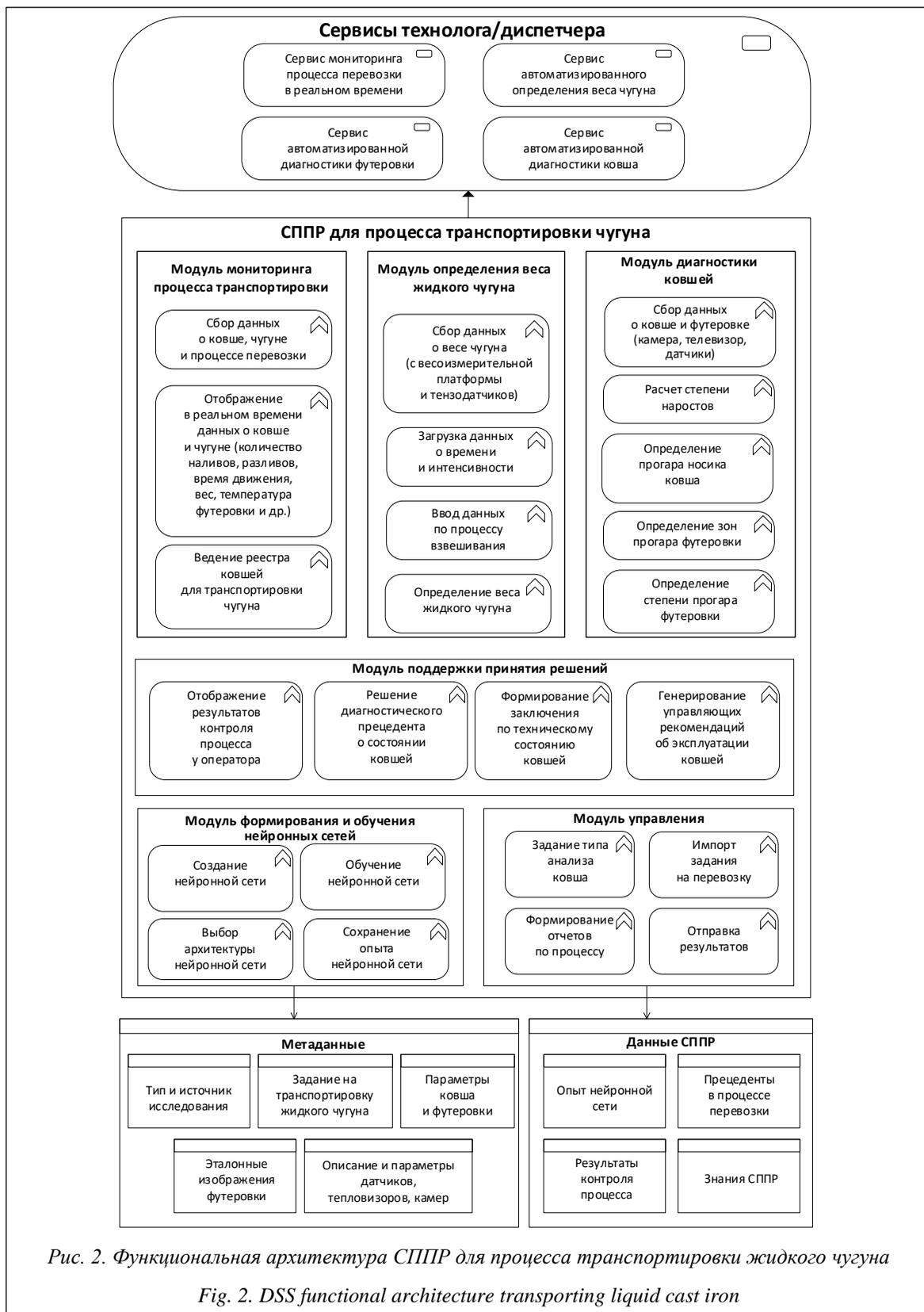


Рис. 2. Функциональная архитектура СППР для процесса транспортировки жидкого чугуна

Fig. 2. DSS functional architecture transporting liquid cast iron

– визуализация значения веса жидкого чугуна у технолога/диспетчера и в конвертерном цеху;

– сохранение результатов определения веса для формирования БД и БЗ для обучения нейронной сети.



Рис. 3. Окно для вычисления веса чугуна

Fig. 3. Window for calculating cast iron weight

Модуль поддержки принятия решений о возможности эксплуатации ковшей (<http://www.swsys.ru/uploaded/image/2024-2/6.jpg>) выполняет следующие функции:

- формирование рекомендаций о режиме эксплуатации (ремонта) ковшей (чугуновозов);
- решение диагностических ситуаций с чугуновозами на основе накопленного опыта;
- сохранение нового диагностического опыта в БД;
- формирование отчетов по операциям диагностики технического состояния ковшей.

Такой возможностью поддерживается принятие решений о допустимости использования ковшей. Данная возможность появилась за счет реализации методов автоматизированной диагностики футеровки ковшей и автоматизированного определения веса чугуна, основанных на использовании нейронных сетей и теории прецедентов для обнаружения зон прогара футеровки и вычисления веса транспортируемого чугуна.

Разработано ПО СППР для процесса транспортировки жидкого чугуна, позволяющее вычислять вес чугуна на основе первичных данных с датчиков и осуществляющее поддержку принятия решений относительно ремонта и эксплуатации ковшей.

### Список литературы

1. Jančar D., Machů M., Velička M., Tvardek P., Kocián L., Vlček J. Use of neural networks for lifetime analysis of teeming ladles. *Materials*, 2022, vol. 15, no. 22, art. 8234. doi: 10.3390/ma15228234.
2. Branca T.A., Fornai B., Colla V., Murri M.M., Streppa E., Schröder A.J. The challenge of digitalization in the steel sector. *Metals*, 2020, vol. 10, no. 2, art. 288. doi: 10.3390/met10020288.
3. Chakraborty B., Sinha B. Process-integrated steel ladle monitoring, based on infrared imaging – a robust approach to avoid ladle breakout. *QIRT J.*, 2019, vol. 17, no. 3, pp. 169–191. doi: 10.1080/17686733.2019.1639112.
4. Li G., Jiang D., Sun Y., Jiang G., Tao B. Life Prediction mechanism of ladle composite structure body based on simulation technology. *Arch. Metall. Mater.*, 2019, vol. 64, no. 4, pp. 1555–1562. doi: 10.24425/amm.2019.130126.
5. Vannucci M., Colla V., Chini M., Gasparido D., Palm B. Artificial intelligence approaches for the ladle predictive maintenance in electric steel plant. *IFAC-PapersOnLine*, 2022, vol. 55, no. 2, pp. 331–336. doi: 10.1016/j.ifacol.2022.04.215.
6. Hanoglu U., Šarler B. Hot rolling simulation system for steel based on advanced meshless solution. *Metals*, 2019, vol. 9, no. 7, art. 788. doi: 10.3390/met9070788.
7. Sawai S., Shiokawa M., Hashimoto K., Yamada Y. Methods of evaluating the damage of steelmaking refractories. *Nippon Steel Tech. Report*, 2020, no. 125, pp. 81–86.
8. Chang W., Sun Y., Li G., Jiang G., Kong J., Jiang D., Liu H. Ladle health monitoring system based on LabVIEW. *IJCSM*, 2018, vol. 9, no. 6, pp. 566–576. doi: 10.1504/IJCSM.2018.096311.
9. Colla V., Martino R., Faes A., Schivalocchi M., Romaniello L. A robot performs the maintenance of the ladle sliding gate. *Stahl und Eisen*, 2019, no. 9, pp. 44–47.
10. Lamm R., Kirchhoff S. Optimization of ladle refractory lining, gap and crack detection, lining surface temperature and sand-filling of the ladle-tap hole by means of a 3d-laserprofile-measurement system that is immersed into a hot ladle to evaluate the entire condition. *Unitecr2017*, 2022. URL: [http://www.unitecr2017.mundodecongresos.com/abstracts/Paper\\_rbofbhfcscxhpgipoispm.pdf](http://www.unitecr2017.mundodecongresos.com/abstracts/Paper_rbofbhfcscxhpgipoispm.pdf) (дата обращения: 10.08.2023).
11. Vannucci M., Ritacco A., Castellano A., Galli F., Vannucci M., Iannino V., Colla V. Flatness defect detection and classification in hot rolled steel strips using convolutional neural networks. In: *LNTCS. Proc. IWANN*, 2019, vol. 11507, pp. 220–234. doi: 10.1007/978-3-030-20518-8\_19.

12. Yang J., Zhang J., Guan M., Hong Y., Gao S., Guo W., Liu Q. Fine description of multi-process operation behavior in steelmaking-continuous casting process by a simulation model with crane non-collision constraint. *Metals*, 2019, vol. 9, no. 10, art. 1078. doi: 10.3390/met9101078.

13. Chernyi S., Emelianov V., Zinchenko E., Zinchenko A., Tsvetkova O., Mishin A. Application of artificial intelligence technologies for diagnostics of production structures. *JMSE*, 2022, vol. 10, no. 2, art. 259. doi: 10.3390/jmse10020259.

14. Yemelyanov V.A., Yemelyanova N.Yu., Nedelkin A.A., Glebov N.B., Tyapkin D.A. Information system to determine the transported liquid iron weight. *Proc. EIConRus*, 2019, pp. 377–380. doi: 10.1109/EIConRus.2019.8656693.

Software &amp; Systems

doi: 10.15827/0236-235X.142.282-288

2024, 37(2), pp. 282–288

### Increasing the level of automation and objectivity of decision-making when transporting liquid cast iron in production

Vitaly A. Yemelyanov<sup>1</sup>, Sergey G. Chernyi<sup>2,3</sup>✉, Nataliya Yu. Yemelyanova<sup>1</sup>

<sup>1</sup> Financial University under the Government of the Russian Federation, Moscow, 125167, Russian Federation

<sup>2</sup> Kerch State Maritime Technological University, Kerch, 298300, Russian Federation

<sup>3</sup> Kherson Technical University, Geniches, 275502, Russian Federation

#### For citation

Yemelyanov, V.A., Chernyi, S.G., Yemelyanova, N.Yu. (2024) 'Increasing the level of automation and objectivity of decision-making when transporting liquid cast iron in production', *Software & Systems*, 37(2), pp. 282–288 (in Russ.). doi: 10.15827/0236-235X.142.282-288

#### Article info

Received: 29.09.2023

After revision: 04.12.2023

Accepted: 23.01.2024

**Abstract.** The study solves the current applied problem of developing a decision support system for a liquid cast iron transporting process. The relevance of the work is due to the need to increase the level of automation and objectivity of decision-making when transporting liquid cast iron in production. The study uses the ArchiMate architecture modeling language to design the functional architecture of a decision support system for the process of transporting liquid cast iron. The authors used an object-oriented approach (including object-oriented analysis, object-oriented design and programming) and a unified modeling language UML to model use cases of a decision support system in the liquid iron transporting process. The study involves designing a model of the functional architecture of a decision support system for the process of transporting liquid cast iron. It also involves a functional analysis and an object-oriented design of the static structure of the decision support system. A decision support system software has been developed for transporting liquid cast iron; it allows calculating the weight of cast iron based on primary data from sensors and supporting decision-making regarding ladle repair and operation. The proposed functional model of the decision support system architecture made it possible to develop specialized software to automate the decision-making process regarding ladles repair and operation. The developed decision support system software was tested on ladle operation data at an iron and steel works.

**Keywords:** decision support system, software, architecture, cast iron transportation, ladle

#### References

1. Jančar, D., Machů, M., Velička, M., Tvardek, P., Kocián, L., Vlček, J. (2022) 'Use of neural networks for lifetime analysis of teeming ladles', *Materials*, 15(22), art. 8234. doi: 10.3390/ma15228234.
2. Branca, T.A., Fornai, B., Colla, V., Murri, M.M., Streppa, E., Schröder, A.J. (2020) 'The challenge of digitalization in the steel sector', *Metals*, 10(2), art. 288. doi: 10.3390/met10020288.
3. Chakraborty, B., Sinha, B. (2019) 'Process-integrated steel ladle monitoring, based on infrared imaging – a robust approach to avoid ladle breakout', *QIRT J.*, 17(3), pp. 169–191. doi: 10.1080/17686733.2019.1639112.
4. Li, G., Jiang, D., Sun, Y., Jiang, G., Tao, B. (2019) 'Life Prediction mechanism of ladle composite structure body based on simulation technology', *Arch. Metall. Mater.*, 64(4), pp. 1555–1562. doi: 10.24425/amm.2019.130126.
5. Vannucci, M., Colla, V., Chini, M., Gaspardo, D., Palm, B. (2022) 'Artificial intelligence approaches for the ladle predictive maintenance in electric steel plant', *IFAC-PapersOnLine*, 55(2), pp. 331–336. doi: 10.1016/j.ifacol.2022.04.215.
6. Hanoglu, U., Šarler, B. (2019) 'Hot rolling simulation system for steel based on advanced meshless solution', *Metals*, 9(7), art. 788. doi: 10.3390/met9070788.
7. Sawai, S., Shiokawa, M., Hashimoto, K., Yamada, Y. (2020) 'Methods of evaluating the damage of steelmaking refractories', *Nippon Steel Tech. Report*, (125), pp. 81–86.
8. Chang, W., Sun, Y., Li, G., Jiang, G., Kong, J., Jiang, D., Liu, H. (2018) 'Ladle health monitoring system based on LabVIEW', *IJCSM*, 9(6), pp. 566–576. doi: 10.1504/IJCSM.2018.096311.
9. Colla, V., Matino, R., Faes, A., Schivalocchi, M., Romaniello, L. (2019) 'A robot performs the maintenance of the ladle sliding gate', *Stahl und Eisen*, (9), pp. 44–47.

10. Lamm, R., Kirchhoff, S. (2022) 'Optimization of ladle refractory lining, gap and crack detection, lining surface temperature and sand-filling of the ladle-tap hole by means of a 3d-laserprofile-measurement system that is immersed into a hot ladle to evaluate the entire condition', *Unitecr2017*, available at: [http://www.unitecr2017.mundodecongresos.com/abstracts/Paper\\_rbofbhxcshpgipoispm.pdf](http://www.unitecr2017.mundodecongresos.com/abstracts/Paper_rbofbhxcshpgipoispm.pdf) (accessed August 10, 2023).
11. Vannocci, M., Ritacco, A., Castellano, A., Galli, F., Vannucci, M., Iannino, V., Colla, V. (2019) 'Flatness defect detection and classification in hot rolled steel strips using convolutional neural networks', in *LNTCS. Proc. IWANN*, 11507, pp. 220–234. doi: 10.1007/978-3-030-20518-8\_19.
12. Yang, J., Zhang, J., Guan, M., Hong, Y., Gao, S., Guo, W., Liu, Q. (2019) 'Fine description of multi-process operation behavior in steelmaking-continuous casting process by a simulation model with crane non-collision constraint', *Metals*, 9(10), art. 1078. doi: 10.3390/met9101078.
13. Chernyi, S., Emelianov, V., Zinchenko, E., Zinchenko, A., Tsvetkova, O., Mishin, A. (2022) 'Application of artificial intelligence technologies for diagnostics of production structures', *JMSE*, 10(2), art. 259. doi: 10.3390/jmse10020259.
14. Yemelyanov, V.A., Yemelyanova, N.Yu., Nedelkin, A.A., Glebov, N.B., Tyapkin, D.A. (2019) 'Information system to determine the transported liquid iron weight', *Proc. EIConRus*, pp. 377–380. doi: 10.1109/EIConRus.2019.8656693.

**Авторы**

**Емельянов Виталий Александрович**<sup>1</sup>, д.т.н.,  
профессор, v.yemelyanov@gmail.com

**Черный Сергей Григорьевич**<sup>1</sup>, к.т.н.,  
доцент, научный сотрудник,  
sergiiblack@gmail.com

**Емельянова Наталия Юрьевна**<sup>1</sup>, к.т.н.,  
доцент, n.yemelianova@gmail.com

**Authors**

**Vitaly A. Yemelyanov**<sup>1</sup>, Dr.Sc. (Engineering),  
Professor, v.yemelyanov@gmail.com

**Sergey G. Chernyi**<sup>1</sup>, Cand. of Sci. (Engineering),  
Associate Professor, Research Associate,  
sergiiblack@gmail.com

**Nataliya Yu. Yemelyanova**<sup>1</sup>, Cand. of Sci. (Engineering),  
Associate Professor, n.yemelianova@gmail.com

<sup>1</sup> Финансовый университет  
при Правительстве Российской Федерации,  
г. Москва, 125167, Россия

<sup>2</sup> Керченский государственный морской  
технологический университет,  
г. Керчь, 298300, Россия

<sup>3</sup> Херсонский технический университет,  
г. Генчешек, 275502, Россия

<sup>1</sup> Financial University under the Government  
of the Russian Federation,  
Moscow, 125167, Russian Federation

<sup>2</sup> Kerch State Maritime Technological University,  
Kerch, 298300,  
Russian Federation

<sup>3</sup> Kherson Technical University,  
Genichesk, 275502, Russian Federation

## Разработка нейросетевого метода и программного инструмента оценки потерь электроэнергии

В.Н. Пичугин <sup>1</sup>✉, А.А. Солдатов <sup>1</sup>, Е.Р. Тюрюшова <sup>1</sup>

<sup>1</sup> Чувашский государственный университет имени И.Н. Ульянова, г. Чебоксары, 428015, Россия

### Ссылка для цитирования

Пичугин В.Н., Солдатов А.А., Тюрюшова Е.Р. Разработка нейросетевого метода и программного инструмента оценки потерь электроэнергии // Программные продукты и системы. 2024. Т. 37. № 2. С. 289–294. doi: 10.15827/0236-235X.142.289-294

### Информация о статье

Группа специальностей ВАК: 1.2.2

Поступила в редакцию: 15.11.2023

После доработки: 11.01.2024

Принята к публикации: 05.02.2024

**Аннотация.** Предприятия электроэнергетики сталкиваются с большими экономическими потерями из-за недоучета электроэнергии. Предметом рассмотрения в статье является процесс сбора удаленных данных учета электроэнергии, например, цифровых подстанций предприятий и сетевых компаний, занимающихся ее передачей. Цель исследования – реализация программного инструмента обработки и предсказания данных, полученных с помощью радиоприемного цифрового модуля для обнаружения, идентификации и удаленной передачи данных на базе фрактальных антенн. Программный инструмент разработан в интегрированной среде Visual Studio 2019 с использованием оригинальных библиотек на платформе Windows. Все свойства объектов доступны для изменения, интерфейсная форма и визуальные компоненты создаются в процессе как проектирования, так и компиляции. Для программной оценки недоучета электрической энергии по показаниям использованы статистический метод и метод искусственной нейронной сети многослойного перцептрона. Подобные методы исследования больших данных, основанные на анализе показаний счетчиков электроэнергии, которые формируются в результате ежемесячного сбора, позволяют судить о их несоответствиях нормам, характеристике и типе недоучета, размерах потерь и т.д. Основным результатом является новый нейросетевой метод выявления данных о потерях электроэнергии, позволяющий обнаруживать неправильные коммерческие данные показаний электросчетчиков. Разработанный программный инструмент применяется для цифрового модуля передачи показаний приборов энергопотребления и использует умные алгоритмы метода для учета несоответствий данных о передаваемой электроэнергии на цифровых подстанциях.

**Ключевые слова:** нейросетевой метод, цифровые подстанции сетевой компании, программная оценка, статистический метод, недостоверный учет электроэнергии

**Благодарности.** Работа выполнена при финансовой поддержке Фонда содействия инновациям, федеральный проект «Платформа университетского технологического предпринимательства», договор № 303ГССС15-L/78984

**Введение.** В России активно автоматизируется сбор информации с приборов учета расхода тепла, газа, воды, электричества и т.п. Основной целью учета электроэнергии является получение достоверной информации о количестве производства, передачи, распределения и потребления электрической энергии на оптовом рынке ЕЭС России и розничном.

Ведутся разработки технологий, позволяющих передавать данные расхода ресурсов по различным проводным и эфирным каналам связи. Созданы приборы и устройства, передающие по существующим проводным сетям показания приборов учета расхода электричества. Основанием для исследований являются экономия времени и усилий по получению необходимой информации для учета показаний электроэнергии с подстанций и повышение скорости принятия решений [1].

Объект исследования – российские предприятия, отвечающие за производство, передачу и сбыт электроэнергии.

В технической литературе описаны способы построения автоматизированных информационно-измерительных систем контроля за функционированием узла энергопотребления. Так, в работе [2] представлен алгоритм анализа достоверности учета электроэнергии балансовым методом, который является основным и стандартным у всех сетевых компаний. Однако достоверность работы единичного рабочего места информационно-измерительной системы этим методом проверить нельзя. Один из авторов данной статьи (Солдатов А.А.) предложил оценивать достоверность измерений посредством сопоставления показаний дублирующих друг друга измерительных узлов учета и улучшил методику определения погрешности работы информационной системы [3]. В рабо-

те [4] предлагается использовать аналитические методы анализа, а именно решения статистического контроля режимов электропотребления, основанного на критерии достоверности учета электроэнергии. Метод предполагает анализ случайной величины, характеризующей режим работы узла учета электроэнергии, зависящей от величины тока на присоединении и равной максимальной по абсолютной величине разности значений углов фазового сдвига между током и напряжением (метод  $3\sigma$ ) [5].

Известно, что важным фактором качественного решения задачи определения достоверности учета электроэнергии является точность расчетной модели, отражающей текущее состояние режима работы узла энергопотребления [6, 7]. В оригинальных статьях [8, 9] представлены возможности использования *искусственной нейронной сети* (ИНС) в качестве основы для метода контроля функционирования информационно-измерительной системы функционирования узла энергопотребления [10].

Авторами настоящей работы спроектирован программный инструмент, позволяющий моделировать картину обработки и предсказания данных электроэнергии, алгоритмы его использования и, таким образом, учитывать несоответствия в показаниях приборов учета электрической энергии.

### Метод исследования

Построение строгой функциональной зависимости величин, определяющих значение мощности, вычисляемой приборами учета электроэнергии, является сложной задачей. В связи с этим для решения задачи верификации режимов работы узлов учета комплекса многопараметрического учета распределенного энергопотребления предложено рассмотреть альтернативный метод контроля, основанный на ИНС многослойного персептрона, способной осуществлять поиск решения в многомерном пространстве многопараметрических параметров.

### Программный инструмент и результаты его работы

На первом этапе выбирается архитектура ИНС многослойного персептрона. Обычно моделирование оптимальной структуры сети сопровождается возникновением вопросов, связанных с определением числа нейронов  $m$  в сети и количества скрытых слоев  $n$ , с использованием функции активации нейрона  $f$ , а также

алгоритма обучения сети. В данном случае число входов ИНС определяется девятью признаками, представимыми действительными числами. Задача нейронной сети состоит в определении на выходе из нее необходимого класса состояния режима работы узла учета комплекса многопараметрического учета распределенного энергопотребления в виде одного из трех возможных номинальных вариантов, определяемых вектором выходных данных:  $P_{out} = [P_{out1}, P_{out2}, P_{out3}]^T$ , где  $P_{out1}$  – режим работы «НОРМА»,  $P_{out2}$  – режим работы «НЕИСПРАВНОСТЬ»,  $P_{out3}$  – режим работы «НЕ ОПРЕДЕЛЕНО».

Выбор структуры ИНС будет опираться на то, что рассматриваются реальные данные, полученные в реальной обстановке при промышленной эксплуатации.

Эвристические правила геометрической пирамиды в вопросе подсчета числа нейронов скрытого слоя для четырехслойного персептрона с двумя скрытыми слоями:

$$r = \sqrt[3]{\frac{n}{m}}, k_1 = mr^2, k_2 = mr,$$

где  $k$  – число нейронов скрытого слоя;  $n$  – число нейронов входного слоя;  $m$  – число нейронов выходного слоя;  $k_1$  – число нейронов первого скрытого слоя;  $k_2$  – число нейронов второго скрытого слоя.

На основании приведенных выводов оптимальная структура нейронной сети определяется как (2, 4, 4, 1) (рис. 1).

В качестве обучающей выборки использованы большие данные двух статистик: первая – с действующего оборудования комплекса учета энергопотребления Алатырского объединения «Чувашэнерго», вторая – в рамках лабораторного эксперимента, проводимого на площадке факультета энергетике и электротехники в Чувашском государственном университете

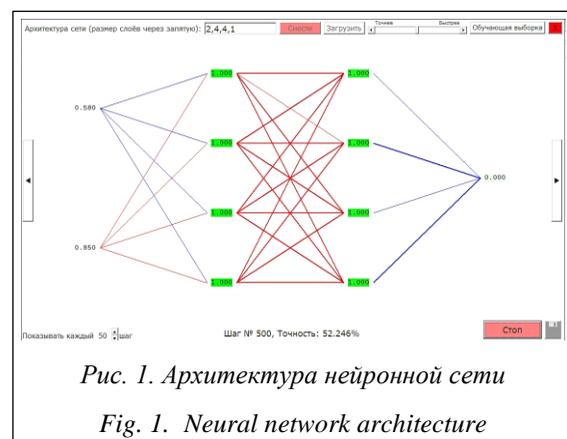


Рис. 1. Архитектура нейронной сети

Fig. 1. Neural network architecture

им. И.Н. Ульянова. Векторы данных, содержащих сведения о режимах работы («НОРМА», «НЕИСПРАВНОСТЬ» и «НЕ ОПРЕДЕЛЕНО»), подаются на разные модели ИНС с различным числом нейронов скрытого слоя. Некоторая часть статистики, участвующей в обучении, представлена в таблице.

На втором этапе осуществляется обучение нейронной сети, заключающееся в подборе весовых коэффициентов для выявления аномальных значений показаний приборов энергопотребления с малой погрешностью появления.

Оконечное устройство, например, электросчетчик, формирует показания в виде текстового файла. Показания электросчетчика отражаются следующим образом: в первой строке – дата предыдущего снятия показания электросчетчика, само показание и заводской номер электросчетчика, во второй – те же параметры, но с текущими датой и показанием счетчика, в третьей – название организации (<http://www.swsys.ru/uploaded/image/2024-2/15.jpg>).

На рисунке 2 представлена диаграмма рассеяния режимов работы приборов учета электропотребления, где в круге выделены аномальные показания счетчиков электроэнергии (неисправность оборудования, выход из строя трансформатора, обрыв линии) либо человеческий фактор (невнимательность оператора, хищение абонентами).

### Анализ результатов работы программного инструмента

Разработанный программный инструмент CR\_Energy применяется для цифрового модуля передачи показаний приборов энергопотребления, схема которого согласована с производителем автоматизированной информационно-измерительной системы дистанционного сбора и передачи удаленных данных (ОАО «Электроавтомат», г. Алатырь).

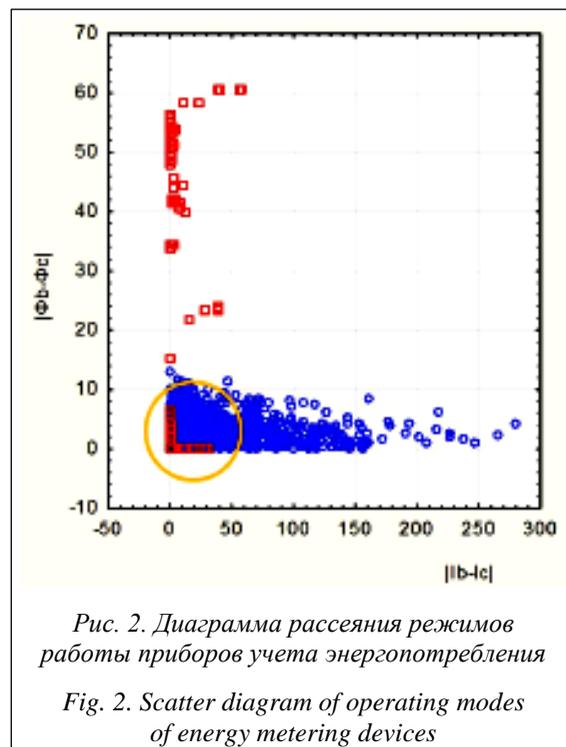


Рис. 2. Диаграмма рассеяния режимов работы приборов учета энергопотребления  
Fig. 2. Scatter diagram of operating modes of energy metering devices

Для этого спроектирована автоматизированная информационно-измерительная система дистанционного сбора и передачи удаленных данных. Система предназначена для дистанционного сбора, обработки, хранения и передачи информации о потреблении и балансе энергоресурсов. Ее основной отличительной особенностью является использование для передачи данных нелицензируемых диапазонов частот 433 МГц, 868 МГц, 2 400 МГц, а также других каналов связи. Развертывание системы сводится к установке индивидуальных, общедомовых и балансовых приборов учета, аппаратуры каналов связи, построению сети и организации диспетчерских пунктов.

Состав оборудования системы: приборы учета (счетчики) – любые приборы учета с им-

### Статистика входных параметров нейронной сети

#### Statistics of neural network input parameters

$ U_A - U_B , В$	$ U_A - U_C , В$	$ U_C - U_A , В$	$ I_A - U_B , мА$	$ I_B - I_C , мА$	$ I_C - I_A , мА$	Заключение
0,01	0,89	0,88	0,7	0,2	0,5	НОРМА
4,4	0,75	3,8	1,2	1,1	0,1	НОРМА
0,28	2,1	1,8	1,4	0,2	1,2	НОРМА
3,9	0,05	3,9	12	12	0	НЕИСПРАВНОСТЬ
0,83	4,8	5,7	0,1	0,5	0,6	НЕИСПРАВНОСТЬ
0,34	1,5	1,1	1,3	0,2	1,5	НЕИСПРАВНОСТЬ
5,3	2,7	2,7	0	0	0	НЕ ОПРЕДЕЛЕНО
4,9	2,7	2,2	0	0	0	НЕ ОПРЕДЕЛЕНО
1,9	2,1	0,21	0	0	0	НЕ ОПРЕДЕЛЕНО

пульсным выходом; приемно-передающий цифровой модуль – устройство сбора, регистрации и передачи данных по радиоканалам; ретранслятор – устройство приема и передачи данных, работающее в режиме ретрансляции; мастер считывания данных – устройство сбора и передачи данных по радиоканалам; стационарный или мобильный компьютер; разработанное ПО CR\_Energy; коммутационные устройства.

Основные возможности системы: дистанционная передача данных; расчеты на основе фактических данных о потребленных энергоресурсах; возможность выявления хищений; возможность дистанционного ограничения или отключения; дистанционная настройка параметров; возможность передачи данных на значительное расстояние с использованием режима ретрансляции; преобразование полученных данных в различные форматы, возможность работы в любых стандартных программах; сохранение полученных данных в течение 30 лет; двухуровневая защита информации – аппаратная и программная; мобильный съем показаний при помощи мастера сбора данных и ноутбука; максимальное количество точек учета – 65 534 или 216 (определяется разрядностью контроллера).

В рамках заключительного пункта проведения исследований сотрудниками предприятия-изготовителя счетчиков электрической энергии совместно с авторами разработан испытательный стенд по тестированию вновь созданной цифровой технологии удаленного сканирования и предсказания показаний счетчиков электрической энергии для отечественных производителей (рис. 3). Данный стенд прошел испытания для счетчиков электрической энергии (ЭЛТА-1, ЭЛТА-3) при передаче данных по радиоканалу.

Проведено тестирование разработанной автоматизированной информационно-измерительной системы дистанционного сбора и передачи удаленных данных на предприятии-изготовителе умных счетчиков электрической энергии с радиоканалом. Выявлены основные достоинства системы: возможность интеграции к установленным счетчикам электрической энергии независимо от марки и производителя; возможность адаптации получаемой информации к любой системе обработки данных; относительно невысокая стоимость оборудования; возможность дистанционного ограничения или отключения энергопотребления при необходимости; ведение учета и контроля потребления

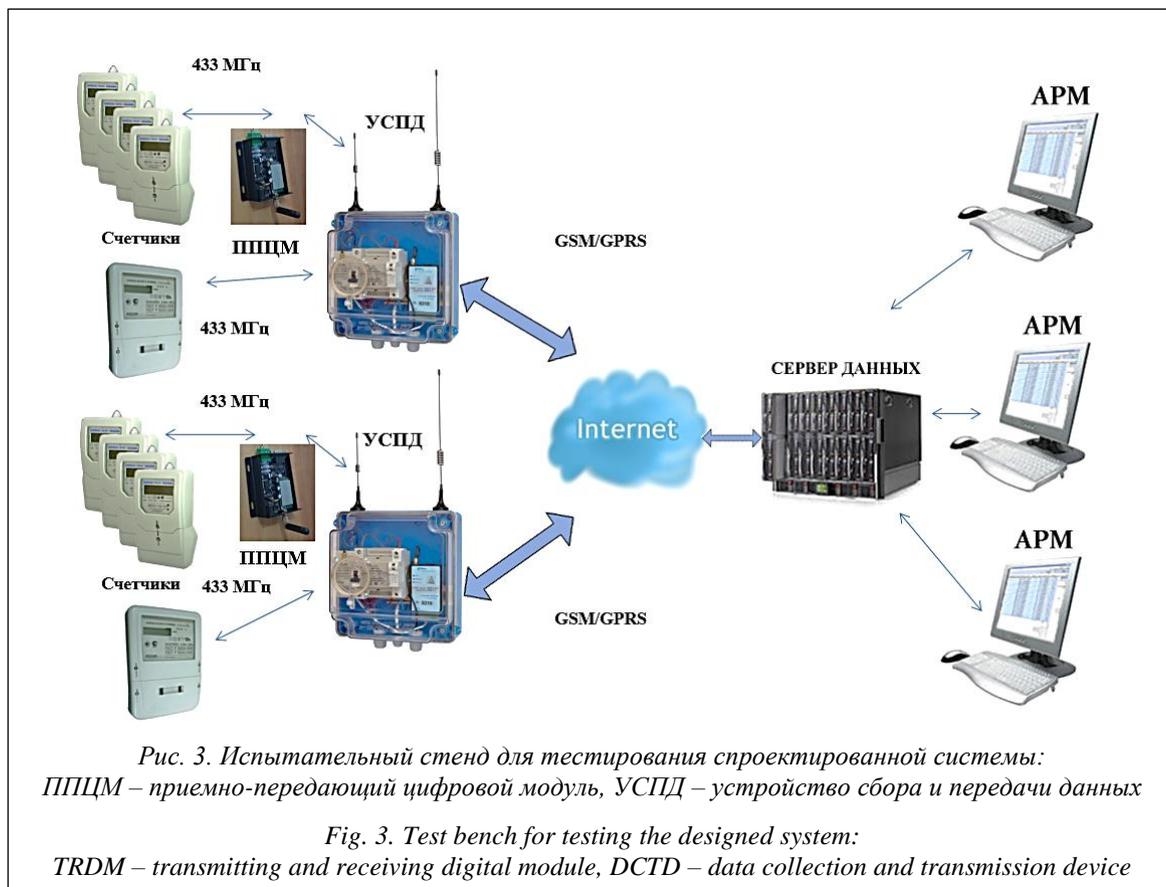


Рис. 3. Испытательный стенд для тестирования спроектированной системы: ППЦМ – приемно-передающий цифровой модуль, УСПД – устройство сбора и передачи данных

Fig. 3. Test bench for testing the designed system: TRDM – transmitting and receiving digital module, DCTD – data collection and transmission device

энергоресурсов в реальном времени, сведение баланса на объектах; использование нелицензированных диапазонов частот; надежность работы системы; простота монтажа и эксплуатации.

### Заключение и дальнейшие перспективы исследования

В результате исследований разработаны оригинальная математическая модель нейронной сети, алгоритм работы и программное средство, позволяющее сократить временные трудозатраты операторов учета и экономические потери предприятия энергосбыта, а также периодически снимать необходимость задействования автотранспорта для доставки электротехнического персонала к местам расположения распределительных подстанций.

Разработана модель нейронной сети и создан прототип программного инструмента, а также осуществлено прогнозирование показателей реальных счетчиков электрической энергии на предприятиях-изготовителях с помощью возможностей разработанного программного инструмента.

На программу для ЭВМ получено свидетельство о госрегистрации № 2022683810.

В перспективе с внедрением разработанной модели и программного инструмента в десятки раз повысится эффективность информационно-измерительных систем контроля за функционированием узла энергопотребления. Внедрение разработанного метода контроля функционирования на основе ИНС с проведением соответствующей модернизации систем учета цифровых подстанций позволит решить проблему недостоверного учета электроэнергии.

### Список литературы

1. Крикунов И.С. Основные тенденции интенсификации электроэнергетики Российской Федерации с помощью информационных технологий // Междунар. журнал информационных технологий и энергоэффективности. 2022. Т. 7. № 4. С. 22–27.
2. Пешнин С.Е. Прорывные технологии прогноза в электроэнергетике анализаторами качества электроэнергии с искусственной нейронной сетью // Электроэнергия. Передача и распределение. 2021. № 4. С. 24–32.
3. Солдатов А.А., Дубровина О.А., Жидяева Т.П. Исследование вопроса недостоверности учета электроэнергии на базовых подстанциях предприятий электроэнергетической отрасли // Актуальные вопросы современной науки: достижения и инновации: матер. Междунар. науч.-практич. конф. 2022. С. 111–115.
4. Юндин К.М., Хорольский В.Я., Исупова А.М., Руденко Н.Б. Методология выбора технических мероприятий по снижению потерь электрической энергии в практике эксплуатации сельских электрических сетей // Энергобезопасность и энергосбережение. 2023. № 4. С. 10–17.
5. Egorov A., Kochneva E., Larionova A., Lyukhanov E., Shender S.E. On-line electrical energy balance monitoring system for power networks enterprise facilities. Proc. RTUCON, 2019, pp. 1–4. doi: 10.1109/RTUCON48111.2019.8982253.
6. Алексеева Н.Н., Дулепов Д.Е. Применение методов теории систем для определения использования оптимального числа устройств синхронизированных фазных измерений в энергосистемах // Вестн. НИЭИ. 2020. № 12. С. 28–37.
7. Вуколов В.Ю., Алексеев Л.Л., Осокин С.В. Оценка достоверности показаний коммерческого и технического учета электроэнергии в сетях промышленных предприятий // Состояние и перспективы развития электро- и теплотехнологии (XXI Бенардосовские чтения): матер. Междунар. науч.-технич. конф. 2021. С. 236–239.
8. Симонов Н., Ивнев Н. Опыт и перспективы применения искусственных нейронных сетей в электроэнергетике // Электроэнергия. Передача и распределение. 2019. № 54. С. 42–48.
9. Унижаев Н.В. Особенности внедрения нейросетей и систем искусственного интеллекта на предприятиях электроэнергетики // Вопросы инновационной экономики. 2023. Т. 13. № 1. С. 215–232. doi: 10.18334/vinec.13.1.116945.
10. Калинин А.Е. Сравнительный анализ методов выявления аномального уровня электропотребления // ППРЭЭ: матер. III науч.-практич. конф. 2021. С. 203–208.

### Neural network method and a software tool for estimating power losses

Vladimir N. Pichugin ✉<sup>1</sup>, Anton A. Soldatov<sup>1</sup>, Evgeniya R. Tyuryushova<sup>1</sup>

<sup>1</sup>I.N. Ulyanov Chuvash state University, Cheboksary, 428015, Russian Federation

#### For citation

Pichugin, V.N., Soldatov, A.A., Tyuryushova, E.R. (2024) 'Neural network method and a software tool for estimating power losses', *Software & Systems*, 37(2), pp. 289–294 (in Russ.). doi: 10.15827/0236-235X.142.289-294

**Article info**

Received: 15.11.2023

After revision: 11.01.2024

Accepted: 05.02.2024

**Abstract.** Electric power companies face the problems of large economic losses due to unreliable electricity metering. The paper considers the process of collecting remote electricity metering data, for example, digital substations of enterprises and network companies engaged in electricity transmission. For this purpose, a digital module is placed at a remote substation that allows scanning electricity meter readings. The purpose of the study is to implement a software tool for processing and predicting data obtained using a digital radio receiver module for detection, identification and remote data transmission based on fractal antennas. The software tool was developed in the Visual Studio 2019 integrated environment using original libraries on the Windows platform. All object properties are available for modification, the interface form and visual components are created both during the design process and during compilation. A statistical method and the method of an artificial neural network of a multilayer perceptron were used to programmatically assess misreporting of electrical energy according to indications. Such methods of big data research based on analyzing electric energy meter readings, which are formed due to monthly collection, allow assuming their inconsistencies with standards, characteristics and a non-accounting type, the amount of losses, etc. The main result of the study is a new neural network method for detecting power loss data, which allows identifying incorrect commercial data of electric meter readings. The main practical result is developing a software tool for a digital module for transmitting readings of energy consumption devices. The module uses smart algorithms of the method to account inconsistencies in data on transmitted electricity at digital substations of PJSC ROSSETI.

**Keywords:** neural network method, unreliable electricity metering, incorrect commercial data of electricity meter readings, digital substations of a network company

**Acknowledgements.** The work was financially supported by the Innovation Promotion Fund in terms of the federal project "University technological entrepreneurship platform", contract no. 303ГССС15-Л/78984

**References**

1. Krikunov, I.S. (2022) 'The main trends in the intensification of the electric power sector of the Russian Federation with the help of information technologies', *Int. J. of Information Technology and Energy Efficiency*, 7(4), pp. 22–27 (in Russ.).
2. Peshnin, S.E. (2021) 'Breakthrough forecasting technologies in the electric power industry with power quality analyzers with an artificial neural network', *Electric Power. Transmission and Distribution*, (4), pp. 24–32 (in Russ.).
3. Soldatov, A.A., Dubrovina, O.A., Zhidyayeva, T.P. (2022) 'Investigation of the question of unreliability of electricity metering at the basic substations of the power industry enterprises', *Proc. Current Issues of Modern Science: Achievements and Innovations*, pp. 111–115 (in Russ.).
4. Yundin, K.M., Khorolskey, V.Ya., Isupova, A.M., Rudenko, N.B. (2023) 'Methodology for selecting technical measures of reducing losses in rural electrical grids', *Energy Security and Energy Saving*, (4), pp. 10–17 (in Russ.).
5. Egorov, A., Kochneva, E., Larionova, A., Lyukhanov, E., Shender, S.E. (2019) 'On-line electrical energy balance monitoring system for power networks. enterprise facilities', *Proc. RTUCON*, pp. 1–4. doi: 10.1109/RTUCON48111.2019.8982253.
6. Alekseeva, N.N., Dulepov, D.E. (2020) 'Application of methods of system theory to determine the use of the optimal number of devices for synchronized phase measurements in power systems', *Bull. NGIEI*, (12), pp. 28–37 (in Russ.).
7. Vukolov, V.Yu., Alekseev, L.L., Osokin, S.V. (2021) 'Assessment of the reliability of commercial and technical metering of electricity in the grids of industrial enterprises', *Proc. State and Prospects of Development of Electrical and Thermal Technology (XXI Benardos Readings)*, pp. 236–239 (in Russ.).
8. Simonov, N., Ivenev, N. (2019) 'Experience and prospects of using artificial neural networks in the electric power industry', *Electric Power. Transmission and Distribution*, (S4), pp. 42–48 (in Russ.).
9. Unizhaev, N.V. (2023) 'Particularities of the introduction of neural networks and artificial intelligence systems at power companies', *Russ. J. of Innovation Economics*, 13, (1), pp. 215–232 (in Russ.). doi: 10.18334/vinec.13.1.116945.
10. Kalinin, A.E. (2021) 'Comparative analysis of methods for detecting anomalous level of electric consumption', *Proc. III Sci-Pract. Conf. PPEE*, pp. 203–208 (in Russ.).

**Авторы****Пичугин Владимир Николаевич**<sup>1</sup>,

к.т.н., доцент, зав. кафедрой,

vladimir\_iii@mail.ru

**Солдатов Антон Александрович**<sup>1</sup>,

к.т.н., доцент, aa.soldatov@bk.ru

**Тюрюшова Евгения Романовна**<sup>1</sup>,

студент, evgenia.kleopatra@yandex.ru

**Authors****Vladimir N. Pichugin**<sup>1</sup>,

Cand. of Sci. (Engineering), Associate Professor,

Head of Chair, vladimir\_iii@mail.ru

**Anton A. Soldatov**<sup>1</sup>, Cand. of Sci. (Engineering),

Associate Professor, aa.soldatov@bk.ru

**Evgeniya R. Tyuryushova**<sup>1</sup>, Student,

evgenia.kleopatra@yandex.ru

<sup>1</sup> Чувашский государственный университет имени И.Н. Ульянова, г. Чебоксары, 428015, Россия

<sup>1</sup> I.N. Ulyanov Chuvash State University, Cheboksary, 428015, Russian Federation

## СО Д Е Р Ж А Н И Е

<b>Агамиров Л.В., Агамиров В.Л., Вестяк В.А.</b> Вычисления обратных функций распределений: алгоритмы и программы .....	137
<b>Родзин С.И.</b> Глобальная оптимизация на основе гибридизации алгоритмов роя саранчи и колонии пауков .....	146
<b>Напольских Д.Л.</b> Представление экономической концепции кластера с помощью специализированной онтологии .....	155
<b>Павленко Д.В., Татарис Ш.Э., Овчаренко В.В.</b> Применение глубокого обучения в интерфейсах мозг–компьютер для распознавания движений .....	164
<b>Гарев К.В.</b> Сравнительный анализ методов сжатия табличных данных .....	170
<b>Зинченко Л.А., Казаков В.В., Карышев Б.В.</b> Применение алгоритмов сжатия информации для задачи когнитивного отображения проектного пространства многослойных экранов защиты от радиации .....	178
<b>Архипов А.Н., Кондаков С.Е.</b> Сегментации файлов неисполняемых форматов для выявления угроз нарушения информационной безопасности, реализуемых в форме эксплоитов .....	186
<b>Иващенко А.В., Александрова М.В., Жейков Д.С., Мазанкина Е.В., Захарова Е.В., Колсанов А.В.</b> Реализация программного комплекса психодиагностики с обратной связью на основе компьютерного зрения .....	193
<b>Сычев О.А., Прокудин А.А., Денисов М.Е.</b> Автоматическая генерации заданий по результатам анализа использования их банка в интеллектуальной обучающей системе .....	201
<b>Мешалкин В.П., Чистякова Т.Б., Петров Д.Ю.</b> Инжиниринг цифрового тренажера для обучения операторов формования листового стекла .....	213
<b>Трубин В.Д., Павлов С.В., Христовуло О.И.</b> Обеспечение полноты и актуальности пространственной информации в распределенных вычислительных системах больших организаций .....	221
<b>Хорев П.Б., Лосев Д.А.</b> Безопасный обмен файлами на основе сетей доверия и сертификатов открытых ключей с помощью разработанного приложения .....	230
<b>Запорожец Д.Ю., Кравченко Ю.А., Кулиева Н.В., Семенова М.М., Кулиев Э.В.</b> Исследование эффективности алгоритма саранчи в задачах сегментации изображений .....	238
<b>Курашкин С.О., Бухтояров В.В., Тынченко В.С., Шуткина Е.В.</b> Разработка программного обеспечения технологического процесса электронно-лучевой сварки тонкостенных изделий .....	247
<b>Чернухин А.В., Богданова Е.А., Савицкая Т.В.</b> Построение модели предиктивной аналитики неисправностей промышленного оборудования .....	254
<b>Пожаркова И.Н.</b> Распознавание траекторий струй огнетушащего вещества из пожарного ствола на основе цифровых изображений .....	262
<b>Карачунов А.Г., Ершов Е.В.</b> Компьютерное моделирование химического состава стали для оптимизации ее производства за счет комбинирования плавок .....	270
<b>Емельянов В.А., Черный С.Г., Емельянова Н.Ю.</b> Повышение уровня автоматизации и объективности принятия решений в процессе транспортировки жидкого чугуна на производстве .....	282
<b>Пичугин В.Н., Солдатов А.А., Тюрюшова Е.Р.</b> Разработка нейросетевого метода и программного инструмента оценки потерь электроэнергии .....	289

## C O N T E N T

<b>Agamirov, L.V., Agamirov, V.L., Vestyak, V.A.</b> Calculating inverse distribution functions: Algorithms and programs.....	137
<b>Rodzin, S.I.</b> Global optimization based on hybridization of locust swarm and spider colony algorithm.....	146
<b>Napolskikh, D.L.</b> Developing a specialized ontology to represent the economic concept of a cluster.....	155
<b>Pavlenko, D.V., Tataris, Sh.E., Ovcharenko, V.V.</b> Applying deep learning in brain-computer interfaces for motion recognition .....	164
<b>Garev, K.V.</b> Compression methods for tabular data: Comparative analysis .....	170
<b>Zinchenko, L.A., Kazakov, V.V., Karyshev, B.V.</b> The problem of cognitive representation of the project space of multilayer radiation shields: Information compression algorithms.....	178
<b>Arkhipov, A.N., Kondakov, S.E.</b> Algorithm for segmentation of non-executable file formats in terms of exploit detection.....	186
<b>Ivaschenko, A.V., Aleksandrova, M.V., Zheikov, D.S., Mazankina, E.V., Zakharova, E.V., Kolsanov, A.V.</b> Psychodiagnostic software with computer vision feedback.....	193
<b>Sychev, O.A., Prokudin, A.A., Denisov, M.E.</b> Intelligent tutoring system with automatic assignment generation according to the results of using an assignment bank .....	201
<b>Meshalkin, V.P., Chistyakova, T.B., Petrov, D.Yu.</b> Digital simulator for training sheet glass forming operators.....	213
<b>Trubin, V.D., Pavlov, S.V., Khristodulo, O.I.</b> Ensuring spatial information completeness, relevance and security in distributed computing systems of large organizations .....	221
<b>Khorev, P.B., Losev, D.A.</b> Secure file sharing based on trust networks and public key certificates using a developed application.....	230
<b>Zaporozhets, D.Yu., Kravchenko, Yu.A., Kulieva, N.V., Semenova, M.M., Kuliev, E.V.</b> Study of the effectiveness of the search algorithm in image segmentation problems.....	238
<b>Kurashkin, S.O., Bukhtoyarov, V.V., Tynchenko, V.S., Shutkina, E.V.</b> Software for the technological process of electron beam welding of thin-walled products.....	247
<b>Chernukhin, A.V., Bogdanova, E.A., Savitskaya, T.V.</b> Development of predictive analytics models for industrial equipment failures .....	254
<b>Pozharkova, I.N.</b> Recognizing jet trajectories of fire monitor extinguishing agent based on digital images .....	262
<b>Karachunov, A.G., Ershov, E.V.</b> Computer modeling of steel chemical composition to optimize steel production by combining heats .....	270
<b>Yemelyanov, V.A., Chernyi, S.G., Yemelyanova, N.Yu.</b> Increasing the level of automation and objectivity of deci-sion-making when transporting liquid cast iron in production.....	282
<b>Pichugin, V.N., Soldatov, A.A., Tyuryushova, E.R.</b> Neural network method and a software tool for estimating power losses.....	289

## *Внимание подписчиков*

Научный журнал  
«Программные продукты и системы»  
распространяется по подписке  
через подписное агентство «УП УРАЛ-ПРЕСС»  
Индекс издания  
в электронном каталоге агентства 70799

*Телефон редакции: (4822) 39-91-49*  
*Электронная почта: red@cps.tver.ru*  
*Сайт журнала: www.swsys.ru*