

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

**Научный журнал Российской академии наук
(издается под руководством Отделения нанотехнологий
и информационных технологий РАН)**

Издается с 1989 года

Журнал выходит ежеквартально

Учредитель:

**Федеральный исследовательский центр
«Информатика и управление» Российской академии наук**

РЕДАКЦИОННЫЙ СОВЕТ

академик РАН И. А. Соколов — председатель Редакционного совета

академик РАН Г. И. Савин

академик РАН А. Л. Стемпковский

профессор Ш. Долев (S. Dolev, Beer-Sheva, Israel)

профессор Ю. Кабанов (Yu. Kabanov, Besancon, France)

профессор В. Ротарь (V. Rotar, San-Diego, USA)

профессор М. Финкельштейн (M. Finkelstein, Bloemfontein, South Africa)

профессор В. Хофкирхнер (W. Hofkirchner, Wien, Austria)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

академик РАН И. А. Соколов — главный редактор

проф., д.ф.-м.н. С. Я. Шоргин — заместитель главного редактора

д.т.н. В. Н. Захаров

д.ф.-м.н. В. И. Синицын

проф., д.ф.-м.н. А. И. Зейфман

проф., д.т.н. И. Н. Синицын

проф., д.т.н. В. Д. Ильин

проф., д.ф.-м.н. В. Г. Ушаков

проф., д.т.н. К. К. Колин

д.ф.-м.н. А. К. Горшенин — отв. секретарь

проф., д.ф.-м.н. В. Ю. Королев

к.ф.-м.н. С. А. Христочевский

д.ф.-м.н. Р. В. Разумчик

Редакция

к.ф.-м.н. Е. Н. Арутюнов

С. Н. Стригина

© Федеральный исследовательский центр «Информатика
и управление» Российской академии наук, 2025

Журнал включен в базу данных Russian Science Citation Index (RSCI),
интегрированную с Web of Science, и в «Белый список» научных журналов

Журнал входит в систему Российского индекса научного цитирования (РИНЦ)

Журнал включен в базу данных CrossRef (систему DOI — Digital Object Identifier),

в базу данных Ulrich's periodicals directory

и в информационную систему «Общероссийский математический портал Math-Net.Ru»

Журнал реферируется в «Реферативном журнале» ВИНИТИ

и в системе Google Scholar

Журнал включен в сформированный Минобрнауки России Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук

<http://www.ipiran.ru/journal/collected>

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 35 № 4 Год 2025

СОДЕРЖАНИЕ

Оптимизация синтеза последовательностных самосинхронных схем по синхронному описанию Л. П. Плеханов, Ю. Г. Дьяченко, Д. В. Хилько, Г. А. Орлов	4
О применении древовидных алгоритмов разрешения конфликтов в процедуре случайного множественного доступа для устройств 5G NB-IoT В. А. Бесчастный, А. М. Тюрликов, Н. В. Степанов, В. С. Шоргин	20
Нейро-нечеткая модель и прецедентная экспертная система трансформационного гибридного интеллекта прогнозирования исходов острого панкреатита С. Б. Румовская	33
Архитектура системы семантического анализа научно-технических документов А. В. Кан, Ал-др А. Хорошилов, И. А. Чечулин, С. А. Ступников, Ю. В. Никитин, Ал-ей А. Хорошилов	45
Движение, скорость и траектории представлений ключевых слов в векторном пространстве языковой модели М. М. Шарнин, Н. В. Сомин	60
Метаэвристические алгоритмы для идентификации параметров трехмерной модели снижения усталостной жесткости композиционных материалов А. В. Пантелеев, Н. В. Турбин, И. С. Надоров, Н. О. Кононов	73

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 35 № 4 Год 2025

СОДЕРЖАНИЕ

Построение и анализ модели прогнозирования
технического состояния буксовых узлов вагонов
с применением интеллектуальных методов
предиктивной аналитики

***О. В. Дружинина, Э. Р. Корепанов,
И. В. Макаренкова, В. В. Максимова, А. А. Петров*** **92**

Картографирование пахотных земель муниципальных
образований Хабаровского края по данным спутника
«Метеор-М» № 2

***А. С. Степанов, Л. В. Илларионова, К. Н. Дубровин,
Е. А. Фомина, А. Л. Верхотуров*** **111**

Современные методы извлечения географического знания
из больших массивов пространственных данных

Д. А. Никишин **129**

Подход к использованию анализа временных рядов
для формирования гипотез в конкретно-историческом
исследовании

И. М. Адамович, О. И. Волков **143**

Сложность алгоритма поиска совпадений в нескольких
последовательностях

***А. А. Грушо, М. И. Забежайло, В. В. Кульченков,
Е. Е. Тимонина*** **154**

Об авторах **164**

Авторский указатель за 2025 г. **167**

2025 Author Index **172**

Правила подготовки рукописей статей **177**

Requirements for manuscripts **181**

ОПТИМИЗАЦИЯ СИНТЕЗА ПОСЛЕДОВАТЕЛЬНОСТНЫХ САМОСИНХРОННЫХ СХЕМ ПО СИНХРОННОМУ ОПИСАНИЮ*

Л. П. Плеханов¹, Ю. Г. Дьяченко², Д. В. Хилько³, Г. А. Орлов⁴

Аннотация: Рассматривается проблема автоматизации проектирования самосинхронных (СС) цифровых схем, представляющих собой альтернативу синхронной схемотехнике, в части синтеза последовательностных устройств. Проектирование СС-схем, обладающих рядом существенных преимуществ в сравнении с синхронными аналогами, требует специфического подхода и учета особенностей дисциплины функционирования СС-схем. Наибольшую трудность представляют формализация и автоматизация синтеза последовательностных СС-устройств. Предлагается использовать для этого метод формализованной подстановки предварительно разработанных адекватных шаблонов. Он заключается в анализе исходного описания синхронного аналога синтезируемой СС-схемы с помощью программных средств логического синтезатора синхронных схем Yosys с открытым кодом, поиске фрагментов, реализуемых устройствами с памятью, и замене их СС-шаблонами, Verilog-описаниями соответствующих последовательностных СС-устройств. Приведены примеры шаблонов и описана разработанная методика их применения в процессе конвертирования исходного синхронного описания синтезируемой схемы в СС-Verilog-описание. Подстановка шаблонов в описание синтезируемой схемы избавляет от необходимости их индивидуального синтеза с учетом специфики СС-схем. Предлагаемый подход обеспечивает минимальные аппаратные затраты и оптимальное быстродействие и гарантирует самосинхронность получаемых схемотехнических реализаций цифровых устройств.

Ключевые слова: самосинхронные схемы; автоматизированный логический синтез; шаблон; последовательностные схемы; конверсия; Verilog

DOI: 10.14357/08696527250401

EDN: THWOVK

1 Введение

Самосинхронные цифровые схемы относятся к классу асинхронных цифровых схем. Благодаря двухфазной дисциплине и обязательной индикации завер-

* Исследование выполнено в рамках гранта Российского научного фонда (проект 25-19-00508).

¹ Федеральний исследовательский центр «Информатика и управление» Российской академии наук, lplekhanov@inbox.ru

² Федеральний исследовательский центр «Информатика и управление» Российской академии наук, diaura@mail.ru

³ Федеральний исследовательский центр «Информатика и управление» Российской академии наук, dhilko@yandex.ru

⁴ Федеральний исследовательский центр «Информатика и управление» Российской академии наук, Orlov.jaja@gmail.com

шения всех переключений при переходе в очередную фазу работы, они обладают уникальными особенностями, недостижимыми в синхронных и прочих асинхронных схемах [1, 2]. Они позволяют получить цифровые устройства с уникальными свойствами, востребованными во многих ответственных применениях:

- полным отсутствием состязаний;
- отказобезопасностью — остановкой работы схемы при возникновении любого числа константных неисправностей типа залипания на 0 или 1.

Эти особенности соблюдаются во всей области физической работоспособности транзисторных структур, т. е. в максимально возможном диапазоне по напряжению питания и температуре окружающей среды. Они были подтверждены прямыми экспериментами (впервые для этого класса схем) [3]. Имеется много приложений, где особенности СС-схем становятся предпочтительными при выборе типа реализации, например: устройства с повышенной надежностью [4], схемы управления устройствами с ограниченными энергетическими ресурсами, схемы с нестабильными источниками питания и др. Самосинхронные схемы имеют высокий потенциал для практического применения и находятся в фокусе современных научно-технических разработок [5–8]. Однако аппаратная избыточность СС-схем и необходимость иметь достаточно высокую квалификацию в области СС-схемотехники для их ручного проектирования сдерживают их широкое внедрение в практические разработки.

Для автоматизированного синтеза СС-схем необходима формализация специальных математических методов [1] и выполнение ряда других условий. Имеющиеся зарубежные программные средства синтеза асинхронных и самосинхронных схем (см., например, [9–14]) ориентированы на использование специальных языков описания поведения асинхронных схем для задания исходного описания синтезируемой схемы (Balsa [10, 13], сети Петри [9], Signal Transition Graph (STG) [11], Communicating Sequential Processes (CSP) [15] и др.), что делает их трудными для применения широким кругом разработчиков. Поэтому автоматизированный синтез СС-схем для многих ответственных применений, не требующий от разработчика глубоких знаний в области СС-схемотехники и специальных языков описания поведения асинхронных схем, становится актуальной задачей.

Теоретические и практические исследования методов синтеза асинхронных и самосинхронных схем [15–17] показали, что эффективный синтез СС-схем возможен и на основе исходного Verilog-описания синхронного прототипа синтезируемой схемы, представленного на любом уровне абстракции: поведенческом, архитектурном, функциональном, структурном [18–20].

В общем случае в схемах могут присутствовать как комбинационные части, так и устройства с памятью: триггеры, регистры хранения, счетчики, регистры сдвига. Синтез комбинационных СС-схем достаточно легко поддается формализации и автоматизации на основе дуализации логических функций и применения универсальных методов минимизации, декомпозиции, композиции и покрытия системы логических функций библиотечными элементами [15]. Необходимым

дополнением при синтезе СС-схемы служит подсистема контроля успешного завершения переключения схемы в очередную фазу работы — индикаторная подсхема [1, 15].

Однако использование аналогичного подхода путем прямого переложения известных методов синтеза синхронных последовательностных устройств [21] на синтез СС-схем или синтез последовательностных СС-устройств с помощью задания их функционирования средствами описания поведения асинхронных и СС-схем [22, 23] оказывается неэффективным и приводит к чрезмерной аппаратной избыточности и худшему быстродействию итоговой СС-схемы. Например, известные в технической литературе методы построения СС-триггеров используют С-элемент для реализации функции хранения и комбинационную логику с парафазным кодированием для формирования его информационных входов [15], что оказывается явно неоптимальным по аппаратным затратам.

В то же время методы оптимальной реализации перечисленных устройств с памятью известны [24–26]. Ручное проектирование в соответствии с этими методами обеспечивает получение эффективных по аппаратным затратам и быстродействию патентоспособных решений (см., например, [27]). Поэтому для автоматизации синтеза СС-схем с памятью целесообразно разработать метод формализованной подстановки оптимизированных по аппаратным затратам и быстродействию реализаций последовательностных СС-устройств, заданных в виде шаблонов на языке Verilog.

Статья посвящена разработке метода формализованного синтеза СС-схем с памятью на основе библиотеки оптимизированных параметризованных шаблонов последовательностных устройств. В литературе известны методы использования шаблонов (templates) при проектировании асинхронных схем. Например, при кодировании состояния для вставки дополнительных сигналов в STG с небольшим номером [28], при встраивании блоков PCNB (Pre-Charge Half Buffer) в схему управления асинхронным конвейером [29], при построении мелкозернистых асинхронных конвейеров с высокой пропускной способностью [30]. Данная статья предлагает метод оптимизации синтеза последовательностных СС-схем на основе исходного Verilog-описания синхронного аналога с помощью шаблонов типовых последовательностных устройств.

2 Постановка задачи

Шаблоном будем называть заранее подготовленное параметризованное описание СС-устройства с памятью на языке высокого уровня (в данном случае — Verilog), гарантирующее близкую к оптимальной по потребительским характеристикам аппаратную реализацию в базе любой заданной библиотеки стандартных элементов, расширенной необходимыми СС-элементами.

Исходной информацией для синтеза СС-схемы служат Verilog-описание синхронного аналога синтезируемой схемы на произвольном уровне абстракции

и библиотека шаблонов последовательностных синхронных и самосинхронных устройств с памятью, также представленных в виде модулей на языке Verilog.

Для повышения эффективности автоматизированного логического синтеза СС-схем с памятью (последовательностных схем) на основе исходного Verilog-описания синхронного аналога необходимо обеспечить:

- (1) конвертирование исходного синхронного описания на произвольном уровне абстракции в описание, содержащее только логические функции и вызовы шаблонов СС-устройств с памятью с конкретными параметрами;
- (2) автоматизированный синтез структурного описания СС-схемы в вызовах библиотечных элементов, в том числе последовательных СС-устройств на основе результата конвертирования.

3 Шаблоны последовательностных устройств

Самосинхронная реализация любой цифровой схемы учитывает особенности построения СС-схем:

- СС-кодирование информационных сигналов;
- двухфазную дисциплину функционирования;
- запрос-ответное взаимодействие частей СС-схемы, связанных друг с другом отношениями «задатчик сигналов – приемник».

Для комбинационных схем существует формальный способ преобразования системы логических функций в СС-реализацию:

- использование парафазного кодирования [1] для всех информационных сигналов;
- дуализация системы функций — для каждой логической функции добавляется ее инверсное представление, которое вместе с исходной функцией формирует парафазную переменную;
- добавление элементов индикации первого уровня, формирующих сигналы, показывающие своим значением фазовое состояние соответствующего парафазного сигнала;
- добавление подсхемы, собирающей все выходы элементов индикации первого уровня в один общий индикаторный выход СС-схемы.

Аналогичный формализованный метод может быть использован и для реализации устройств с памятью. Стандартные логические синтезаторы синхронных схем синтезируют многоразрядные устройства с памятью единообразно:

- функция хранения реализуется с помощью регистра хранения на двухтактных триггерах соответствующей разрядности с параллельной записью;

- алгоритм изменения состояния устройства по приходе тактового или счетного сигнала реализуется с помощью комбинационной схемы, формирующей сигналы, подаваемые на информационные входы разрядов устройства с памятью.

В случае регистров хранения комбинационная часть используется только для реализации условий записи в регистр и установки его в определенное состояние и не приводит к аппаратной избыточности. В случае синхронных регистров сдвига и счетчиков получается параллельная реализация, обладающая повышенным быстродействием за счет некоторой аппаратной избыточности.

Формальное преобразование параллельного регистра сдвига или счетчика в СС-реализацию возможно, но приводит к схемотехническому решению с чрезмерной избыточностью и низким быстродействием. Причина — индикаторная подсхема, которая в этом случае получается намного более сложной и медленной.

Самосинхронные регистры хранения, например регистры на выходе ступени конвейера, также могут быть реализованы путем замены синхронных двухтактных триггеров их СС-аналогами [24, 26]. Однако двухфазная дисциплина СС-схем позволяет упростить реализацию СС-регистра хранения: вместо двухтактных триггеров с задачей хранения вполне справляются одноктактные триггеры. Более того, использование в составе одного разряда двух гистерезисных триггеров [1] обеспечивает хранение как рабочего состояния парафазного сигнала, формируемого комбинационной частью СС-схемы, так и его спейсерного состояния [27].

Самосинхронные счетчики в отличие от синхронных аналогов не могут быть построены путем последовательного соединения двухтактных триггеров D-типа с замыканием инверсного информационного выхода на информационный вход этого же триггера из-за проблем с индикацией многоразрядного счетчика. Схемотехнические решения для реализации последовательных СС-счетчиков были предложены группой В. И. Варшавского [1]. Они обеспечивают минимальные аппаратные затраты и приемлемое быстродействие для СС-счетчиков любой разрядности. Например, моделирование схем параллельного и последовательного счетчика [25] показало, что среднее быстродействие последовательного варианта СС-счетчика в 1,7 раза выше, чем параллельного. Чем выше разрядность счетчика, тем больше преимущество в быстродействии. Поэтому для гарантированного получения СС-схемы с параметрами, близкими к оптимальным, целесообразно на начальном этапе логического синтеза заменять многоразрядные устройства с памятью, генерируемые синхронным синтезатором, готовыми СС-решениями (шаблонами), подготовленными заранее и включенными в библиотеку сложнофункциональных блоков синтезатора СС-схем [31].

Разработанные шаблоны устройств с памятью содержат готовые схемотехнические СС-реализации типовых последовательностных схем: триггеров, регистров хранения и сдвига, счетчиков. Они реализуют функциональность соответствующего устройства в структурном виде — в базисе библиотеки стандартных

ячеек. В качестве такой библиотеки может быть использована любая промышленная библиотека, дополненная специфическими элементами, необходимыми для проектирования СС-схем. Verilog-код в шаблонах содержит конкретные имена библиотечных ячеек, поэтому он настраивается индивидуально под ту библиотеку, с которой предстоит работать логическому синтезатору СС-схем.

Шаблоны последовательностных СС-устройств преобразуются в схему на библиотечных элементах любым логическим синтезатором. Логика СС-поведения и необходимая индикаторная подсхема уже реализованы в шаблоне. Поэтому полученный синхронным синтезатором результат не требует дополнительной доработки средствами синтезатора СС-схем.

4 Процедура синтеза

В предлагаемом методе используется известная открытая программная система Yosys [32] логического синтеза цифровых схем, использующая в качестве исходного задания описание алгоритма работы схемы на языке Verilog любого уровня абстракции. Применение этой системы значительно упрощает первый этап синтеза, обеспечивая преобразование поведенческого описания синтезируемой схемы в систему логических функций и процедурных блоков `always` и избавляя от необходимости разработки собственных аналогичных программных средств [31]. Система логических функций описывает комбинационную часть синтезируемой схемы. Блоки `always` задают поведение устройств с памятью.

Опыт показывает, что система Yosys преобразует описание синхронной схемы, заданное на любом уровне, в функциональное описание в операторах `assign`, описывающих логические функции с числом аргументов не более трех. С точки зрения СС-схемы такая реализация не оптимальна, поскольку сильная декомпозиция схемы приводит к увеличению сложности и индикаторной подсхемы и к ухудшению быстродействия всей СС-схемы. Кроме того, многоразрядные последовательностные устройства, как было отмечено выше, Yosys синтезирует в виде регистра хранения на двухтактных триггерах и комбинационного окружения, формирующего информационные входы регистра. В результате получается СС-схема с чрезмерной избыточностью.

В связи с этим в рамках предлагаемого подхода к синтезу СС-схем реализован метод шаблонов.

4.1 Метод шаблонов

Метод шаблонов синтеза СС-схем с памятью заключается в последовательном выполнении следующих этапов.

1. Синтаксический и семантический контроль и анализ исходного описания синхронного аналога синтезируемой СС-схемы на языке Verilog с помощью программных средств логического синтезатора синхронных схем Yosys.

2. Поиск фрагментов исходного Verilog-кода, реализуемых устройствами с памятью. Как правило, устройства с памятью в языке Verilog реализуются процедурными блоками `always`, заголовков и тело которых описывают особенности функционирования устройств.
3. Замена найденных фрагментов кода вызовами типовых модулей синхронных триггеров, регистров и счетчиков, реализующих соответствующую функциональность, с помощью инструментария Yosys.
4. Замена типовых модулей синхронных устройств с памятью шаблонами адекватных последовательностных СС-устройств с помощью табличного метода однозначного соответствия [24] на основе разработанной библиотеки шаблонов последовательностных СС-устройств.
5. Внедрение в подставляемые СС-шаблоны дополнительных атрибутов, специфицирующих СС-свойства, типа спейсера входов и выходов, типа сигналов сброса и установки (асинхронный или самосинхронный) и т. д. на основе анализа взаимодействия функциональных блоков в составе синтезируемой схемы. Атрибуты облегчают последующий синтез комбинационной части СС-схемы и организацию запрос-ответных связей между блоками схемы.

Метод шаблонов позволяет формализовать и автоматизировать конвертирование синхронных последовательностных устройств, описанных на языке Verilog, в их СС-представление, сохранив функциональность устройств с учетом особенностей реализации СС-схем. Разработанные вручную шаблоны гарантируют самосинхронность отдельно взятых последовательностных устройств в составе синтезированной схемы и обеспечивают оптимальность их характеристик.

4.2 Преобразование исходного описания

Один из ключевых этапов синтеза СС-схемы — преобразование исходного синхронного описания в промежуточное представление [31], которое удовлетворяет указанным в разд. 2 требованиям. Функциональные возможности Yosys позволяют автоматизировать данный процесс без необходимости разработки дополнительного программного обеспечения, что является несомненным преимуществом предлагаемого подхода.

Система синтеза Yosys функционирует аналогично многопроходному компилятору. Процесс синтеза организуется в виде исполнения последовательности преобразований над текущим состоянием схемы. Каждый этап преобразований в документации Yosys называется `pass`. Исходное описание на языке Verilog или VHDL преобразуется во внутренний формат Register Transfer Level Intermediate Language (RTLIL). Текущее состояние схемы всегда представляется в формате RTLIL, который передается между вызовами отдельных `pass`'ов. Ключевые `pass`'ы — `proc`, `opt`, `extract`, `techmap`, `dfflibmap`.

Этап `proc` осуществляет обработку блоков `always` исходного Verilog-описания, извлекая из них всю комбинационную логику, в результате чего каждый остав-

шийся блок `always` содержит описание логики, обеспечивающей хранение состояния (т. е. регистры и триггеры).

Этап `opt` осуществляет оптимизацию промежуточного состояния схемы и используется практически на всех этапах синтеза.

Этап `extract` — это один из вариантов технологического маппинга. Данный этап осуществляет поиск в промежуточном представлении после этапа `proc` подсхем, изоморфных заданным библиотечным описаниям. Другими словами, именно этот этап позволяет найти все фрагменты схемы, которые соответствуют искомому шаблону.

Этап `techmap` — это непосредственный технологический маппинг, который осуществляет разбиение многоразрядных элементов на группы одnorазрядных и их замену представленными в библиотеке базовыми элементами. По умолчанию используется базовая библиотека элементов, встроенная в `Yosys`, однако существует возможность разработки собственных библиотек маппинга, которые будут позволять осуществлять замену требуемыми шаблонами.

Этап `dfflibmap` — это преобразование (маппинг) триггеров в заданные библиотечные элементы. На данном этапе оставшиеся блоки `always` (а это могут быть только элементы с памятью) заменяются библиотечными триггерами.

Таким образом, для осуществления подстановки шаблонов необходимо обеспечить:

- (1) библиотеку СС-шаблонов;
- (2) библиотеку синхронных шаблонов, равномошную по числу элементов библиотеке СС-шаблонов. Каждому элементу данной библиотеки соответствует (в логическом смысле) только один элемент СС-библиотеки;
- (3) библиотеку синхронных триггеров;
- (4) библиотеку соответствующих СС-триггеров;
- (5) технологический маппер, который будет осуществлять замену обнаруженных синхронных шаблонов СС-эквивалентами.

В процессе синтеза в `Yosys` этап `extract` может быть осуществлен многократно. Очевидно, что чем полнее будут сформированы библиотеки синхронных и СС-шаблонов, тем больше будет обнаружено элементов и тем более оптимальными будут результаты преобразований. Тем не менее вероятность того, что не все шаблоны будут распознаны, не равна нулю. Однако это не помешает успешному завершению синтеза, так как после выполнения этапа `dfflibmap` нераспознанные на ранних этапах элементы памяти будут заменены соответствующими триггерами из синхронной библиотеки, для которых в библиотеке СС-триггеров уже определены эквивалентные элементы.

Технологический маппер обеспечивает корректную замену синхронных шаблонов их СС-эквивалентами. Данный подход достаточно гибкий, так как позволяет помимо непосредственной замены обеспечить генерацию требуемой для СС-элемента инфраструктуры (по сути, дополнительной логики). Однако

разработка хорошего технологического маппера представляется нетривиальной задачей.

4.3 Синтез на нижнем уровне

Задача этого этапа — распознавание вызовов шаблонов, выделение и СС-синтез комбинационных фрагментов схемы.

При вызовах шаблонов необходимо обеспечить СС-интерфейс в соответствии с их параметрами по месту вызовов информационными самосинхронными входами и выходами, а также фазовыми управляющими и индикаторными сигналами.

В финальной части этапа создается индикаторная подсхема всей создаваемой схемы.

Данный этап выполняется с помощью ранее разработанных методов [33].

5 Заключение

Применение системы Yosys с открытым кодом значительно упрощает первый этап синтеза СС-схем по исходному синхронному описанию, обеспечивая преобразование поведенческого описания синтезируемой схемы в систему логических функций и процедурных конструкций языка Verilog, задающих поведение устройств с памятью.

Предлагаемый метод шаблонов позволяет значительно упростить синтез СС-схем с памятью за счет использования заранее подготовленной библиотеки параметризованных шаблонов типовых СС-устройств с памятью: триггеров, регистров хранения и сдвига, счетчиков. Шаблоны вставляются в промежуточное Verilog-описание синтезируемой схемы с помощью механизма поиска и подстановки указанных пользователем модулей соответствующих функциональных устройств вместо фрагментов Verilog-кода. Для этого используется инструментарий синхронного логического синтезатора Yosys. Последующая интеграция последовательностных СС-устройств в синтезируемую СС-схему, согласование типов спейсера частей схемы и организация СС-управления выполняются разработанными авторами программными средствами.

Литература

1. Варшавский В. И., Кишиневский М. А., Мараховский В. Б. и др. Автоматное управление асинхронными процессами в ЭВМ и дискретных системах / Под ред. В. И. Варшавского. — М.: Наука, 1986. 400 с. doi: 10.13140/RG.2.1.2230.6644.
2. Fant K. M. Logically determined design: Clockless system design with NULL convention logic. — Hoboken, NJ, USA: John Wiley & Sons, 2005. 292 p. doi: 10.1002/0471702897.
3. Плеханов Л. П., Степченков Ю. А. Экспериментальная проверка некоторых свойств строго самосинхронных схем // Системы и средства информатики, 2006. Вып. 16. С. 476–485. EDN: KZUWOX.

4. Соколов И. А., Степченко Ю. А., Дьяченко Ю. Г., Рождественский Ю. В. Оценка надежности синхронного и самосинхронного конвейеров // Информатика и её применения, 2022. Т. 16. Вып. 4. С. 2–7. doi: 10.14357/19922264220401. EDN: GWXJHM.
5. Yoshikawa S., Sannomiya S., Iwata M., Nishikawa H. Pipeline stage level simulation method for self-timed data-driven processor on FPGA // 8th Electrical Engineering Congress (International) Proceedings. — Piscataway, NJ, USA: IEEE, 2020. Art. 229515. 5 p. doi: 10.1109/IEECON48109.2020.229515.
6. Chikarenko S. K., Ivanova K. M., Skornyakova A. Y., Tyurin S. F. Self-timed FPGA design perspectives // Conference (International) on Information and Digital Technologies Proceedings. — IEEE, 2021. P. 106–112. doi: 10.1109/IDT52577.2021.9497620.
7. Kushnerov A., Medina M., Yakovlev A. Towards hazard-free multiplexer based implementation of self-timed circuits // 27th IEEE Symposium (International) on Asynchronous Circuits and Systems Proceedings. — Piscataway, NJ, USA: IEEE, 2021. P. 17–24. doi: 10.1109/ASYNC48570.2021.00011.
8. Nautiyal V., Singla G., Maiti B., Kinkade M. Self-timed write aid circuit for tall memories in advanced CMOS technologies // Symposium (International) on Circuits and Systems Proceedings. — IEEE, 2021. Art. 9401420. 4 p. doi: 10.1109/ISCAS51556.2021.9401420.
9. Cortadella J., Kishinevsky M., Kondratyev A., Lavagno L., Yakovlev A. Petri-fy: A tool for manipulating concurrent specifications and synthesis of asynchronous controllers // IEICE T. Inf. Syst., 1997. Vol. E80-D. No. 3. P. 315–325.
10. Edwards D., Bardsley A. Balsa: An asynchronous hardware synthesis language // Comput. J., 2002. Vol. 45. No. 1. P. 12–18. doi: 10.1093/comjnl/45.1.12.
11. Poliakov I., Sokolov D., Mokhov A. Workcraft: A static data flow structure editing, visualisation and analysis tool // Petri nets and other models of concurrency / Eds. J. Kleijn, A. Yakovlev. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer, 2007. Vol. 4546. P. 505–514. doi: 10.1007/978-3-540-73094-1_30.
12. Taubin A., Cortadella J., Lavagno L., Kondratyev A., Peeters A. Design automation of real-life asynchronous devices and systems // Foundations Trends Electronic Design Automation, 2007. Vol. 2. No. 1. P. 1–133. doi: 10.1561/1000000006.
13. Bardsley A., Tarazona L., Edwards D. Teak: A token-flow implementation for the Balsa language // 9th Conference (International) on Application of Concurrency to System Design Proceedings. — IEEE, 2009. P. 23–31. doi: 10.1109/ACSD.2009.15.
14. Tailor R. A., Reese R. B. UNCLE — Unified NCL Environment — an NCL design tool // Asynchronous circuit applications. — Institute of Engineering and Technology, 2019. P. 293–307. doi: 10.1049/PBCS061E.ch14.
15. Sparsø J. Introduction to asynchronous circuit design. — Copenhagen, Denmark: DTU Compute, Technical University of Denmark, 2020. 255 p. https://backend.orbit.dtu.dk/ws/files/215895041/JSPA_async_book_2020_PDF.pdf.
16. Oliveira D., Cardoso N., Batista G. A new method for synthesis of self-timed combinational circuits with strong indication // 5th Ecuador Technical Chapters Meeting Proceedings. — IEEE, 2021. Art. 9590822. 6 p. doi: 10.1109/ETCM53643.2021.9590822.

17. Fiorentino M., Thibeault C., Savaria Y. Introducing KeyRing self-timed microarchitecture and timing-driven design flow // IET Comput. Digit. Tec., 2021. Vol. 15. P. 409–426. doi: 10.1049/cdt2.12032.
18. Andrikos N., Lavagno L., Pandini D., Sotiriou C. P. A fully-automated desynchronization flow for synchronous circuits // 44th Annual Design Automation Conference Proceedings. — New York, NY, USA: Association for Computing Machinery, 2007. P. 982–985. doi: 10.1145/1278480.1278722.
19. Zhou R., Chong K.-S., Gwee B.-H., Chang J. S., Ho W.-G. Synthesis of asynchronous QDI circuits using synchronous coding specifications // Symposium (International) on Circuits and Systems Proceedings. — Piscataway, NJ, USA: IEEE, 2014. P. 153–156. doi: 10.1109/ISCAS.2014.6865088.
20. Chadzynski T. An efficient design methodology for complex sequential asynchronous digital circuits: Master Thesis. — San Jose State University, 2020. 213 p. doi: 10.31979/etd.gy4n-x9sz.
21. Аббясов А. М., Терехов А. И. Синтез последовательностных устройств: методические указания. — Иваново: ИГЭУ, 2022. 24 с. <https://e.lanbook.com/book/369575>.
22. Czekalski P., Tokarz K., Pochopien B. A modern approach to the asynchronous sequential circuit synthesis // Theoretical Applied Informatics, 2014. Vol. 26. No. 1-2. P. 25–37.
23. Andaloussi I., Sedra M. B. Reversible design of asynchronous sequential circuits // Int. J. Engineering Technology, 2020. Vol. 12. No. 2. P. 101–107. doi: 10.21817/ijet/2020/v12i2/201202017.
24. Степченко Ю. А., Степченко Д. Ю., Дьяченко Ю. Г., Морозов Н. В., Плеханов Л. П. Замена синхронных триггеров самосинхронными аналогами в процессе десинхронизации схемы // Системы и средства информатики, 2023. Т. 33. № 4. С. 4–15. doi: 10.14357/08696527230401. EDN: VPLSHI.
25. Степченко Ю. А., Дьяченко Ю. Г., Морозов Н. В., Степченко Д. Ю., Дьяченко Д. Ю. Формализация синтеза самосинхронных счетчиков // Системы и средства информатики, 2024. Т. 34. № 2. С. 66–82. doi: 10.14357/08696527240205. EDN: KDIEOJ.
26. Дьяченко Ю. Г., Плеханов Л. П., Морозов Н. В., Степченко Д. Ю., Орлов Г. А., Дьяченко Д. Ю. Реализация функциональности синхронных триггеров в самосинхронном базисе // Системы и средства информатики, 2025. Т. 35. № 3. С. 3–16.
27. Sokolov I., Stepchenkov Y., Diachenko Y. Synthesis of self-timed circuits with memory // Russian Smart Industry Conference (International) Proceedings. — IEEE, 2024. P. 511–516. doi: 10.1109/SmartIndustryCon61328.2024.10516224.
28. Sudeng S., Thongtak A. Synthesis of complicated asynchronous control circuits using template based technique // World Congress on Engineering Proceedings. — London, U.K.: Newswood Ltd., 2008. Vol. I. P. 462–467.
29. Sakib A. A., Smith S. C., Srinivasan S. K. Formal modeling and verification of PCHB asynchronous circuits // IEEE T.VLSI Syst., 2019. Vol. 27. No. 12. P. 2911–2924. doi: 10.1109/TVLSI.2019.2937087.
30. Ozdag R. O. Template based asynchronous design: PhD Thesis. — Los Angeles, CA, USA: University of Southern California, 2003. 121 p. <https://scispace.com/pdf/template-based-asynchronous-design-2dasq9z4y7.pdf>.
31. Степченко Ю. А., Хилько Д. В., Дьяченко Ю. Г., Морозов Н. В., Степченко Д. Ю., Орлов Г. А. Методика десинхронизации при синтезе самосинхронных

схем // Системы и средства информатики, 2024. Т. 34. № 1. С. 33–43. doi: 10.14357/08696527240103. EDN: XGZCWU.

32. Yosys Open Synthesis Suite. <https://yosyshq.net/yosys>.

33. Плеханов Л. П. Оптимизация синтеза самосинхронных схем // Российский форум «Микроэлектроника-2023». 9-я Научная конференция «Электронная компонентная база и микроэлектронные модули»: Сборник тезисов. — М: Техносфера, 2023. 5 с.

Поступила в редакцию 05.06.2025

Принята к публикации 15.10.2025

TEMPLATE METHOD IN SYNTHESIS OF SELF-TIMED DIGITAL CIRCUITS

L. P. Plekhanov, Yu. G. Diachenko, D. V. Khilko, and G. A. Orlov

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The article considers the problem of self-timed (ST) digital circuit design automation. Self-timed circuits are an alternative to the synchronous ones. In spite of significant advantages, especially in terms of operational reliability in a wide range of operating conditions under the influence of unfavorable factors, ST circuits have not yet found wide application. In part, this is due to the complexity of their design which requires a specific approach and consideration of the ST circuit’s functioning discipline features. The greatest difficulty is the formalization and automation of sequential ST unit synthesis. For this purpose, it is proposed to use the template method. It includes an analysis of the synchronous counterpart’s original description of the synthesized ST circuit using Yosys, the open-source logical synthesizer of synchronous circuits, searching for fragments implemented by units with memory, and replacing them with pre-prepared templates, namely, ST Verilog descriptions of sequential units adequate to the prototype in terms of operational features. The templates contain the synchronous and ST implementations of the corresponding units. The article provides template examples and describes the method of their application in the process of converting the original synchronous description of the synthesized circuit into an ST Verilog description. Substituting templates into the synthesized circuit description eliminates the need for their individual synthesis taking into account the specifics of the ST circuits. The proposed approach ensures minimal hardware costs and optimal performance and guarantees the ST nature of the resulting circuit implementations of digital units.

Keywords: self-timed circuits; automated logic synthesis; template; sequential circuits; conversion; Verilog

DOI: 10.14357/08696527250401

EDN: THWOVK

Acknowledgments

The research was financially supported by the grant from the Russian Science Foundation (project No. 25-19-00508).

References

1. Varshavskiy, V. I., V. A. Kishinevskiy, V. B. Marakhovskiy, *et al.* 1986. *Avtomatnoe upravlenie asinkhronnymi protsessami v EVM i diskretnykh sistemakh* [Automata control of asynchronous processes in computers and discrete systems]. Ed. V. I. Varshavskiy. Moscow: Nauka. 400 p. doi: 10.13140/RG.2.1.2230.6644.
2. Fant, K. M. 2005. *Logically determined design: Clockless system design with NULL convention logic*. Hoboken, NJ: John Wiley & Sons. 292 p. doi: 10.1002/0471702897.
3. Plekhanov, L. P., and Yu. A. Stepchenkov. 2006. Eksperimental'naya proverka nekotorykh svoystv strogo samosinkhronnykh skhem [Experimental verification of some properties of strictly self-timed circuits]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 16:476–485. EDN: KZUWOX.
4. Sokolov, I. A., Yu. A. Stepchenkov, Yu. G. Diachenko, and Yu. V. Rogdestvenski. 2022. Otsenka nadezhnosti sinkhronnogo i samosinkhronnogo konveyerov [Synchronous and self-timed pipeline's reliability estimation]. *Informatika i ee Primeneniya — Inform. Appl.* 16(4):2–7. doi: 10.14357/19922264220401. EDN: GWXJHM.
5. Yoshikawa, S., S. Sannomiya, M. Iwata, and H. Nishikawa. 2020. Pipeline stage level simulation method for self-timed data-driven processor on FPGA. *8th Electrical Engineering Congress (International) Proceedings*. Piscataway, NJ: IEEE. Art. 229515. 5 p. doi: 10.1109/IEECON48109.2020.229515.
6. Chikarenko, S. K., K. M. Ivanova, A. Y. Skornyakova, and S. F. Tyurin. 2021. Self-timed FPGA design perspectives. *Conference (International) on Information and Digital Technologies Proceedings*. IEEE. 106–112. doi: 10.1109/IDT52577.2021.9497620.
7. Kushnerov, A., M. Medina, and A. Yakovlev. 2021. Towards hazard-free multiplexer based implementation of self-timed circuits. *27th Symposium (International) on Asynchronous Circuits and Systems Proceedings*. Piscataway, NJ: IEEE. 17–24. doi: 10.1109/ASYNC48570.2021.00011.
8. Nautiyal, V., G. Singla, B. Maiti, and M. Kinkade. 2021. Self-timed write aid circuit for tall memories in advanced CMOS technologies. *IEEE Symposium (International) on Circuits and Systems Proceedings*. IEEE. Art. 9401420. 4 p. doi: 10.1109/ISCAS51556.2021.9401420.
9. Cortadella, J., M. Kishinevsky, A. Kondratyev, L. Lavagno, and A. Yakovlev. 1997. Petrify: A tool for manipulating concurrent specifications and synthesis of asynchronous controllers. *IEICE T. Inf. Syst.* E80-D(3):315–325.
10. Edwards, D., and A. Bardsley. 2002. Balsa: An asynchronous hardware synthesis language. *Comput. J.* 45(1):12–18. doi: 10.1093/comjnl/45.1.12.
11. Poliakov, I., D. Sokolov, and A. Mokhov. 2007. Workcraft: A static data flow structure editing, visualisation and analysis tool. *Petri nets and other models of concurrency*. Eds. J. Kleijn and A. Yakovlev. Lecture notes in computer science ser. Berlin, Heidelberg: Springer. 4546:505–514. doi: 10.1007/978-3-540-73094-1_30.

12. Taubin, A., J. Cortadella, L. Lavagno, A. Kondratyev, and A. Peeters. 2007. Design automation of real-life asynchronous devices and systems. *Foundations Trends Electronic Design Automation* 2(1):1–133. doi: 10.1561/1000000006.
13. Bardsley, A., L. Tarazona, and D. Edwards. 2009. Teak: A token-flow implementation for the Balsa language. *9th Conference (International) on Application of Concurrency to System Design Proceedings*. IEEE. 23–31. doi: 10.1109/ACSD.2009.15.
14. Tailor, R. A., and R. B. Reese. 2019. UNCLE — Unified NCL Environment — an NCL design tool. *Asynchronous circuit applications*. Institute of Engineering and Technology. 293–307. doi: 10.1049/PBCS061E_ch14.
15. Sparsø, J. 2020. *Introduction to asynchronous circuit design*. Copenhagen, Denmark: DTU Compute, Technical University of Denmark. 269 p. Available at: https://orbit.dtu.dk/files/215895041/JSPA_async_book_2020_PDF.pdf (accessed October 13, 2025).
16. Oliveira, D. L., N. N. M. Cardoso, and G. C. Batista. 2021. A new method for synthesis of self-timed combinational circuits with strong indication. *5th Ecuador Technical Chapters Meeting Proceedings*. IEEE. Art. 9590822. 6 p. doi: 10.1109/ETCM53643.2021.9590822.
17. Fiorentino, M., C. Thibeault, and Y. Savaria. 2021. Introducing KeyRing self-timed microarchitecture and timing-driven design flow. *IET Comput. Digit. Tec.* 15:409–426. doi: 10.1049/cdt.2.12032.
18. Andrikos, N., L. Lavagno, D. Pandini, and C. P. Sotiriou. 2007. A fully-automated desynchronization flow for synchronous circuits. *44th Annual Design Automation Conference Proceedings*. New York, NY: Association for Computing Machinery. 982–985. doi: 10.1145/1278480.1278722.
19. Zhou, R., K.-S. Chong, B.-H. Gwee, J. S. Chang, and W.-G. Ho. 2014. Synthesis of asynchronous QDI circuits using synchronous coding specifications. *Symposium (International) on Circuits and Systems Proceedings*. Piscataway, NJ: IEEE. 153–156. doi: 10.1109/ISCAS.2014.6865088.
20. Chadzynski, T. 2020. An efficient design methodology for complex sequential asynchronous digital circuits. San Jose State University. Master Thesis. 213 p. doi: 10.31979/etd.gy4n-x9sz.
21. Abbyasov, A. M., and A. I. Terekhov. 2022. *Sintez posledovatel’nostnykh ustroystv: metodicheskie ukazaniya* [Synthesis of sequential devices: Methodological guidelines]. Ivanovo: IGEU. 24 p.
22. Czekalski, P., K. Tokarz, and B. Pochopien. 2014. A modern approach to the asynchronous sequential circuit synthesis. *Theoretical Applied Informatics* 26(1-2):25–37.
23. Andaloussi, I., and M. B. Sedra. 2020. Reversible design of asynchronous sequential circuits. *Int. J. Engineering Technology* 12(2):101–107. doi: 10.21817/ijet/2020/v12i2/201202017.
24. Stepchenkov, Yu. A., D. Yu. Stepchenkov, Yu. G. Diachenko, N. V. Morozov, and L. P. Plekhanov. 2023. Zamena sinkhronnykh triggerov samosinkhronnymi analogami v protsesse desinkhronizatsii skhemy [Replacing synchronous triggers with self-timed counterparts during circuit desynchronization]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(4):4–15. doi: 10.14357/08696527230401. EDN: VPLSHI.

25. Stepchenkov, Yu. A., Yu. G. Diachenko, N. V. Morozov, D. Yu. Stepchenkov, and D. Yu. Diachenko. 2024. Formalizatsiya sinteza samosinkhronnykh schetchikov [Self-timed counter synthesis formalization]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 34(2):66–82. doi: 10.14357/08696527240205. EDN: KDIEOJ.
26. Diachenko, Yu. G., L. P. Plekhanov, N. V. Morozov, D. Yu. Stepchenkov, G. A. Orlov, and D. Yu. Diachenko. 2025. Realizatsiya funktsional'nosti sinkhronnykh triggerov v samosinkhronnom baze [Implementation of synchronous flip-flop and latch functionality in a self-timed basis]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 35(3):3–16.
27. Sokolov, I., Y. Stepchenkov, and Y. Diachenko. 2024. Synthesis of self-timed circuits with memory. *Russian Smart Industry Conference (International) Proceedings*. IEEE. 511–516. doi: 10.1109/SmartIndustryCon61328.2024.10516224.
28. Sudeng, S., and A. Thongtak. 2008. Synthesis of complicated asynchronous control circuits using template based technique. *World Congress on Engineering Proceedings*. London, U.K.: Newswood Ltd. I:462–467.
29. Sakib, A. A., S. C. Smith, and S. K. Srinivasan. 2019. Formal modeling and verification of PCHB asynchronous circuits. *IEEE T. VLSI Syst.* 27(12):2911–2924. doi: 10.1109/TVLSI.2019.2937087.
30. Ozdag, R. O. 2003. Template based asynchronous design. Los Angeles, CA: University of Southern California. PhD Thesis. 121 p. Available at: <https://scispace.com/pdf/template-based-asynchronous-design-2dasq9z4y7.pdf> (accessed October 13, 2025).
31. Stepchenkov, Yu. A., D. V. Khilko, Yu. G. Diachenko, N. V. Morozov, D. Yu. Stepchenkov, and G. A. Orlov. 2024. Metodika desinkhronizatsii pri sinteze samosinkhronnykh skhem [Desynchronization methodology at self-timed circuit synthesis]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 34(1):33–43. doi: 10.14357/08696527240103. EDN: XGZCWU.
32. Yosys open synthesis suite. Available at: <https://yosyshq.net/yosys> (accessed October 13, 2025).
33. Plekhanov, L. P. 2023. Optimizatsiya sinteza samosinkhronnykh skhem [Optimization of the synthesis of self-timed circuits]. *Rossiyskiy forum «Mikroelektronika-2023». 9-ya Nauchnaya konferentsiya «Elektronnaya komponentnaya baza i mikroelektronnye moduli»: Sbornik tezisev* [Russian Forum “Microelectronics 2023,” 9th Scientific Conference “Electronic Component Base and Microelectronic Modules” Abstracts]. Moscow: Tekhnosfera. 5 p.

Received June 5, 2025

Accepted October 15, 2025

Contributors

Plekhanov Leonid P. (b. 1943)—Candidate of Science (PhD) in technology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; lplekhanov@inbox.ru

Diachenko Yuri G. (b. 1958)—Candidate of Science (PhD) in technology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian

Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; diaura@mail.ru

Khilko Dmitri V. (b. 1987) — Candidate of Science (PhD) in technology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; dhilko@yandex.ru

Orlov Georgii A. (b. 1994) — junior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; orlov.jaja@gmail.com

О ПРИМЕНЕНИИ ДРЕВОВИДНЫХ АЛГОРИТМОВ РАЗРЕШЕНИЯ КОНФЛИКТОВ В ПРОЦЕДУРЕ СЛУЧАЙНОГО МНОЖЕСТВЕННОГО ДОСТУПА ДЛЯ УСТРОЙСТВ 5G NB-IoT*

В. А. Бесчастный¹, А. М. Тюрликов², Н. В. Степанов³, В. С. Шоргин⁴

Аннотация: Современные IoT-приложения характеризуются более строгими требованиями к задержке, чем те, что определены для 5G Massive Machine-Type Communications в ITU-R M.2410, например менее 10 мс. Для этого консорциум 3GPP (3rd Generation Partnership Project) определил несколько процедур для снижения задержки доступа в системе 5G NB-IoT (Narrowband Internet-of-Things). Однако эти процедуры чисто ALOHA-типа и не используют информацию, которая может быть доступна на базовой станции (БС) NB-IoT. В данной работе рассмотрены два расширения процедуры случайного доступа на основе древовидных алгоритмов разрешения конфликтов. Полученные результаты показывают, что блокируемый алгоритм характеризуется большей на 25%–35% пропускной способностью по сравнению с блокируемым алгоритмом. При этом с точки зрения задержки рассматриваемые алгоритмы могут быть эффективно использованы, когда предлагаемая нагрузка не превышает 70%–80% от максимальной пропускной способности.

Ключевые слова: 5G; NB-IoT (Narrowband Internet-of-Things); канал случайного доступа (RACH); задержка; древовидный алгоритм разрешения конфликтов

DOI: 10.14357/08696527250402

EDN: TXPHCY

1 Введение

В настоящее время рынок приложений Интернета вещей (Internet of things, IoT) развивается беспрецедентными темпами. Это обусловлено не только ростом числа потребительских товаров, оснащенных беспроводными модулями, но

* Исследование выполнено за счет гранта Российского научного фонда № 25-79-10142, <https://rscf.ru/project/25-79-10142/> (исполнитель В. А. Бесчастный, разд. 1, 2, 5), а также при финансовой поддержке Министерства науки и высшего образования Российской Федерации, соглашение № FSRF-2023-0003, «Фундаментальные основы построения помехозащищенных систем космической и спутниковой связи, относительной навигации, технического зрения и аэрокосмического мониторинга» (получатели А. М. Тюрликов и Н. В. Степанов, разд. 3 и 4).

¹ Российский университет дружбы народов им. Патриса Лумумбы, beschastnyy-va@rudn.ru

² Санкт-Петербургский государственный университет аэрокосмического приборостроения, turlikov@guap.ru

³ Санкт-Петербургский государственный университет аэрокосмического приборостроения, stepanov.nikita@guap.ru

⁴ Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vshorgin@ipiran.ru

и появлением новых приложений, связанных с различными отраслями промышленности. В настоящее время NB-IoT стал де-факто стандартом для технологий сотового IoT (CIoT), недавно принятым ITU-R в качестве основного для сервисов массовых межмашинных коммуникаций (Massive Machine Type Communications, mMTC).

Современный рынок Интернета вещей характеризуется появлением новых приложений, которые занимают промежуточное положение между сервисами со сверхнадежной связью и сверхнизкой задержкой (Ultra Reliable and Low Latency Communications, URLLC) и mMTC. Такие сценарии использования, как промышленная автоматизация на производстве, связь между беспилотными летательными аппаратами (БПЛА), мониторинг энергетических сетей, отличаются более строгими требованиями к задержке, более высокой надежностью и более высокой пропускной способностью, чем те, что определены для mMTC [1]. Хотя эти требования все еще не так строги, как для URLLC, существующие решения CIoT, такие как LTE-M (Long-Term Evolution for Machines) и NB-IoT, не соответствуют требованиям этих сервисов по задержке и надежности.

В ответ на эти новые требования 3GPP в недавних стандартах предложил несколько усовершенствований для связи с пониженной задержкой, особенно в части процедур канала случайного доступа (RACH, Random Access Channel). В частности, в Release 16 была предложена двухшаговая процедура, позволяющая одновременно передавать преамбулу доступа и данные в одном кадре, гарантируя, что передача данных будет завершена за два кадра. Однако, как было показано в [2], такой подход может приводить к субоптимальной производительности в условиях высокой загрузки сети.

Чтобы решить указанную проблему двухшаговой процедуры RACH в NB-IoT, в данной работе рассматриваются варианты расширения метода доступа без резервирования ресурсов в NB-IoT на основе древовидных алгоритмов разрешения конфликтов. Предложенный подход не изменяет саму двухшаговую процедуру случайного доступа и требует минимальных модификаций текущей версии стандарта NB-IoT.

2 Методы случайного доступа в NB-IoT

Процедура случайного доступа в NB-IoT — это способ, которым абонентское устройство (АУ) устанавливает связь с БС. В ее основе лежит четырехшаговый процесс: устройство отправляет короткий сигнал (преамбулу), БС отвечает, предоставляя ресурсы для передачи следующего сообщения. Затем устройство отправляет свое сообщение, а БС подтверждает успешное завершение процедуры. Основная проблема — коллизии (наложение сообщений от разных устройств), возникающие, когда несколько устройств передают одну и ту же преамбулу одновременно. Если имеется возможность динамически переопределять частотно-временные ресурсы, которые выделяются для передачи преамбул

и сообщений, то можно за счет такого перераспределения снизить вероятность коллизий и уменьшить задержку [3].

Хотя вышеупомянутая традиционная процедура доступа зарекомендовала себя как надежная, она иногда приводит к ненужным задержкам на радиointерфейсе. Недавно 3GPP предложила так называемое двухшаговое случайное предоставление доступа без рукопожатия [4]. Согласно этой процедуре преамбулы теперь жестко связаны с возможностями передачи в канале PUSCH (Physical Uplink Shared Channel), а описанная 4-шаговая процедура сокращена всего до двух фаз. На первой фазе *сообщение 1* и данные передаются по каналам NPRACH (Narrowband Physical RACH) и PUSCH. Их успешная доставка подтверждается с помощью *сообщения 2*. Это изменение позволяет значительно сократить задержку доставки информации в сетях с хорошо спроектированной конфигурацией [2, 5]. Однако теперь процедура подвержена потенциальным коллизиям между пакетами данных в канале PUSCH даже в том случае, когда разные устройства выбрали разные преамбулы и преамбулы были правильно обнаружены. Чтобы еще больше повысить производительность этого подхода, был предложен вариант возврата к традиционной 4-шаговой процедуре. Соответственно, если преамбулы были правильно обнаружены, но пакеты данных не могут быть распознаны из-за коллизии, система переходит к *сообщениям 3 и 4* 4-шаговой процедуры. Тем не менее это приводит к дополнительным задержкам, связанным с процедурой возврата и выделением ресурсов в PUSCH.

3 Древовидные алгоритмы разрешения конфликтов

Древовидные алгоритмы разрешения конфликтов были предложены более 40 лет назад в работах [6, 7]. В этих работах была введена модель системы случайного множественного доступа (СМД) с потенциально неограниченным числом АУ. Основные особенности данной модели состоят в том, что процесс появления абонентов в системе описывается однородным пуассоновским потоком. У абонента имеется одно-единственное сообщение, время передачи принято за единицу времени, все время передачи по каналу разделено на слоты единичной длины, а моменты разделения окон известны абонентам, и абонент может начинать передачу сообщения только в начале слота. Если в одном слоте передают два или более абонентов, то ни одно из сообщений этих абонентов не может быть успешно принято, возникает так называемый конфликт. Алгоритмом случайного множественного доступа называется правило, в соответствии с которым абоненты, наблюдая выход канала, принимают решение, передавать сообщение или нет. В работах [6, 7] это правило представлялось в виде двоичного дерева разрешения конфликта. Корень и промежуточные вершины этого дерева соответствовали слотам с ситуацией «конфликт», а концевые вершины соответствовали окнам с ситуациями «пусто» или «успех». Все абоненты, наблюдая выход канала, строят дерево разрешения конфликта. Абоненты, передававшие в слоте, который соответствует корневой вершине дерева, в процессе построения дерева выбирают

слоты, в которых будут делать повторную передачу. Все остальные абоненты также строят дерево разрешения конфликта, но откладывают (блокируют) свою передачу до тех пор, пока конфликт не будет разрешен, т. е. дерево разрешения конфликта построено. Предложенные в работах [7, 8] алгоритмы получили название древовидных алгоритмов СМД и в рамках этой модели обеспечивали устойчивую работу при интенсивности входного потока порядка 0,375 сообщений в единицу времени. В дальнейшем древовидные алгоритмы развивались в трех направлениях:

- (1) создание альтернативного описания древовидных алгоритмов в форме стека и появление версии алгоритма, в которой абоненты сразу могут передавать в канал (алгоритм свободного доступа, СД);
- (2) усовершенствование способа построения дерева (или эквивалентного правила работы со стеком) с целью повышения предельной интенсивности входного потока, до которой алгоритм может обеспечить устойчивую работу системы в рамках модели из [8]. В русскоязычных работах эта предельная интенсивность получила название «скорость алгоритма», используется и прямой перевод с английского — «пропускная способность алгоритма» (the throughput of an algorithm). В работе [7] была предложена модификация древовидного алгоритма, так называемый алгоритм дробления (part-and-try algorithm), который имеет скорость 0,4877. До настоящего времени алгоритмов, которые в рамках модели из [8] имеют большую скорость, не найдено;
- (3) изменение допущений модели из [8] с учетом свойств физического уровня и усовершенствование способа построения дерева (или эквивалентного правила работы со стеком) с учетом особенностей видоизмененной модели.

С развитием концепции mMTC в настоящее время появилось достаточно большое число работ, связанных именно с третьим направлением развития древовидных алгоритмов разрешения конфликтов. При этом, по мнению авторов статьи, в настоящее время с учетом современных подходов к построению физического уровня можно использовать подход, предложенный более 40 лет назад в работе [9]. Далее кратко опишем этот подход.

В работе [9] в модель из [8] вносится следующее изменение. В слоте перед сообщением абонент передает служебную двоичную последовательность из H бит, среди которых $(H - 1)$ бит имеют значение 0 и один случайно выбранный абонентом при каждой передаче бит равен 1. При этом предполагается, что в канале имеется идеальная побитовая синхронизация и при наложении сообщений от разных абонентов двоичные последовательности дизъюнктивно складываются. С учетом видоизмененной модели дерево разрешения конфликта перестает быть двоичным. Теперь у каждой вершины, в которой произошел конфликт, столько потомков, сколько единиц в служебной последовательности.

В [9] показывается, что $\lambda_{\text{СД}}(H)$ — пропускная способность алгоритма случайного доступа — является корнем уравнения

$$1 = \frac{\psi}{1 - \psi} e^{-\psi/H} \sum_{m=0}^{\infty} e^{\psi/H^m} \left\{ (1 - H^{-1}) (1 - e^{-\lambda/H^m}) + (H^m - 1) \left(1 - \frac{\lambda}{H^m} - e^{-\lambda/H^m} \right) + \frac{1 - \psi}{\psi} H^m \left(\frac{\lambda}{H^m} + H (e^{-\lambda/H^{m+1}} - 1) \right) \right\},$$

где $\psi \triangleq \lambda/(1 - H^{-1})$. Также отмечается, что $\lambda_{\text{СД}}(H) \rightarrow \lambda_0$ при $H \rightarrow \infty$, где $\lambda_0 = 0,567$ — корень уравнения $\lambda = e^{-\lambda}$.

Также в работе [9] приводится приближенное выражение для расчета пропускной способности алгоритма блокируемого доступа (БД):

$$\lambda_{\text{БД}}(H) \approx \ln H (1 + \ln H)^{-1}.$$

Из анализа алгоритмов следует, что для блокируемого алгоритма пропускную способность с увеличением значения H можно сделать сколь угодно близкой к единице. При этом реализация алгоритмов не сложнее, чем реализация других древовидных алгоритмов разрешения конфликтов.

Однако предположение о том, что в канале имеется идеальная побитовая синхронизация и при наложении сообщений от разных абонентов двоичные последовательности дизъюнктивно складываются, не реалистично для беспроводных систем, и, возможно, поэтому на работу [9] не было обращено внимание.

Данные алгоритмы могут быть использованы в рамках 2-шаговой процедуры случайного множественного доступа, которая применяется в сетях 5-го поколения. При этом вместо двоичных последовательностей с идеальной побитовой синхронизацией будут использоваться преамбулы (последовательности Задова–Чу), применяемые в современных сотовых сетях. Тогда число таких преамбул будет равняться числу бит в двоичной последовательности, рассматриваемой в работе [9]. Таким образом, для алгоритма свободного доступа, увеличивая число преамбул, можно достичь пропускной способности $\lambda_0 = 0,567$, а для алгоритма БД пропускная способность может быть достигнута сколь угодно близкой к единице.

4 Предлагаемый подход и модель системы

4.1 Модифицированные алгоритмы доступа для NB-IoT

Основная проблема доступа без резервирования ресурсов заключается в том, что без явных сообщений Message 3 (идентификатор оконечного устройства) и Message 4 (разрешение коллизий) обработка коллизий становится затруднительной. Если несколько устройств передают данные, их преамбулы и пакеты

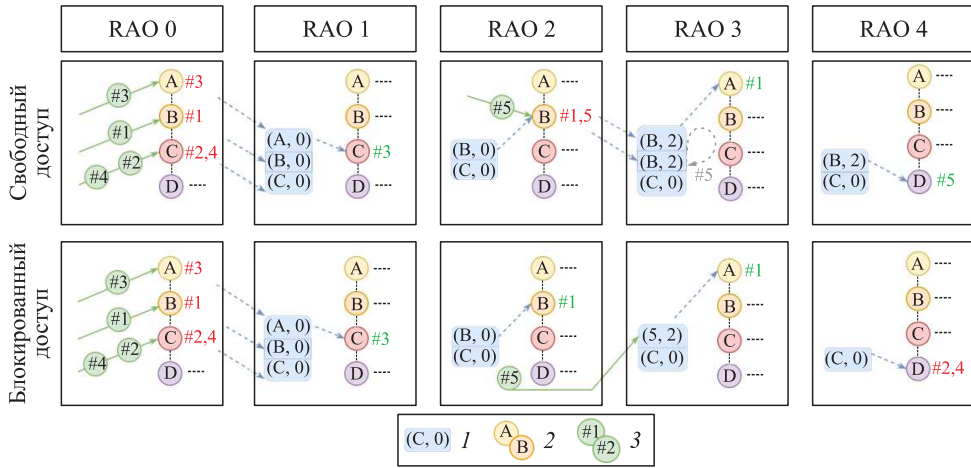


Рис. 1 Иллюстрация применения древовидных алгоритмов разрешения конфликтов с СД и БД для двухшаговой процедуры случайного доступа в NB-IoT: 1 — стек; 2 — преамбулы; 3 — поступающие заявки

данных накладываются друг на друга и БС может не суметь их различить. Предложенные древовидные алгоритмы разрешения конфликтов можно интегрировать в процедуру случайного доступа NB-IoT. Рассмотрим процедуру на примере СД.

На стороне АУ при попытке доступа оно сохраняет как номер преамбулы, так и номер кадра, в котором она была отправлена. Если ни одно другое АУ не выбрало ту же преамбулу и БС успешно декодирует данные в выделенном канале PUSCH, она отправляет АУ *сообщение В*, и АУ покидает систему.

В противном случае, если несколько АУ выбрали одну и ту же преамбулу, на БС инициируется древовидный алгоритм. Рассмотрим стек, показанный на рис. 1, где самая верхняя ячейка имеет индекс 0. Изначально стек пуст. Если $n > 1$ и АУ одновременно выбрали одну и ту же преамбулу, ячейки стека начиная с индекса 1 сдвигаются вниз на $n - 1$ позиций. Нулевая ячейка и новые $n - 1$ освободившихся ячеек заполняются n номерами полученных преамбул и номером кадра. Если только одно АУ выбрало преамбулу, ячейки стека начиная с индекса 1 сдвигаются вниз на одну позицию. Нулевая ячейка и новая освободившаяся ячейка с индексом 1 заполняются тем же номером полученной преамбулы и номером слота.

Затем канал NPSSCH используется для информирования АУ (1) о неудаче передачи, (2) номере кадра, (3) номере преамбулы, (4) общем числе различных преамбул n , переданных на БС. На стороне АУ при получении этого сообщения, если $n > 1$, АУ повторяет попытку доступа.

В случае если данные успешно приняты, нулевая ячейка стека очищается, и все ячейки начиная с индекса 1 сдвигаются вверх на одну позицию. Информация

об успешном приеме данных и сами данные из нулевой ячейки стека передаются по каналу NRPSCH.

При этом отличие протокола БД состоит в том, что новые АУ при первой попытке доступа отправляют преамбулу в текущем слоте только в случае пустого стека, а в случае непустого стека блокируются и помещаются на его вершину, т. е. передают преамбулы уже в последующих слотах. Таким образом сокращается вероятность возникновения конфликтов между вновь пришедшими и уже находящимися в системе АУ.

4.2 Система предположений

Для анализа предложенных алгоритмов были приняты следующие допущения. Во-первых, учитывая специфику сервиса mMTC, предполагается, что суммарный процесс поступления запросов в систему описывается процессом Пуассона с интенсивностью λ . Единицей времени (слотом) в рассматриваемой системе служит подкадр длительностью 1 мс, и каждый подкадр включает возможность случайного доступа (Random Access Occasion, RAO). Это соответствует индексу конфигурации PRACH 27 [10]. Этот индекс также определяет использование формата преамбулы 0 на радиointерфейсе. Однако стоит отметить, что предложенный подход может быть использован и для других конфигурационных индексов, где только ограниченное число подкадров может содержать возможности RAO.

В работе также приняты следующие допущения относительно декодирования преамбул и данных на БС:

- (1) когда один пользователь передает преамбулу и данные, они оба принимаются корректно;
- (2) когда два или более пользователей передают преамбулу и данные, пакеты данных всегда принимаются с ошибкой, при этом БС распознает тип переданной преамбулы.

Важно, что после возникновения коллизии дополнительные пользователи могут активироваться во время фазы ее разрешения и присоединиться к процессу.

4.3 Имитационная модель

Для расчета задержек в системе в работе использованы методы имитационного моделирования. Программа расчета была реализована на языке программирования Python. В рамках моделирования рассматривалась только процедура случайного доступа, но с учетом временных ограничений, накладываемых протоколом NB-IoT [11].

Поскольку для интенсивностей, не превышающих теоретической границы, возможно достижение стационарного режима, его начало определяется с помощью метода экспоненциально взвешенного скользящего среднего с параметром сглаживания 0,01. Сбор данных осуществлялся в стационарном режиме

порционно по 1000 слотов наблюдения, после чего для каждого интервала вычислялись усредненные характеристики, что позволяло избежать возникновения эффекта остаточной корреляции [12]. Согласно использованному методу, каждая полученная выборка делится на партии, каждая из которых содержит по 1000 наблюдений. Среднее значение каждой партии статистически не зависит от средних значений других партий, и, таким образом, выборка средних значений может быть обработана с использованием стандартных статистических методов. При этом в каждом эксперименте набирается достаточное число партий, чтобы доверительные интервалы не отклонялись от точечных оценок более чем на 1% при установленном уровне значимости $\alpha = 0,05$.

Стоит отметить, что по мере приближения к предельной интенсивности достижение стационарного режима перестает быть возможным. По этой причине было введено ограничение на размер стека в 10 000 заявок, при достижении которого интенсивность входящего потока считалась предельной. Таким образом были получены оценки сверху на пропускную способность системы.

5 Численный анализ

В данном разделе проводится численный анализ предлагаемых подходов к улучшению метода доступа без резервирования ресурсов в NB-IoT. Рассмотрены неблокируемый и блокируемый алгоритмы.

На рис. 2 представлено сравнение пропускной способности алгоритмов БД и СД для системы NB-IoT как функции от числа преамбул. Теоретические результаты получены с помощью модели, предложенной в [9]. В то же время имитационная модель реализует 2-шаговую процедуру с использованием рассматриваемых алгоритмов разрешения конфликтов. Анализируя полученные данные, отметим, что теоретические результаты, представленные в разд. 2, подтверждаются имитационным моделированием. При этом для СД наблюдается

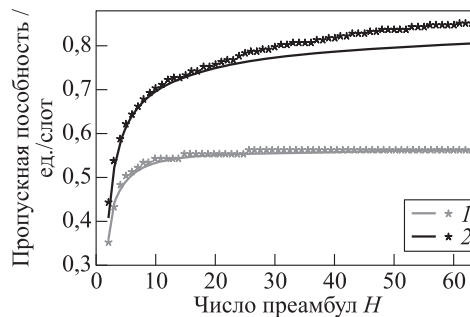


Рис. 2 Сравнение пропускной способности алгоритмов СД (1) и БД (2): кривые — теоретические результаты; значки — имитационная модель

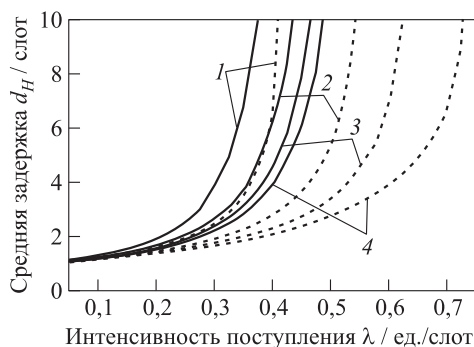


Рис. 3 Зависимость средней задержки алгоритмов СД (сплошные кривые) и БД (пунктирные кривые) от интенсивности поступления заявок: 1 — $H = 2$; 2 — 4; 3 — 8; 4 — $H = 32$

полное соответствие, тогда как для БД наблюдается расхождение теоретической модели начиная с $H = 20$. При этом отметим, что древовидный алгоритм с БД значительно превосходит алгоритм с СД. Разница превышает 25% уже при $H = 10$ и далее увеличивается до 35% при $H > 40$. Однако реализация БД требует дополнительной сигнальной информации в нисходящем канале связи, что не поддерживается на текущем этапе стандартизации (серия документов 3GPP 18) систем связи NB-IoT.

На рис. 3 представлено сравнение средней задержки алгоритмов БД и СД для системы NB-IoT как функции от интенсивности поступления заявок для различного числа преамбул. Представленная зависимость отражает разницу в задержке рассматриваемых алгоритмов. Отметим, что задержка демонстрирует экспоненциальный рост, и фактически каждый из рассматриваемых алгоритмов может быть эффективно использован, когда предлагаемая нагрузка не превышает 70%–80% от максимальной пропускной способности.

6 Заключение

Рассмотрены варианты расширения метода доступа без резервирования ресурсов в NB-IoT на основе древовидных алгоритмов разрешения конфликтов. Рассмотрены два варианта реализации древовидных алгоритмов — с СД и БД.

Сравнительный анализ показал, что пропускная способность алгоритма с БД на 25%–35% выше, чем у СД. Что касается задержки, оба подхода эффективно работают при нагрузке, не превышающей 70%–80% от максимальной пропускной способности системы.

Литература

1. ITU-R. Minimum requirements related to technical performance for IMT-2020 radio interface. Report M.2410-0, 2017. <https://www.itu.int/pub/R-REP-M.2410-2017>.

2. *Peralta E., Levanen T., Frederiksen F., Valkama M.* Two-step random access in 5G New radio: Channel structure design and performance // 93rd Vehicular Technology Conference Proceedings. — IEEE, 2021. Art. 9449057. 7 p. doi: 10.1109/VTC2021-Spring51267.2021.9449057.
3. *Stepanov N., Turlikov A., Begishev V.* Balancing the data transmission and random access phases in 6G MMTC radio technologies // IEEE Commun. Lett., 2023. Vol. 27. Iss. 12. P. 3419–3423. doi: 10.1109/LCOMM.2023.3322598.
4. *Holma H., Toskala A.* Industrial internet of things // 5G technology: 3GPP evolution to 5G-advanced / Eds. H. Holma, A. Toskala, T. Nakamura. — 2nd ed. — Wiley, 2024. P. 521–545. doi: 10.1002/9781119816058.ch18.
5. *Seo J.-B., Toor W.T., Jin H.* Analysis of two-step random access procedure for cellular ultra-reliable low latency communications // IEEE Access, 2021. Vol. 9. P. 5972–5985. doi: 10.1109/ACCESS.2020.3048824.
6. *Capetanakis J.* Tree algorithms for packet broadcast channels // IEEE T. Inform. Theory, 1979. Vol. 25. Iss. 5. P. 505–515. doi: 10.1109/TIT.1979.1056093.
7. *Цыбаков Б. С., Михайлов В. А.* Случайный множественный доступ пакетов. Алгоритм дробления // Пробл. передачи информ., 1980. Т. 16. № 4. С. 65–79.
8. *Цыбаков Б. С., Михайлов В. А.* Свободный синхронный доступ пакетов в широко-вещательный канал с обратной связью // Пробл. передачи информ., 1978. Т. 14. № 4. С. 32–59.
9. *Керекеш И., Михайлов В. А.* Стек-алгоритм СМД в канал с посимвольной синхронизацией // 6-й Международный симпозиум по теории информации: Тезисы докладов. — Москва–Ташкент, 1984. Т. 1. С. 80–82.
10. 3GPP TR 38.211. NR; Physical channels and modulation, 2025. Release 19.0.0. https://www.3gpp.org/ftp/Specs/archive/38_series/38.211/38211-j00.zip.
11. 3GPP TR 36.763. Study on narrow-band internet of things (NB-IoT)/enhanced machine type communication (eMTC) support for non-terrestrial networks (NTN), 2021. Release 17.0.0. https://www.3gpp.org/ftp/Specs/archive/36_series/36.763/36763-h00.zip.
12. *Mezzavilla M., Zhang M., Polese M., Ford R., Dutta S., Rangan S., Zorzi M.* End-to-end simulation of 5G mmWave networks // IEEE Commun. Surv. Tut., 2018. Vol. 20. Iss. 3. P. 2237–2263. doi: 10.1109/COMST.2018.2828880.

Поступила в редакцию 22.08.2025

Принята к публикации 15.10.2025

TREE-SPLITTING ALGORITHMS FOR MASSIVE RANDOM ACCESS OF 5G NB-IoT COMMUNICATIONS

V. A. Beschastnyi¹, A. M. Turlikov², N. V. Stepanov², and V. S. Shorgin³

¹Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

²Saint Petersburg State University of Aerospace Instrumentation, 67 Bolshaya Morskaya Str., Saint Petersburg 190000, Russian Federation

³Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: State-of-the-art Internet of Things (IoT) applications are characterized by exclusive requirements on latency which are much higher than those defined for 5G Massive Machine-Type Communications in ITU-R M.2410 (e. g., less than 10 ms). For this reason, 3GPP (3rd Generation Partnership Project) has defined several procedures to reduce access latency in the 5G NB-IoT systems. However, these procedures are purely ALOHA-type and do not use information that may be available on the NB-IoT base station. In this paper, two extensions of the random access procedure based on tree-splitting algorithms have been considered. The results show that the algorithm with blocked access is characterized by a 25%–35% higher throughput compared to the free-access algorithm. At the same time, the considered algorithms can be effectively employed in the case when the traffic load does not exceed 70%–80% of the available throughput.

Keywords: 5G; NB-IoT; RACH; delay; tree splitting

DOI: 10.14357/08696527250402

EDN: TXPHCY

Acknowledgments

The study was funded by the Russian Science Foundation, project number 25-79-10142, <https://rscf.ru/en/project/25-79-10142/> (recipient V. A. Beschastnyi; Sections 1, 2, 5, and 6) and with the financial support of the Ministry of Science and Higher Education of the Russian Federation, grant agreement No. FSRF-2023-0003, "Fundamental principles of building of noise-immune systems for space and satellite communications, relative navigation, technical vision and aerospace monitoring" (recipients A. M. Turlikov and N. V. Stepanov; Sections 3 and 4).

References

1. ITU-R. 2017. Minimum requirements related to technical performance for IMT-2020 radio interface(s). Report M.2410-0. Available at: <https://www.itu.int/pub/R-REP-M.2410-2017> (accessed October 14, 2025).
2. Peralta, E., T. Levanen, F. Frederiksen, and M. Valkama. 2021. Two-step random access in 5G new radio: Channel structure design and performance. *93rd Vehicular*

- Technology Conference Proceedings*. IEEE. Art. 9449057. 7 p. doi: 10.1109/VTC2021-Spring51267.2021.9449057.
3. Stepanov, N., A. Turlikov, and V. Begishev. 2023. Balancing the data transmission and random access phases in 6G mMTC radio technologies. *IEEE Commun. Lett.* 27(12):3419–3423. doi: 10.1109/LCOMM.2023.3322598.
 4. Holma, H., and A. Toskala. 2024. Industrial internet of things. *5G technology: 3GPP evolution to 5G-advanced*. Eds. H. Holma, A. Toskala, and T. Nakamura. 2nd ed. Wiley. 521–545. doi: 10.1002/9781119816058.ch18.
 5. Seo, J.-B., W.T. Toor, and H. Jin. 2021. Analysis of two-step random access procedure for cellular ultra-reliable low latency communications. *IEEE Access* 9:5972–5985. doi: 10.1109/ACCESS.2020.3048824.
 6. Capetanakis, J. 1979. Tree algorithms for packet broadcast channels. *IEEE T. Inform. Theory* 25(5): 505–515. doi: 10.1109/TIT.1979. 1056093.
 7. Tsybakov, B. S., and V. A. Mikhailov. 1980. Random multiple packet access: Part-and-try algorithm. *Probl. Inform. Transm.* 16(4):305–317.
 8. Tsybakov, B. S., and V. A. Mikhailov. 1978. Free synchronous packet access in a broadcast channel with feedback. *Probl. Inform. Transm.* 14(4):259–280. EDN: BWSVDW.
 9. Kerekesh, I., and V. A. Mikhaylov. 1984. Stek-algoritm SMD v kanale s posimvol'noy sinkhronizatsiey [Stack algorithm of SMD in a channel with symbolwise synchronization]. *6th Symposium (International) on Information Theory*. Moscow–Tashkent. 1:80–82.
 10. 3GPP. 2025. NR; Physical channels and modulation (Release 19): Technical Specification 38.211 V19.0.0. Available at: https://www.3gpp.org/ftp/Specs/archive/38_series/38.211/38211-j00.zip (accessed October 14, 2025).
 11. 3GPP. 2021. Study on narrow-band internet of things (NB-IoT)/enhanced machine type communication (eMTC) support for non-terrestrial networks (NTN) (Release 17). Technical report 36.763 V17.0.0. Available at: https://www.3gpp.org/ftp/Specs/archive/36_series/36.763/36763-h00.zip (accessed October 14, 2025).
 12. Mezzavilla, M., M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi. 2018. End-to-end simulation of 5G mmWave networks. *IEEE Commun. Surv. Tut.* 20(3):2237–2263. doi: 10.1109/COMST.2018.2828880.

Received August 22, 2025

Accepted October 15, 2025

Contributors

Beschastnyi Vitalii A. (b. 1992) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Probability Theory and Cyber Security, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; beschastnyy-va@rudn.ru

Turlikov Andrey M. (b. 1957) — Doctor of Science in technology, professor, head of Department of Infocommunication Technologies and Communication Systems, Saint

Petersburg State University of Aerospace Instrumentation, 67 Bolshaya Morskaya Str., Saint Petersburg 190000, Russian Federation; turlikov@guap.ru

Stepanov Nikita V. (b. 1996) — assistant professor, Department of Infocommunication Technologies and Communication Systems, Saint Petersburg State University of Aerospace Instrumentation, 67 Bolshaya Morskaya Str., Saint Petersburg 190000, Russian Federation; stepanov.nikita@guap.ru

Shorgin Vsevolod S. (b. 1978) — Candidate of Science (PhD) in technology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vshorgin@ipiran.ru

НЕЙРО-НЕЧЕТКАЯ МОДЕЛЬ И ПРЕЦЕДЕНТНАЯ ЭКСПЕРТНАЯ СИСТЕМА ТРАНСФОРМАЦИОННОГО ГИБРИДНОГО ИНТЕЛЛЕКТА ПРОГНОЗИРОВАНИЯ ИСХОДОВ ОСТРОГО ПАНКРЕАТИТА*

*С. Б. Румовская*¹

Аннотация: При полиморбидных, полиэтиологических заболеваниях со сложной клинической картиной и фазовым течением, подобных острому панкреатиту (ОП), в случае развития инфекционных осложнений и при отсутствии адекватной тактики лечения и хирургического пособия летальность может достигать 100%. Улучшение результатов лечения пациентов с ОП возможно путем раннего прогнозирования инфекционных осложнений и возможного летального исхода, выделяя таким образом наиболее тяжелую группу пациентов с целью своевременной комплексной интенсивной терапии и адекватного хирургического лечения. Используемые методы и подходы в большинстве случаев максимально информативны только к концу вторых–третьих суток, в то время как эффективность терапевтического воздействия наибольшая в первые–вторые сутки. Следовательно, необходимо перейти к персонализированным интегрированным алгоритмам прогнозирования. В работе представлены результаты разработки гетерогенного модельного поля (ГМП) трансформационной модели, разрабатывающейся как «нейро-нечеткая прецедентная экспертная система (ЭС)». Данная модель представляет собой часть разрабатываемых кооперативных самонастраивающихся гибридных интеллектуальных систем (ГиИС) персонализированной оценки состояния пациента (на примере ОП).

Ключевые слова: гибридная интеллектуальная система; нейро-нечеткая система; прецедентная экспертная система; прогнозирование в медицине; острый панкреатит

DOI: 10.14357/08696527250403

EDN: WGIRDY

1 Введение

Прогнозирование состояния здоровья пациента осложняется полиморбидностью и гетерогенностью состояния объекта исследования, неоднородностью, неточностью и нечеткостью как симптоматической информации, так и информации о самом процессе оценки и прогнозировании, а также об этиологии и патогенезе заболеваний. Прогноз [1] — один из показателей понимания и объяснения

*Исследование выполнено за счет гранта Российского научного фонда № 24-21-00392, <https://rscf.ru/project/24-21-00392/>.

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sophiyabr@gmail.com

состояния организма пациента (его систем и подсистем). В [2] представлены результаты анализа особенностей проблемной среды оценки (как диагностической, так и прогностической) состояния организма пациента с подозрением на ОП, в том числе статистические данные: (1) по смертности от заболеваний поджелудочной железы, в частности от ОП — до 15%, а для его тяжелых форм, а также послеоперационной — до 70% и выше; (2) о врачебных ошибках при поступлении подобных больных, которые достигают 30% ввиду неверных оценок состояния пациента, а также прогнозирования тяжелых состояний и осложнений. Представление о динамике и направлении развития заболевания формируется для организма как единого целого в рамках дифференциальной прогностики — для формирования окончательного заключения рассматриваются варианты будущего течения болезни, взвешиваются все за и против и оцениваются. Необходимы технологии виртуализации принятия подобных решений, использующие инструментарий «мягких» вычислений [3], в частности нечеткие системы, искусственные нейронные сети и генетические алгоритмы (ГА). Эту необходимость подтверждают результаты анализа в [4], отечественных и зарубежных публикаций в научной литературе по теме интеллектуальных систем поддержки принятия решений в медицине, в том числе прогнозных, разработанных с применением искусственных нейронных сетей, глубокого обучения, ансамблевых методов, логистической регрессии, ЭС и ГиИС, интегрирующих, например, многоагентные системы, онтологии и моделирование иммунной сети; экспертные производственные системы, ГА, искусственные нейронные сети и нечеткие системы; искусственные нейронные сети и нечеткие системы; сверточные нейронные сети и ГА.

В [5] предлагается разработка нового класса интеллектуальных систем — кооперативных самонастраивающихся ГиИС «Виртуальные консилиумы персонализированной оценки тяжести и прогнозирования состояния пациентов» (ВК ПОТ ПСП) на примере ОП. ВК ПОТ ПСП комбинируют модели ГиИС двух типов:

- (1) функциональную интегрированную модель оценки тяжести состояния сложного объекта и выработки решений о дальнейшем развитии заболевания, основываясь на разнородных сотрудничающих знаниях об объекте, проблеме, а также этиологии и патогенезе заболевания на всех фазах его течения на протяжении всего процесса ведения пациента;
- (2) трансформационную интегрированную модель прогнозирования исходов заболевания с обоснованием прогноза.

Проблематика и перспективы разработки подобных систем рассмотрены в [5]. В [2, 6, 7] представлены результаты системного анализа проблемы персонализированной оценки тяжести и прогнозирования состояния сложного объекта (ПОТ ПС СО) — организма пациента, его систем и подсистем на примере ОП: идентификации проблемы, ее редукции, а также спецификации и выбора методов автоматизированного решения функциональных и технологических задач из декомпозиции, полученной в результате редукции. В данной декомпозиции

подмножество задач лица, принимающего решения, включает, помимо прочих, такие задачи, как p_{PO}^h — прогноз исходов ОП (панкреатическая инфекция и летальный исход); p_{tehPO}^h — построение входного множества информативных признаков для прогнозирования исходов; p_{intP}^h — интерпретация прогноза. В настоящей работе рассмотрены результаты разработки моделей решения данных задач, интеграция которых образует трансформационную ГиИС, а именно: функциональной модульной нейро-нечеткой модели прогнозирования исходов ОП и интерпретирующей ее прогноз функциональной модели прецедентной ЭС с базой прецедентов (БП) и гибридной фреймово-продукционной базой знаний (БЗ), а также ГА оптимизации набора показателей состояния здоровья для прогнозирования исходов ОП (технологическая модель). Все модели будут включены в ГМП будущего ВК ПОТ ПСП.

2 Снижение размерности множества факторов, учитываемых при формировании прогностических оценок

Проблемы типа ПОТ ПС СО, в частности прогнозирования в медицине, — многофакторные [5, 6]. Однако восприятие данных высокой размерности в клинической практике невозможно, и врачи при построении прогнозов фокусируются только на ограниченном наборе факторов, формируемом с позиции их практического опыта. Это невербальные знания [8]. Признаковое пространство размерностью 203 построено по клиническим рекомендациям относительно ОП [9], учебно-методическим пособиям [10–12] и результатам анализа бесед с экспертом (хирургом с опытом работы более 10 лет).

Технологическая модель решает задачу безусловной оптимизации на множестве оцениваемых признаков и исключает из рассмотрения избыточные и взаимозависимые показатели, значения которых определяют прогноз, что повышает эффективность и качество принятия решения при прогнозировании исходов ОП. Для реализации технологической модели, решающей задачу «Построение входного множества информативных признаков для прогнозирования исходов» (p_{tehPO}^h), предлагается ввиду низких трудозатрат и отсутствия необходимости в обучении модели использовать ГА [7] — модель $\text{mod}_{g\text{tehPO}}$.

Обозначим основные характеристики модели $\text{mod}_{g\text{tehPO}}$. Размер популяции — 50 особей. Размер индивидуумов равен общему числу учитываемых признаков N_F — 203 (ген принимает значение 1, если признак учитывается, и 0 — если нет). В качестве оператора скрещивания используется ориентированный на размер подмножества оператор пересечения общих признаков (Subset Size-Oriented Common Feature Crossover), часто применяемый при отборе существенных признаков [13–16] и сохраняющий полезные информативные шаблоны — общие для родителей признаки (гены «1») сохраняются особями-потомками, а необщие признаки наследуются от i -го родителя с вероятностью $n_i - n^c/n^u$, где n_i — число выбранных признаков i -го родителя; n^c — число

общих выбранных признаков; n^u — число необщих выбранных признаков. Оператор мутации работает с вероятностью 0,6 для генов «1» и 0,3 для генов «0». Метод селекции — турнирный (размер турнира — 2): выбираются лучшие особи, из которых порождаются новые особи-потомки с помощью операторов мутации и скрещивания (вероятность 0,7), и порождается новая популяция. Функция приспособленности предложена по аналогии с [16]:

$$FF = 1 \cdot \frac{(1 - MC)(N_F/15 - 15SF)}{N_F} + 2 \cdot \frac{MC(N_F/15 + 15SF)}{N_F}, \quad (1)$$

где $MC = |A \cap B \cap C \cap \dots| / |A \cup B \cup C \dots|$ (A , B и C — множества экземпляров, для которых соответствующие выбранные признаки имеют значение «1», или активны) — отношение размера множества экземпляров, для которых все выбранные признаки активны к размеру множества экземпляров, для которых хотя бы один из выбранных признаков активен, т. е. мера перекрытия или коллинеарности между выбранными признаками; N_F — общее число признаков; SF — число выбранных значимых признаков, которые «не слишком близки» друг к другу, т. е. не кодируемые соседними битами в хромосоме.

Функция приспособленности устанавливает баланс между двумя аспектами:

- (1) сокращением размера набора признаков. Об этом говорит член $N_F/15 - 15SF$, который стремится к увеличению, когда SF уменьшается. Здесь значимость общего числа признаков (N_F) уменьшена за счет деления на 15, что приводит к более сильному акценту на сокращение размерности. Деление на N_F в (1) — нормализация по размеру набора данных;
- (2) выбором значимых и неколлинеарных признаков посредством MC . Член $1 \cdot (1 - MC)(N_F/15 - 15SF)/N_F$ штрафует за высокую коллинеарность (MC близко к 1) и за большое число значимых признаков ($SF > N_F/225$) — стремится максимизировать пригодность, выбирая признаки, которые вносят разный вклад в классификацию (чем меньше SF и чем меньше корреляция между признаками, тем выше оценка). Член $2 \cdot (MC(N_F/15 + 15SF))/N_F$ вознаграждает за низкую коллинеарность (MC близко к 0) и за большое число значимых признаков — коэффициент 2 усиливает влияние этого члена. Штраф за большое число значимых признаков (MC) в первом члене значительно сильнее, чем вознаграждение за большое число значимых признаков во втором члене.

Функция приспособленности агрессивно отбирает признаки, акцентируя внимание на сокращении размерности и отбирая только самые значимые признаки, даже если это означает отказ от менее важных, но неколлинеарных признаков.

Результат работы алгоритма используется для снижения размерности входных данных модели $\text{mod}_{\text{ANFISPO}}$ (см. разд. 3) на этапе разработки моделей ГМП и как маскирующая информация при инициализации ВК ПОТ ПСП (опционально) — аргумент функции модификации прецедентной ЭС mod_{erPO} (см. разд. 4), по умолчанию содержащей весь отобранный набор признаков.

Работает ГА следующим образом: (1) проводится инициализация начальной популяции индивидуумов в 50 особей; (2) вычисляются функции приспособленности индивидуумов; (3) проводится отбор родителей методом турнирной селекции; (4) избранные порождают новые особи-потомки операторами мутации и скрещивания, и формируется новая популяция; (5) после 100 итераций выбирается наилучшая особь с оптимальным набором признаков — показателей состояния здоровья пациента для прогнозирования исходов.

Генетический алгоритм реализован средствами MatLab. Результат работы ГА сравнивался с объемом входной информации для $mod_{ANFISPO}$ и mod_{erPO} (для данной модели применяется опционально, по запросу врача-пользователя), который был получен, исходя из рассчитанных по теории подтверждения Карнапа [17] факторов уверенности отобранных экспертом признаков. В результате работы ГА остался 21 признак, а с использованием теории Карнапа — 35. Анализ множеств признаков, полученных обоими методами, показал, что разница объясняется тем, что эксперты-врачи в общей массе симптомов так или иначе учитывают также и те, что имеют достоверность, равную нулю, и не влияют на прогноз. Генетический алгоритм же исключает из рассмотрения показатели состояния здоровья с нулевым коэффициентом достоверности.

3 Моделирование процесса решения задачи прогнозирования исходов острого панкреатита

Для решения задачи p_{PO}^h «Прогноз исходов ОП» (панкреатическая инфекция и летальный исход) предлагается функциональная модульная нейро-нечеткая модель $mod_{ANFISPO}$. Она включает два модуля для прогнозирования развития панкреатической инфекции (1 — да; 0 — нет) и вероятности летального исхода (1 — высокий риск; 0 — низкий риск). Каждый модуль — нечеткая система вывода Сугено, реализованная в адаптивных сетях, — ANFIS (Adaptive-Network-Based Fuzzy Inference System) [18, 19]. Модули не связаны друг с другом. Входная информация (значения 21 показателя состояния здоровья, отобранных $mod_{g_{tehPO}}$) поступает на оба модуля одновременно. Такая структура позволяет модернизировать модель при необходимости повышения числа прогнозируемых исходов путем добавления новых модулей без переобучения всех модулей сети. Каждый модуль состоит из пяти слоев: первый — термы входных переменных; второй — antecedentes (посылки) нечетких правил; третий — нормализация степеней выполнения правил; четвертый — заключения правил; пятый — агрегирование результата, полученного по различным правилам.

Формат входной информации: (1) качественные переменные — определяют показатель состояния здоровья по лингвистической шкале оценок (например, «интенсивность боли» описывается степенью принадлежности к множеству низкая, средняя, высокая, очень высокая); (2) категориальные (например, «тип боли» — тупая, острая, пульсирующая, жгучая, ноющая, схваткообразная) —

закодированы с использованием метода One-Hot Encoding: созданы отдельные бинарные признаки для каждого возможного значения; (3) количественные — нормализованы методом Min-Max Scaling, чтобы привести их к диапазону от 0 до 1; (4) бинарные, отражающие наличие, «1», или отсутствие, «0», признака. Размер обучающей выборки — 100 записей, проверочной — 40 и тестовой — 40. Проверочная выборка используется для предотвращения переобучения во время обучения ANFIS. Метод генерации нечеткой системы (Fuzzy Inference System, FIS) — алгоритм вычитающей кластеризации Subtractive Clustering (SC). Для SC радиус влияния кластеров (Influence Range) — от 0,2 до 0,8 — будет настраиваться с шагом 0,1 на этапе исследования работы всех моделей ГМП и синтезируемого над ГМП лабораторного прототипа ВК ПОТ ПСП. Если SC не даст хороших результатов, то возможна смена метода на метод нечеткой кластеризации С-средних (Fuzzy C-Means Clustering). Для каждой входной переменной генерируются по две гауссовы функции принадлежности, а для категориальных/бинарных по две треугольные (для «Нет» — пик в 0, для «Да» — пик в 1). Число правил будет равно числу кластеров, т. е. будет зависеть от радиуса (начальное число правил при радиусе 0,2–25). Параметры обучения: число циклов обучения (эпох) — 100; метод — метод обратного распространения ошибки (Backpropagation) ввиду бинарной классификации.

Метод $\text{mod}_{\text{ANFISPO}}$ реализуется средствами ANFIS Editor пакета MatLab поэтапно: (1) формирование последовательности входов и выходов, их предобработка и разделение на выборки; (2) определение структуры ANFIS в ANFIS Editor (загрузка данных, выбор метода генерации нечеткой системы и настройка параметров структуры FIS); (3) генерация ANFIS; (4) моделирование; (5) обучение, адаптация (здесь возможна ручная настройка функций принадлежности и коррекция/удаление правил совместно с экспертом) и оценка. Выполнены четыре этапа. Пятый этап запланирован в рамках исследования всех моделей ГМП будущего виртуального консилиума.

4 Экспертная система, интерпретирующая прогноз

Модель ЭС $\text{mod}_{\text{erIntP}}$ задачи «Интерпретация прогноза» (p_{IntP}^h) формирует объяснение полученного от модели $\text{mod}_{\text{ANFISPO}}$ прогноза (панкреатическая инфекция и летальный исход). Экспертная система разрабатывается на Python и включает БЗ с гибридной моделью представления знаний (продукционно-фреймовой) и БП.

Множество фреймов используется для структурированного представления знаний о ресурсах задачи p_{IntP}^h — база фактов размерностью 26 фреймов [20]. Для слотов, которые участвуют в подсчете баллов шкал, используемых при прогнозировании развития панкреатической инфекции и летального исхода (НИИ СП им. И. И. Джанелидзе, BISAP — Bedside Index of Severity in Acute Pancreatitis, SOFA — Sepsis-related Organ Failure Assessment и др.) [9–12], определены

соответствующие демоны, срабатывающие на on-set — установку значения в слот. Совокупность продукционных правил — БЗ, включающая правила-продукции вида «ЕСЛИ *антецедент*, ТО *консеквент*», имитирующие минимально возможную конструкцию точных рассуждений эксперта-врача. Размерность БЗ — 78 правил. База прецедентов также представляется в виде экспертных правил продукционного типа и пополняется по результатам работы $\text{mod}_{\text{ANFISPO}}$: входные сигналы $\text{mod}_{\text{ANFISPO}}$ — антецедент, а выходные — консеквент. После работы ANFIS новые случаи накапливаются как потенциальные прецеденты, врач-эксперт анализирует их, расширяет их консеквенты *действием* — выдачей рекомендаций, соответствующих прогнозу, и добавляет в БП. Возможно ручное заполнение БП при разработке и при поддержке, если в практике появляются нестандартные случаи.

База фактов и БЗ для ЭС разработаны по результатам анализа медицинских текстовологических источников по диагностике и прогнозированию состояния пациентов с ОП [9–12] и бесед с экспертом (хирургом с опытом работы более 10 лет). При интерпретации прогноза ЭС работает в режиме обратного вывода. Затем запускается прямой вывод на БП для поиска похожих случаев и получения рекомендаций: (1) для каждого правила в БП проводится проверка: выполняется ли условие (полностью/частично) для данных пациента — если условие выполняется (правило срабатывает), то правило считается применимым к данному случаю; (2) среди всех применимых правил выбираются три прецедента, которые имеют наибольшее число совпадений с данными пациента (чем больше признаков из условия правила соответствуют данным пациента, тем более «похожим» считается прецедент). Механизм разрешения конфликтов — выборочная стратегия.

5 Заключение

Разрабатываемая трансформационная модель на всем протяжении ведения больного предоставит возможность определить прогноз будущего направления течения заболевания и его исходов, а значит, впоследствии подобрать/скорректировать адекватную конкретной ситуации терапию с целью противодействия выявленным аномалиям (патологиям) в организме пациента, а также начать лечение на этапе управляемости патологических(ого) процессов(а) [21]. ANFIS позволит обеспечить формализацию размытых понятий и связей (нечеткость в рассуждениях врачей), вырабатывая прогноз, который будет интерпретироваться ЭС, на основе поступившей входной информации о пациенте и результатах работы виртуального консилиума. Экспертная система объединяет преимущества применения продукций и фреймов — наглядность, гибкость, модульность, наследование свойств, использование процедурных привязок и легкость внесения изменений. При этом у трансформационной ГиИС предполагается наличие двух режимов работы: «модульная ANFIS–прецедентная ЭС» — решается непосредственно проблема прогнозирования исходов заболевания и «прецедентная

ЭС – модульная ANFIS» — для обучения ANFIS в ходе разработки и поддержки, тем самым сократив время на обучение и объем обучающей выборки. Истории болезни пациентов будет содержать база данных, доступная для всех моделей ГМП и разрабатываемая на основе системы управления базами данных PostgreSQL ввиду большого разнообразия поддерживаемых типов данных и наличия универсального механизма индексирования.

Литература

1. Малая медицинская энциклопедия / Под ред. В. И. Покровского. — М.: Медицина, 1996. Т. 5. С. 44.
2. Rumovskaya S., Paramzin F. Virtual training simulator and medical decision support for personalized assessment of severity and prediction of the patients' state (the case of acute pancreatitis): Investigation of the problem area and identification of the problem // 4th Conference (International) on Technology Enhanced Learning in Higher Education Proceedings. — IEEE, 2024. P. 100–105. doi: 10.1109/TELE62556.2024.10605702.
3. Румовская С. Б. Интеллектуальные системы поддержки принятия решений в медицине: понятие, проблемы, подходы к разработке // Системы и средства информатики, 2024. Т. 34. № 2. С. 107–122. doi: 10.14357/08696527240208. EDN: IJDVVV.
4. Кривенко М. П. Мягкие вычисления в задачах медицинской диагностики // Информатика и её применения, 2021. Т. 15. Вып. 2. С. 52–59. doi: 10.14357/19922264210208. EDN: VFNFR.
5. Румовская С. Б., Кириков И. А. Моделирование выработки полного раннего дифференциального диагноза и прогноза состояния сложных объектов в медицине. Проблемы и перспективы // Системы и средства информатики, 2022. Т. 32. № 3. С. 26–35. doi: 10.14357/08696527220303. EDN: FWJRWB.
6. Румовская С. Б., Парамзин Ф. Н. Кооперативные самонастраивающиеся гибридные интеллектуальные системы персонализированной диагностики и прогнозирования в медицине: концепция, подход к разработке, декомпозиция проблемы // Системы и средства информатики, 2024. Т. 34. № 4. С. 59–72. doi: 10.14357/08696527240405. EDN: YLQTQU.
7. Rumovskaya S., Listopad S., Luchko A. Study of functional and instrumental heterogeneity of the problem of assessing the patients' body state // 6th Conference (International) on Control Systems, Mathematical Modeling, Automation and Energy Efficiency Proceedings. — IEEE, 2024. P. 724–730. doi: 10.1109/SUMMA64428.2024.10803724.
8. Поспелов Д. А., Кузнецов О. П. Знания и рассуждения в гуманитарных науках // Новости искусственного интеллекта, 1996. № 2. С. 93–98. EDN: CAGIAQ.
9. Клинические рекомендации. Острый панкреатит. https://cr.minzdrav.gov.ru/view-cr/903_1.
10. Топчиев М. А., Кчибеков Э. А., Паршин Д. С. и др. Острый панкреатит: комплексная диагностика, прогнозирование осложнений и лечение. — Астрахань: АГМУ, 2017. 120 с. EDN: WTCHRT.

11. Грекова Н. М., Бордуновский В. Н., Наймушина Ю. В., Бухвалов А. Г. Острый панкреатит. — Челябинск: Уралполисервис, 2019. 96 с. EDN: AYYFDU.
12. Попандопуло К. И., Авакмян В. А., Авакмян С. В. и др. Острые и хронические заболевания поджелудочной железы. — Краснодар: КубГМУ Минздрава России, 2020. 105 с. EDN: VEWCSV.
13. Emmanouilidis C., Hunter A., MacIntyre J. A multiobjective evolutionary setting for feature selection and a commonality-based crossover operator // Congress (International) on Evolutionary Computation Proceedings. — IEEE, 2000. Vol. 1. P. 309–316. doi: 10.1109/CEC.2000.870311.
14. Hamon J., Gaël E., Dassonneville R., Jacques J., Dhaenens C. Use of a novel evolutionary algorithm for genomic selection, 2015. <https://hal.inria.fr/hal-01100660>.
15. Jourdan L., Dhaenens C., Talbi El-Gh., Gallina S. A data mining approach to discover genetic and environmental factors involved in multifactorial diseases // Knowl.-Based Syst., 2002. Vol. 15. No. 4. P. 235–242. doi: 10.1016/S0950-7051(01)00145-9.
16. Turabieh H. GA-based feature selection with ANFIS approach to breast cancer recurrence // IJCSI Int. J. Computer Science Issues, 2016. Vol. 13. Iss. 1. P. 36–41. doi: 10.20943/IJCSI-201602-3641.
17. Новикова В. А., Андреева Е. Ю., Туйкина Д. К. Искусственный интеллект и экспертные системы. — Курск: КГУ, 2004. 45 с.
18. Jang J.-S. R. ANFIS: Adaptive-network-based fuzzy inference system // IEEE T. Syst. Man Cyb., 1993. Vol. 23. No. 3. P. 665–685. doi: 10.1109/21.256541.
19. Negnevitsky M. Artificial intelligence: A guide to intelligent systems. — Reading, MA, USA: Addison-Wesley: Pearson Education, 2005. 415 p.
20. Minsky M. A framework for representing knowledge // Frame conceptions and text understanding / Ed. D. Metzger. — Berlin, Boston: De Gruyter, 1979. P. 1–25. doi: 10.1515/9783110858778-003.
21. Забежайло М. И., Грушо А. А., Грушо Н. А., Тимонина Е. Е. Поддержка решения задач диагностического типа // Системы и средства информатики, 2021. Т. 31. № 1. С. 69–81. doi: 10.14357/08696527210106. EDN: DMYQIG.

Поступила в редакцию 03.07.2025

Принята к публикации 15.10.2025

NEURO-FUZZY MODEL AND PRECEDENT EXPERT SYSTEM OF THE TRANSFORMATIONAL HYBRID INTELLIGENCE FOR PREDICTING OUTCOMES OF THE ACUTE PANCREATITIS

S. B. Rumovskaya

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: In polymorbid and polyetiological diseases with complex clinical presentations and phasic progression such as acute pancreatitis, the development of infectious complications combined with inadequate treatment strategies and surgical interventions can lead to mortality rates reaching 100%. Improving

outcomes for patients with acute pancreatitis requires the early prediction of infectious complications and potential lethal outcomes to identify the most severe patient cohort for timely comprehensive intensive therapy and adequate surgical treatment. Current diagnostic methods and approaches are maximally informative only by the end of the second to third days. However, the highest therapeutic efficacy is observed within the first 24–48 h. Thus, transitioning to personalized integrated prediction algorithms is critical. The paper presents the results of the developing a heterogeneous field of models within the transformational model designed as a “neuro-fuzzy – case-based expert system. This model is the part of the future cooperative self-adjusting hybrid intelligent systems for personalized patient state assessment (the case of acute pancreatitis).

Keywords: hybrid intelligent system; neuro-fuzzy system; precedent expert system; predicting in medicine; acute pancreatitis

DOI: 10.14357/08696527250403

EDN: WGIRDY

Acknowledgments

The study was funded by the Russian Science Foundation according to the research project No. 24-21-00392, <https://rscf.ru/project/24-21-00392/>.

References

1. Pokrovskiy, V. I., ed. 1996. *Malaya meditsinskaya entsiklopediya* [Small medical encyclopedia]. Moscow: Meditsina. 5:44.
2. Rumovskaya, S., and F. Paramzin. 2024. Virtual training simulator and medical decision support for personalized assessment of severity and prediction of the patients' state (the case of acute pancreatitis): Investigation of the problem area and identification of the problem. *4th Conference (International) on Technology Enhanced Learning in Higher Education Proceedings*. IEEE. 100–105. doi: 10.1109/TELE62556.2024.10605702.
3. Rumovskaya, S. B. 2024. Intellektual'nye sistemy podderzhki prinyatiya resheniy v meditsine: ponyatie, problemy, podkhody k razrabotke [Intelligent decision support systems in medicine: Concept, problems, and approaches to the development]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 34(2):107–122. doi: 10.14357/08696527240208. EDN: IJDVVV.
4. Krivenko, M. P. 2021. Myagkie vychisleniya v zadachakh meditsinskoy diagnostiki [Soft computing in problems of medical diagnostics]. *Informatika i ee Primeneniya — Inform. Appl.* 15(2):52–59. doi: 10.14357/19922264210208. EDN: VFNFR.
5. Rumovskaya, S. B., and I. A. Kirikov. 2022. Modelirovanie vyrabotki polnogo rannego differentsial'nogo diagnoza i prognoza sostoyaniya slozhnykh ob"ektov v meditsine. Problemy i perspektivy [Modeling of the development of a complete early differential diagnosis and prognosis of the state of complex objects in medicine. Problems and prospects]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(3):26–35. doi: 10.14357/08696527220303. EDN: FWJRWB.
6. Rumovskaya, S. B., and F. N. Paramzin. 2024. Kooperativnye samonastrayayushchi- esya gibridnye intellektual'nye sistemy personalizirovannoy diagnostiki i prognozirovaniya v meditsine: kontseptsiya, podkhod k razrabotke, dekompozitsiya

- problem [Cooperative self-configuring hybrid intelligent systems for personalized diagnostics and prognosis in medicine: Conceptual idea, development approach, and problem decomposition]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 34(4):59–72. doi: 10.14357/08696527240405. EDN: YLQTQU.
7. Rumovskaya, S., S. Listopad, and A. Luchko. 2024. Study of functional and instrumental heterogeneity of the problem of assessing the patients' body state. *6th Conference (International) on Control Systems, Mathematical Modeling, Automation and Energy Efficiency Proceedings*. IEEE. 724–730. doi: 10.1109/SUMMA64428.2024.10803724.
8. Pospelov, D. A., and O. P. Kuznetsov. 1996. Znaniya i rassuzhdeniya v gumanitarnykh naukakh [Knowledge and reasoning in the humanities]. *Novosti iskusstvennogo intellekta* [Artificial Intelligence News] 2:93–98. EDN: CAGIAQ.
9. Ministerstvo zdravookhraneniya Rossiyskoy Federatsii [Ministry of Health of the Russian Federation]. 2024. Klinicheskie rekomendatsii. Ostryy pankreatit [Clinical guidelines. Acute pancreatitis]. Available at: https://cr.minzdrav.gov.ru/view-cr/903_1 (accessed October 17, 2025).
10. Topchiev, M. A., E. A. Kchibekov, D. S. Parshin, *et al.* 2017. *Ostryy pankreatit: kompleksnaya diagnostika, prognozirovanie oslozhneniy i lechenie* [Acute pancreatitis: Complex diagnosis, prediction of complications, and treatment]. Astrakhan: AGMU. 120 p. EDN: WTCERT.
11. Grekova, N. M., V. N. Bordunovskiy, Yu. V. Naymushina, and A. G. Bukhvalov. 2019. *Ostryy pankreatit* [Acute pancreatitis]. Chelyabinsk: Uralpoliservis. 96 p. EDN: AYYFDU.
12. Popandopulo, K. I., V. A. Avakimyan, S. V. Avakimyan, *et al.* 2020. *Ostrye i khronicheskie zabolevaniya podzhelechnoy zhelezy* [Acute and chronic diseases of the pancreas]. Krasnodar: KubGMU Minzdrava Rossii. 105 p. EDN: VEWCSV.
13. Emmanouilidis, C., A. Hunter, and J. MacIntyre. 2000. A multiobjective evolutionary setting for feature selection and a commonality-based crossover operator. *Congress (International) on Evolutionary Computation Proceedings*. IEEE. 1:309–316. doi: 10.1109/CEC.2000.870311.
14. Hamon, J., E. Gaél, R. Dassonneville, J. Jacques, and C. Dhaenens. 2015. Use of a novel evolutionary algorithm for genomic selection. Available at: <https://hal.inria.fr/hal-01100660> (accessed October 17, 2025).
15. Jourdan, L., C. Dhaenens, E.-G. Talbi, and S. Gallina. 2002. A data mining approach to discover genetic and environmental factors involved in multifactorial diseases. *Knowl.-Based Syst.* 15(4):235–242. doi: 10.1016/S0950-7051(01)00145-9.
16. Turabieh, H. 2016. GA-based feature selection with ANFIS approach to breast cancer recurrence. *Int. J. Computer Science Issues* 13(1):36–41. doi: 10.20943/IJCSI-201602-3641.
17. Novikova, V. A., E. Yu. Andreeva, and D. K. Tuykina. 2004. *Iskusstvennyy intellekt i ekspertnye sistemy* [Artificial intelligence and expert systems]. Kursk: KGU. 45 p.
18. Jang, J.-S. R. 1993. ANFIS: Adaptive-network-based fuzzy inference systems. *IEEE T. Syst. Man Cyb.* 23(3):665–685. doi: 10.1109/21.256541.
19. Negnevitsky, M. 2005. *Artificial intelligence: A guide to intelligent systems*. Reading, MA: Addison-Wesley: Pearson Education. 415 p.

20. Minsky, M. 1979. A framework for representing knowledge. *Frame conceptions and text understanding*. Ed. D. Metzinger. Berlin, Boston: De Gruyter. 1–25. doi: 10.1515/9783110858778-003.
21. Zabezhailo, M. I., A. A. Grusho, N. A. Grusho, and E. E. Timonina. 2021. Podderzhka resheniya zadach diagnosticheskogo tipa [Support for solving diagnostic type problems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(1):69–81. doi: 10.14357/08696527210106. EDN: DMYQIG.

Received July 3, 2025

Accepted October 15, 2025

Contributor

Rumovskaya Sophiya B. (b. 1985) — Candidate of Science (PhD) in technology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sophiyabr@gmail.com

АРХИТЕКТУРА СИСТЕМЫ СЕМАНТИЧЕСКОГО АНАЛИЗА НАУЧНО-ТЕХНИЧЕСКИХ ДОКУМЕНТОВ*

*А. В. Кан¹, Ал-др А. Хорошилов², И. А. Чечулин³, С. А. Ступников⁴,
Ю. В. Никитин⁵, Ал-ей А. Хорошилов⁶*

Аннотация: Рассматривается новая функциональная архитектура системы семантического анализа (ССА) научно-технических документов (НТД). Система относится к категории интеллектуальных платформ для агрегации и анализа НТД. Основное методологическое решение, отличающее ее от существующих систем и нацеленное на повышение эффективности и качества обработки НТД, состоит в использовании адаптируемого под различные отрасли науки и техники комплекса декларативных средств, реализующих концепцию фразеологического концептуального анализа текстов (ФКАТ), предполагающую объединение результатов графематического, морфологического, концептуального и семантико-синтаксического анализа текста в формализованную модель метаданных (ФММ). Рассмотрены составляющие архитектуру подсистемы и модули, описано их функционирование и взаимодействие. Один из наиболее важных компонентов системы — автоматически формируемые и настраиваемые на отрасль знания морфологические, семантико-синтаксические и концептуальные словари, обеспечивающие актуальность комплекса лингвистических ресурсов. Рассмотрена классификация ССА текстов и реализуемых ими методов, место системы новой архитектуры в классификации и ее отличительные черты.

Ключевые слова: фразеологический концептуальный анализ текстов; семантический анализ текстов; обработка научно-технических текстов

DOI: 10.14357/08696527250404

EDN: YTYHOU

1 Введение

Стратегией научно-технологического развития Российской Федерации (Указ Президента РФ № 145 от 28.02.2024) предусмотрено обеспечение доступа иссле-

* Исследование выполнено в рамках государственного задания FFFU-2025-0010.

¹Всероссийский институт научной и технической информации Российской академии наук; Московский авиационный институт, kanav@viniti.ru

²Всероссийский институт научной и технической информации Российской академии наук; Московский авиационный институт, khoroshilov@viniti.ru

³Всероссийский институт научной и технической информации Российской академии наук, chechulin@viniti.ru

⁴Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Всероссийский институт научной и технической информации Российской академии наук, sstupnikov@frccsc.ru

⁵Национальный исследовательский центр «Институт имени Н. Е. Жуковского», yuri.v.nikitin@gmail.com

⁶Всероссийский институт научной и технической информации Российской академии наук, khoroshilovaa@viniti.ru

дователей и исследовательских коллективов к национальным и международным информационным ресурсам, результатам научных исследований, научной инфраструктуре, к технологическим разработкам на всей территории страны. Достижение поставленных целей невозможно без развития Государственной системы научно-технической информации (НТИ), основная задача которой заключается в обеспечении достоверной и актуальной НТИ научно-исследовательских, конструкторских и промышленно-технологических организаций, поддержке научных исследований и технологического прогресса.

Несмотря на мировые достижения последних лет в области обработки и анализа текстов, технологический процесс аналитико-синтетической обработки документов (в частности, в государственных информационных системах) остается далеким от полной автоматизации. К примеру, домен «Наука и инновации»¹ предлагает поиск среди 4 млн научных отчетов, диссертаций, научно-исследовательских и опытно-конструкторских работ и результатов интеллектуальной деятельности, однако только по названию, ключевым словам, аннотации и некоторым метаданным, но не полнотекстовый семантический поиск. Практическая задача выбора существующей или разработки новой ССА НТД для повышения эффективности этого процесса делает актуальной научную задачу сравнительного анализа известных систем обработки текстов и, в частности, НТД, а также разработку новой функциональной архитектуры ССА НТД, направленной на повышение эффективности и качества обработки НТД, выбора ее методологической основы.

В настоящее время основным инструментом искусственного интеллекта для автоматической обработки текстовой информации становятся большие языковые модели (БЯМ). Как показывает практика использования БЯМ, они могут при обработке текстовой информации наряду с результатами высокого качества генерировать выдуманную и недостоверную информацию [1]. Ключевыми проблемами нейросетевых технологий остаются необходимость создания высококачественных обучающих наборов данных больших объемов, недетерминированность и неинтерпретируемость получаемых результатов, т. е. отсутствие прозрачности в установлении причинно-следственных связей при сквозной обработке текстовых данных [2]. Известно также, что для эффективных атак отравления (подмешивания измененных данных в обучающую выборку), в частности для продуцирования нежелательного поведения модели по фразе-триггеру в запросе (например, когда модель отвечает на опасные запросы, которые при ее нормальном функционировании отвергаются), достаточно лишь ограниченного числа обучающих примеров, не зависящего от размера обучающей выборки². Тем не менее быстро развиваются основанные на БЯМ системы обработки и анализа

¹<https://gisnauka.ru/>.

²Souly A., Rando J., Chapman E., Davies X., Hasircioglu B., Shereen E., Mougan C., Mavroudis V., Jones E., Hicks C., Carlini N. Poisoning attacks on LLMs require a near-constant number of poison samples. — Cornell University, 2025. arXiv:2510.07192.

НТД. Например, INVENTORUS¹, позиционируемая как интеллектуальная аналитическая платформа для ускорения и оптимизации процессов поиска и анализа мировой НТИ, осуществляет поиск среди более чем 437 млн документов². Сервисы мониторинга научной и инновационной активности авторов и организаций, сравнения организаций, формирования дайджеста публикаций и аналитических отчетов, вопросно-ответный сервис, предлагаемые системой, основаны на БЯМ. Открытыми остаются вопросы качества и эффективности поиска и анализа НТД в системах этой категории.

Методы компьютерной лингвистики, в основе которых лежат современные лингвистические модели смысловой структуры текстов [2–4], машинные словарные ресурсы, тезаурусы и онтологии, почти лишены упомянутых недостатков БЯМ. Они могут допускать неточности, но галлюцинации им не свойственны. Основанные на методах компьютерной лингвистики системы могут быть разумной альтернативой огромным языковым моделям, обученным на десятках терабайт текстовой информации, при смысловом анализе описаний отраслевых технологических процессов, когда в отраслевом текстовом информационном пространстве циркулируют объемы текстовой информации на несколько порядков ниже, синтаксическая структура гораздо беднее текстовых ситуаций, на которых обучается БЯМ, но в то же время структура отраслевых текстов более специфична и насыщена отраслевой терминологией.

Системы семантического анализа текстов развиваются и применяются в различных отраслях, и их можно разбить на две крупные категории. К первой категории относятся, в частности, системы крупных облачных провайдеров (Google, Amazon, Microsoft), которые реализуют методы *базового семантического анализа* (БСА). Ко второй категории относятся промышленные и академические системы, реализующие методы *специализированного семантического анализа*: например, в системе Leximancer³ реализованы методы контентного анализа, а в системе TITANIS [5] реализованы методы психолингвистического анализа, анализа эмоциональной структуры, определения фрустрации. Среди систем этой категории, реализующих методы обработки и анализа НТД, можно выделить группу систем, реализующих *базовые методы структуризации* НТД (например, GROBID⁴). Наиболее же зрелые в методическом смысле системы решают задачи *специализированного семантического анализа НТД*. Такие системы позиционируются как интеллектуальные платформы для агрегации и анализа НТД: например, SciApp⁵, основанная на программно-аппаратном комплексе интеллектуального поиска и анализа больших массивов текстов TextAppliance [5]. В данной работе рассматривается новая функциональная архитектура ССА НТД (разд. 4), отно-

¹<https://inventorus.ru/inventorusproduct>.

²По информации на сайте компании-разработчика системы.

³<https://www.leximancer.com/>.

⁴<https://grobid.readthedocs.io/>.

⁵<https://sciapp.ru/>.

сящейся к этой же группе. Архитектура направлена на повышение эффективности и качества обработки НТД. Ее методологическое решение состоит в создании относительно простого, малозатратного, быстро адаптируемого под новую отрасль науки и техники комплекса декларативных средств, реализующих концепцию ФКАТ [6] (разд. 3). В разд. 2 предложена классификация ССА текстов и реализуемых ими методов с концентрацией на системах обработки и анализа НТД.

2 Классификация систем семантического анализа текстов и реализуемых ими методов

Проведенный анализ современных ССА текстов показывает, что их можно разбить на две крупные категории по спектру реализуемых ими методов. К *первой категории* относятся системы крупных облачных провайдеров (Google Cloud NL¹, Amazon Comprehend², Azure AI Language³), проприетарные системы (Dandelion⁴, Semanticase⁵) и некоторые академические системы (Texterra⁶ [7]), нацеленные на реализацию следующих методов БСА:

- *базовый лингвистический анализ* (токенизация, лемматизация, морфологический анализ, выделение частей речи, разделение текста на предложения);
- *извлечение сущностей*. Сущности могут быть из ограниченного встроенного списка (персоны, организации, места и т. д.) либо из списка, настраиваемого пользователем (custom). При этом для извлечения сущностей создается модель машинного обучения на основе размеченной выборки, предоставленной пользователем;
- *анализ тональности*. Тональность обычно бывает позитивная, негативная, нейтральная или смешанная, для каждой из категорий дается числовая оценка. Оценка дается сущности либо тексту в целом. Отслеживаются временные тренды тональности;
- *классификация*. Текст классифицируется на предопределенные категории, которых может быть достаточно много (например, в системе Google Cloud NL их поддерживается более 1000), либо на категории, настраиваемые пользователем (в этом случае на основе размеченной выборки создается модель машинного обучения);
- *выделение и маскировка персональной информации* (например, данных о кредитных картах, номерах документов, сетевых адресах);
- *выделение ключевых слов* (обычно — существительных с некоторым модификатором);

¹<https://cloud.google.com/natural-language>.

²<https://aws.amazon.com/comprehend/>.

³<https://learn.microsoft.com/en-us/azure/ai-services/language-service/overview>.

⁴<https://dandelion.eu/>.

⁵<https://www.semanticcase.it/>.

⁶<https://texterra.ispras.ru/>.

- *определение языка* (до 100 языков в системах облачных провайдеров);
- *кластеризация и моделирование тем*. Используются специализированные методы кластеризации и моделирования тем (латентное размещение Дирихле, стохастические модели совместного появления слов); проводится выделение категорий, определение названия тем, их суммаризация;
- *суммаризация*. *Экстрактивная* суммаризация предполагает выделение наиболее характерных предложений из текста и из позиций. При *абстрактивной* суммаризации происходит генерация связного реферата, не обязательно в точности цитирующего исходный текст;
- *связывание сущностей* со статьями Википедии и/или понятиями dbpedia;
- *определение семантической схожести текстов*;
- *формирование ответов на вопросы* позволяет формировать «разговорный» слой над текстовыми данными с настройкой по корпусу текстов и разметкой вопросно-ответных пар.

Ко *второй категории* относятся промышленные и академические системы, реализующие методы специализированного семантического анализа. Отдельную подкатегорию образуют системы, реализующие методы *обработки и анализа НТД*. Среди них можно выделить группу систем, реализующих *базовые методы структуризации НТД*:

- *разбор заголовочной части НТД и выделение специализированных сущностей*: название, аннотация, авторы, аффилиации, адреса, ключевые слова, финансирование;
- *извлечение и разбор библиографических ссылок, их консолидация и разрешение*. Применяются модели глубокого обучения. Поддерживается широкий набор метаданных, включая DOI и PMID;
- *структуризация текста и классификация предложений*. Используются модели сегментации документа и структуризации текста (абзацы, названия разделов, ссылки и сноски, рисунки, таблицы и т. д.) Проводится классификация предложений НТД: цель исследования, новизна, методы, результаты с использованием моделей глубокого обучения.

К этой группе относятся, в частности, системы GROBID [8], SciWING¹ [9], Scim² [10].

Наиболее зрелые в методическом смысле системы помимо методов БСА и методов базовой структуризации НТД реализуют методы специализированного семантического анализа НТД:

- *концептуальный анализ* [11];
- *семанτικο-синтаксический анализ* [2, 11];

¹<https://sciwing.readthedocs.io/en/latest/index.html>.

²<https://github.com/rayfok/scim>.

- дискурсивный анализ [2];
- формирование модели метаданных текста [11];
- иерархическая классификация с заданными классификаторами (ГРНТИ, УДК, . . .);
- выделение тезауруса;
- обнаружение кросс-языковых заимствований и близких по смыслу высказываний;
- семантический поиск по запросу на естественном языке с учетом ограничений на метаданные, поиск по документу-образцу, поиск по базе знаний по запросу на декларативном языке;
- агрегация метаданных.

К этой категории относятся, в частности, система SciApp и система новой архитектуры (см. разд. 4). Эти системы также нацелены на реализацию слоя проблемно-ориентированных сервисов, решающих следующие аналитические задачи на основе методов семантического анализа:

- оценка научных проектов;
- построение научных ландшафтов;
- обнаружение трендов в областях знаний;
- мониторинг рыночных трендов, конкурентная разведка;
- аналитический научный обзор предметных областей;
- «трансфер технологий» — поиск технологических решений под задачи промышленности;
- отслеживание цепочки «исследование–разработка–продукт»;
- определение центров компетенций;
- поиск экспертов по заданной проблематике;
- формирование, автоматизированное пополнение и анализ пользовательских подборок документов;
- выделение научных направлений;
- выделение коллективов исследователей.

3 Модель фразеологического концептуального анализа текстов

Языковая модель ФКАТ представляет собой эволюционное развитие традиционных лингвистических подходов, базирующихся на модели системы флексивных классов русского языка, принципе лингвистической аналогии, центроидно-контекстной модели и методах разноуровневого анализа текстов, используемых в технологических процессах автоматической обработки текстов [6]. С теоретической точки зрения ФКАТ опирается на исследования семантической организации

мультиязычных текстов, выявляющие следующие лингвистические закономерности: (а) в языковых системах объективно существует разноуровневая иерархия смысловых единиц; (б) семантика вышестоящих единиц не является аддитивной суммой значений составляющих элементов; (в) фразеологические текстовые конструкции представляют наиболее устойчивые и воспроизводимые смысловые образования [3].

Формализация смыслового содержания текста происходит путем преобразования его текстового представления в структурированную формализованную модель представления его смысла — ФММ текста [6]. Эта модель состоит из двух типов структурированных данных: описания текстовой структуры документа и описания его понятийной структуры. Текстовая структура состоит из элементов библиографических описаний документов и названий разделов документа. Понятийная структура ФММ включает ряд моделей нижестоящих уровней: графематической, морфологической, семантико-синтаксической и концептуальной моделей, обеспечивающих решение ряда высокоуровневых прикладных задач.

Графематическая модель текста (ГМТ) концептуализирует текст как упорядоченную последовательность токенов и разделительных элементов, где каждому токену присваиваются соответствующие типологические и графематические характеристики. Процедура формирования ГМТ выполняет трансформацию символьной последовательности в структурированный набор текстовых информационных единиц. Параллельно с токенизацией выполняется многоуровневая классификация токенов по следующим параметрам: типологическая принадлежность (простые/составные лексемы, числовые выражения, пунктуационные маркеры); видовая спецификация (дробные величины, сетевые идентификаторы, контактные данные); графематические характеристики (алфавитно-символьный состав и регистровые особенности).

Морфологическая модель текста (ММТ) формализует текст как последовательность токенов, каждый из которых сопровождается комплексом грамматических и семантических атрибутов, определяемых без учета контекстного окружения. Данная модель обеспечивает строгое разделение между грамматическими и семантическими признаками слов, формируя их исчерпывающий морфологический портрет. Процесс построения ММТ осуществляется предобученной морфологической языковой моделью, включающей систему декларативных языковых ресурсов (словарей словоформ, буквосочетаний, семантических признаков, грамматических таблиц преобразования словоформ).

Концептуальная модель текста представляет понятийную структуру текста в виде системы понятий текста и их смысловых отношений (парадигматических и синтагматических). Концептуальная модель текста базируется на методах точного и предиктивного выделения понятий. *Точный* вариант модели базируется на использовании *эталонного концептуального словаря*, включающего основной понятийный состав широкого спектра тематических областей. *Предиктивный* вариант алгоритма выделения наименований понятий позволяет решить проблему выявления всех возможных понятий в тексте на основе метода *лингвистической*

аналоги представления мнемонических форм текстовых понятий в текстах различных тематик с установлением грамматических характеристик текстовых форм слов, входящих в состав словосочетаний.

Семантико-синтаксическая модель (ССМ) текста состоит из последовательности предложений, сопровождаемых синтаксической и семантической информацией. Процесс формирования ССМ реализует извлечение и классификацию синтаксических конструкций предложения, а также устанавливает их роль в предложении и определяет их структуру с использованием обеспечивающих субмоделей грамматических классов слов, дерева зависимостей слов, структуры каркаса предложения, членов предложения, предикатно-актантной структуры предложения и центроидно-контекстной модели, предназначенной для однозначного разрешения неоднозначной текстовой ситуации на основе анализа ее контекстного окружения [11, 12].

На конечном этапе формализации текста грамматические, синтаксические и семантические характеристики всей иерархии моделей исходного текста преобразуются в многослойную ФММ текста.

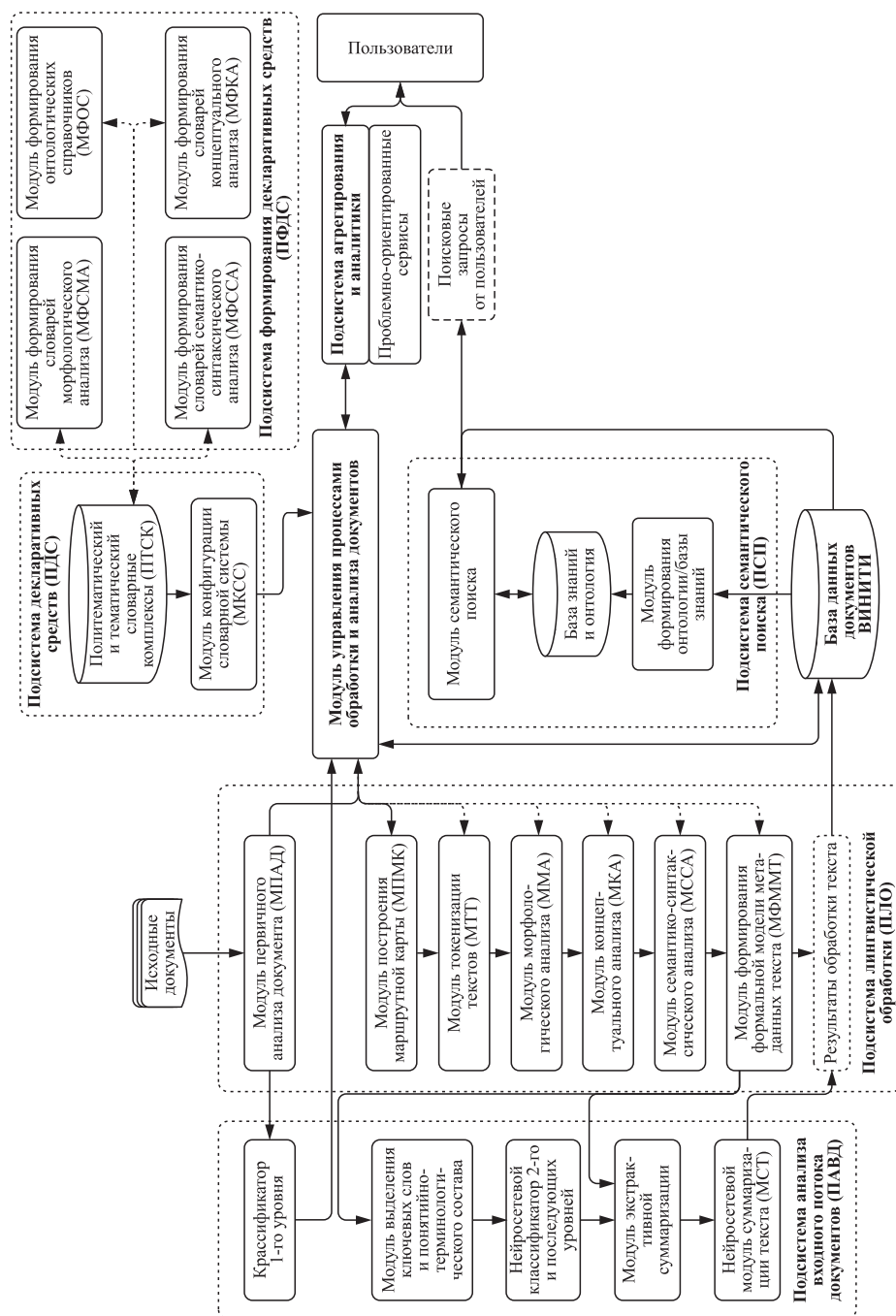
Разработаны методы создания комплекса словарей по отраслевым текстам, описание морфологической, синтаксической, концептуальной моделей, наборов морфологических, синтаксических и концептуальных признаков [3, 6]. Качество работы методов компьютерной лингвистики зависит от полноты и качества словарей и грамматических таблиц, создание и актуализация которых в различных областях науки и техники остается трудоемкой задачей и требует значительной вовлеченности экспертов. Концепция ФКАТ позволяет реализовать технологию взаимосвязанных базовых методов обработки текстов в режиме постоянной актуализации комплекса отраслевых лингвистических ресурсов (словарей). Данная особенность очень важна, она обеспечивает возможность обработки и анализа НТД с отражением новых знаний и новых терминологических понятий.

4 Функциональная архитектура системы семантического анализа текстов

Функциональная архитектура ССА текстов, нацеленная на реализацию концепции ФКАТ, изображена на рисунке.

Архитектура включает состоящие из модулей подсистемы декларативных средств (ПДС), лингвистической обработки текстов (ПЛО), анализа входного потока документов (ПАВД), формирования декларативных средств (ПФДС), семантического поиска (ПСА), агрегирования и аналитики. Функционирование основных модулей системы выглядит следующим образом.

Входящие документы поступают в модуль первичного анализа документа (МПАД), очищаются от ненужной информации, проверяются на наличие орфографических ошибок и приводятся к внутрисистемному формату. В модуле построения маршрутной карты (МПМК) определяются языки документов и их



Функциональная архитектура системы

типы и автоматически формируются маршрутные карты документов, в которых указываются последовательности выполнения технологических операций над документами. В модуле классификации документы классифицируются согласно Государственному рубриктору НТИ (ГРНТИ). Классификация осуществляется с использованием каскада предобученных нейросетевых моделей (это обусловлено большим числом рубрик — около 7 тыс.), для чего предварительно выполняется векторизация текстов документов. После классификации документов по рубрикам 1-го уровня с некоторой периодичностью проводится автоматическое конфигурирование тематических словарных систем для рубрик 1-го уровня (рассматриваемых как отдельные области) в модуле МКСС на основе информации о тематике документов и загрузка словарного комплекса соответствующей тематической области.

Далее в модуле МТТ выполняется токенизация текстов и формирование их графематической модели, в модуле ММА — морфологический анализ и формирование морфологической модели, в модуле МКА — концептуальный анализ и формирование концептуальной модели, в модуле МССА — семантико-синтаксический анализ и формирование ССМ, в модуле МФММ — формирование формальной модели метаданных, из которой выделяются ключевые слова и понятийно-терминологический состав документов (в выделении могут принимать участие эксперты [13]). Полученные данные используются в качестве исходных для нейросетевого модуля 2-го и последующих уровней классификации.

В модуле МСТ осуществляется суммаризация текста — составляется предварительный расширенный реферат документа с использованием экстрактивных методов на основании результатов лингвистического анализа документа, после чего данный текст обрабатывается с применением нейросетевой модели суммаризации. Такой подход помогает решить проблему выделения слишком общего, несодержательного реферата, что характерно для суммаризации с применением языковых моделей.

Для обеспечения семантического поиска по отраслевой коллекции документов отдельный модуль создает отраслевую онтологию и базу знаний на основании информации о семантических отношениях, извлеченных в результате анализа текстов в ПЛО [14]. В модуле семантического поиска предполагается гибридное применение нескольких подходов для повышения качества результатов поиска: поиск по метаданным, полнотекстовый поиск, поиск по онтологии и базе знаний с использованием декларативных запросов на языке SPARQL.

Проблемно-ориентированные сервисы агрегирования и аналитики предназначены для решения широкого спектра задач, перечисленных в разд. 2.

5 Заключение

В статье рассмотрена новая функциональная архитектура ССА НТД. Сама система относится к категории интеллектуальных платформ для агрегации и анализа НТД. Методологическую основу системы составляет концепция ФКАТ,

предполагающая объединение результатов графематического, морфологического, концептуального и семантико-синтаксического анализа текста в ФММ. Одни из наиболее важных компонентов системы — автоматически формируемые и настраиваемые на отрасль знания морфологические, семантико-синтаксические и концептуальные словари, обеспечивающие актуальность комплекса лингвистических ресурсов. Эти особенности отличают предлагаемую архитектуру от известных ССА текстов и нацелены на повышение качества обработки и анализа НТД. Применение относительно компактных наборов отраслевых декларативных средств позволяет надеяться на более высокую эффективность работы, чем в системах, основанных на БЯМ. Применение гибридного семантического поиска, объединяющего поиск по метаданным, полнотекстовый поиск и поиск по онтологии и базе знаний, позволит повысить качество поиска — релевантность результатов запросу.

Дальнейшим естественным шагом представляется эффективная программная реализация предложенной архитектуры в виде специализированного лингвистического процессора. Это позволит экспериментально проверять гипотезы о преимуществах предлагаемой архитектуры в качестве и эффективности обработки и анализа НТД над архитектурами, реализованными в существующих системах. Для проведения экспериментов необходимо также сформировать бенчмарк для сравнительного тестирования систем обработки НТД, оценить качество и производительность разрабатываемой системы на бенчмарке и сравнить с существующими системами и БЯМ по качеству и потребляемым ресурсам. Вместе с тем быстрое распространение БЯМ во многих областях человеческой деятельности — это мотивация для исследования и разработки архитектуры гибридного варианта ССА НТД на основе совместного использования БЯМ и методов компьютерной лингвистики для повышения качества результатов автоматических операций аналитико-синтетической обработки НТИ и работы проблемно-ориентированных сервисов.

Литература

1. Huang L., Yu W., Ma W., Zhong W., Feng Z., Wang H., Chen Q., Peng W., Feng X., Qin B., Liu T. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions // ACM T. Inform. Syst., 2025. Vol. 43. No. 2. Art. 42. 55 p. doi: 10.1145/3703155.
2. Смирнов И. В. Интеллектуальный анализ текстов на основе методов разноразмерной обработки естественного языка. — М.: ФИЦ ИУ РАН, 2023. 356 с. EDN: VCDQEW.
3. Хорошилов Ал-др А., Кан А. В., Хорошилов Ал-ей А. Фразеологический машинный перевод текстов: теоретические основы и технологические решения. — М.: Директ-Медиа, 2019. 466 с. doi: 10.23681/563869. EDN: PYQZHX.
4. Mel'čuk I., Milićević J. An advanced introduction to semantics: A meaning-text approach. — Cambridge University Press, 2020. 450 p.

5. Smirnov I., Stankevich M., Kuznetsova Y., Suworova M., Larionov D., Nikitina E., Savelov M., Grigoriev O. TITANIS: A tool for intelligent text analysis in social media // Artificial intelligence / Eds. S. M. Kovalev, S. O. Kuznetsov, A. I. Panov. — Lecture notes in computer science ser. — Cham: Springer, 2021. Vol. 12948. P. 232–247. doi: 10.1007/978-3-030-86855-0_16. EDN: DQVODN.
6. Ананьева М. И., Девяткин Д. А., Зубарев Д. В., Осипов Г. С., Смирнов И. В., Соченков И., Тихомиров И. А., Швеиц А. В., Шелманов А. О. TextAppliance: поиск и анализ больших массивов текстов // Труды 15-й Национальной конференции по искусственному интеллекту с международным участием. — Смоленск: Универсум, 2016. Т. 3. С. 220–228. EDN: WZPBPB.
7. Turdakov D. Y., Astrakhantsev N. A., Nedumov Y. R., et al. Texterra: A framework for text analysis // Programing Computer Software, 2014. Vol. 40. P. 288–295. doi: 10.1134/S0361768814050090.
8. Lopez P. GROBID: Combining automatic bibliographic data recognition and term extraction for scholarship publications // Research and advanced technology for digital libraries / Eds. M. Agosti, J. L. Borbinha, S. Kapidakis, C. Papatheodorou, G. Tsakonas. — Lecture notes in computer science ser. — Berlin, Heidelberg: Springer, 2009. Vol. 5714. P. 473–474. doi: 10.1007/978-3-642-04346-8_62.
9. Kashyap A. R., Kan M.-Y. SciWING — a software toolkit for scientific document processing // 1st Workshop on Scholarly Document Processing Proceedings. — Association for Computational Linguistics, 2020. P. 113–120. doi: 10.18653/v1/2020.sdp-1.13.
10. Fok R., Kambhamettu H., Soldaini L., Bragg J., Lo K., Hearst M., Head A., Weld D. S. Scim: Intelligent skimming support for scientific papers // 28th Conference (International) on Intelligent User Interfaces Proceedings. — New York, NY, USA: Association for Computing Machinery, 2023. P. 476–490. doi: 10.1145/3581641.3584034.
11. Никитин Ю. В., Хорошилов А. А. Интеллектуальный текстовый процессор // Искусственный интеллект: теория и практика, 2023. Т. 1. Вып. 1. С. 56–75. EDN: KOZSNI.
12. Захаров В. Н., Хорошилов Ал-др А., Хорошилов Ал-ей А. Автоматическое построение синтаксических моделей языка для систем обработки текстовой информации // Системы и средства информатики, 2018. Т. 28. № 4. С. 4–9. doi: 10.14357/08696527180401. EDN: PJMVOD
13. Зацман И. М. Модель процесса извлечения новых терминов и тональных слов из текстов // Системы и средства информатики, 2022. Т. 32. № 2. С. 115–127. doi: 10.14357/08696527220211. EDN: VPHYVNK.
14. Хорошилов А. А., Кан А. В., Филиппов Г. В. Методы автоматизированного формирования комплекса онтологических справочников по авиакосмической тематике // Научно-технический вестник Поволжья. — Казань: Рашин Сайнс, 2024. Вып. 5. С. 150–154. EDN: CPKZTE.

Поступила в редакцию 29.07.2025

Принята к публикации 15.10.2025

AN ARCHITECTURE OF A SYSTEM FOR THE SEMANTIC ANALYSIS OF THE SCIENTIFIC AND TECHNICAL DOCUMENTS

*A. V. Kan^{1,2}, Alexander A. Khoroshilov^{1,2}, I. A. Chechulin¹, S. A. Stupnikov^{1,3},
Yu. V. Nikitin⁴, and Alexey A. Khoroshilov¹*

¹All-Russian Institute for Scientific and Technical Information of the Russian Academy of Sciences, 20 Usievicha Str., Moscow 125190, Russian Federation

²Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation

³Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

⁴National Research Center “Zhukovsky Institute,” 1 Zhukovsky Str., Zhukovsky 140180, Russian Federation

Abstract: The paper considers the functional architecture of the system of semantic analysis of scientific and technical documents. The system belongs to the category of intelligent platforms for the aggregation and analysis of scientific and technical documents. The main technological solution of the system is to create a set of declarative tools adaptable to various domains of science and technology, implementing the concept of phraseological conceptual analysis of texts. This assumes combining the results of graphemic, morphological, conceptual, and semantic and syntactic analysis of texts into a formalized metadata model. The subsystems and modules constituting the architecture are considered, their functions and interaction are described. One of the most important components of the system is a set of morphological, semantic, syntactic, and conceptual dictionaries that are automatically generated and customized to domains of knowledge, ensuring the relevance of a set of linguistic resources. A classification of systems for semantic analysis of texts and the problems they solve are considered. The system being developed is positioned in the classification and its distinctive features are distinguished.

Keywords: phraseological conceptual analysis of texts; semantic analysis of texts; scientific and technological text processing

DOI: 10.14357/08696527250404

EDN: YTYHOU

Acknowledgments

The work was financially supported by the research topic FFFU-2025-0010 of the VINITI RAS.

References

1. Huang, L., W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, and T. Liu. 2025. A survey on hallucination in large language

- models: Principles, taxonomy, challenges, and open questions. *ACM T. Inform. Syst.* 43(2):42. 55 p. doi: 10.1145/3703155.
2. Smirnov, I. V. 2023. *Intellektual'nyy analiz tekstov na osnove metodov raznourovnevoy obrabotki estestvennogo yazyka* [Intelligent text analysis based on multilevel natural language processing methods]. Moscow: FRC CSC RAS. 356 p. EDN: VCDQEW.
 3. Khoroshilov, Al-dr A., A. V. Kan, and Al-ey A. Khoroshilov. 2019. *Frazeologicheskii mashinnyy perevod tekstov: teoreticheskie osnovy i tekhnologicheskie resheniya* [Phraseological machine translation: Theoretical foundations and technical solutions]. Moscow: Direkt-Media. 466 p. doi: 10.23681/563869. EDN: PYQZHX.
 4. Mel'čuk, I., and J. Milićević. 2020. *An advanced introduction to semantics: A meaning-text approach*. Cambridge University Press. 450 p.
 5. Smirnov, I., M. Stankevich, Y. Kuznetsova, M. Suvorova, D. Larionov, E. Nikitina, M. Savelov, and O. Grigoriev. 2021. TITANIS: A tool for intelligent text analysis in social media. *Artificial intelligence*. Eds. S. M. Kovalev, S. O. Kuznetsov, and A. I. Panov. Lecture notes in computer science ser. Cham: Springer. 12948:232–247. doi: 10.1007/978-3-030-86855-0_16. EDN: DQVODN.
 6. Ananieva, M. I., D. A. Devyatkin, D. V. Zubarev, G. S. Osipov, I. V. Smirnov, I. V. Sochenkov, I. A. Tikhomirov, A. V. Shvets, and A. O. Shelmanov. 2016. TextAppliance: poisk i analiz bol'shikh massivov tekstov [TextAppliance: Retrieval and analysis of large text datasets]. *Trudy 15-y Natsional'noy konferentsii po iskusstvennomu intellektu* [15th National Conference on Artificial Intelligence Proceedings] 3:220–228. EDN: WZPBPB.
 7. Turdakov, D. Y., N. A. Astrakhantsev, Y. R. Nedumov, et al. 2014. Textterra: A framework for text analysis. *Programing and Computer Software* 40:288–295. doi: 10.1134/S0361768814050090.
 8. Lopez, P. 2009. GROBID: Combining automatic bibliographic data recognition and term extraction for scholarship publications. *Research and advanced technology for digital libraries*. Eds. M. Agosti, J. L. Borbinha, S. Kapidakis, C. Papatheodorou, and G. Tsakonas. Lecture notes in computer science ser. Berlin, Heidelberg: Springer. 5714:473–474. doi: 10.1007/978-3-642-04346-8_62.
 9. Kashyap, A. R., and M.-Y. Kan. 2020. SciWING — a software toolkit for scientific document processing. *1st Workshop on Scholarly Document Processing Proceedings*. Association for computational linguistics. 113–120. doi: 10.18653/v1/2020.sdp-1.13.
 10. Fok, R., H. Kambhamettu, L. Soldaini, J. Bragg, K. Lo, M. Hearst, A. Head, and D. S. Weld. 2023. Scim: Intelligent skimming support for scientific papers. *28th Conference (International) on Intelligent User Interfaces Proceedings*. New York, NY: Association for Computing Machinery. 476–490. doi: 10.1145/3581641.3584034.
 11. Nikitin, Yu. V., and A. A. Khoroshilov. 2023. Intellektual'nyy tekstovyy protsessor [Intelligent word processor]. *Iskusstvennyy intellekt: teoriya i praktika* [Artificial intelligence: Theory and practice] 1(1):56–75. EDN: KOZSNI.
 12. Zakharov, V. N., Al-dr A. Khoroshilov, and Al-ey A. Khoroshilov. 2018. Avtomaticheskoe postroenie sintaksicheskikh modeley yazyka dlya sistem obrabotki tekstovoy informatsii [Automatic construction of syntactic language models for text processing systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(4):4–9. doi: 10.14357/08696527180401. EDN: PJMVOD.

13. Zatsman, I. 2022. Model' protsessy izvlecheniya novykh terminov i tonal'nykh slov iz tekстов [A model of discovering novel terms and sentiments in texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(2):115–127. doi: 10.14357/08696527220211. EDN: VPHYVVK.
14. Khoroshilov, A. A., A. V. Kan, and G. V. Filippov. 2024. Metody avtomatizirovannogo formirovaniya kompleksa ontologicheskikh spravochnikov po aviakosmicheskoy tematike [Methods for automated generation of a complex of ontological reference books on aerospace topics]. *Nauchno-tekhnicheskiiy vestnik Povolzh'ya* [Scientific and Technical Volga Region Bulletin]. Kazan: Rashin Sayns. 5:150–154. EDN: CPKZTE.

Received July 29, 2025

Accepted October 15, 2025

Contributors

Kan Anna V. (b. 1977) — Candidate of Science (PhD) in technology, associate professor, Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; deputy director, All-Russian Institute for Scientific and Technical Information of the Russian Academy of Sciences, 20 Usievicha Str., Moscow 125190, Russian Federation; kanav@viniti.ru

Khoroshilov Alexander A. (b. 1952) — Doctor of Science in technology, leading scientist, All-Russian Institute for Scientific and Technical Information of the Russian Academy of Sciences, 20 Usievicha Str., Moscow 125190, Russian Federation; professor, Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; khoroshilov@viniti.ru

Chechulin Ivan A. (b. 1988) — deputy head of department, All-Russian Institute for Scientific and Technical Information of the Russian Academy of Sciences, 20 Usievicha Str., Moscow 125190, Russian Federation; chechulin@viniti.ru

Stupnikov Sergey A. (b. 1978) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sstupnikov@frcsc.ru

Nikitin Yuri V. (b. 1977) — head of department, National Research Center “Zhukovsky Institute,” 1 Zhukovsky Str., Zhukovsky 140180, Russian Federation; yuri.v.nikitin@gmail.com

Khoroshilov Alexey A. (b. 1958) — Candidate of Science (PhD) in technology, head of laboratory, All-Russian Institute for Scientific and Technical Information of the Russian Academy of Sciences, 20 Usievicha Str., Moscow 125190, Russian Federation; alex_khor@viniti.ru

ДВИЖЕНИЕ, СКОРОСТЬ И ТРАЕКТОРИИ ПРЕДСТАВЛЕНИЙ КЛЮЧЕВЫХ СЛОВ В ВЕКТОРНОМ ПРОСТРАНСТВЕ ЯЗЫКОВОЙ МОДЕЛИ

М. М. Шарнин¹, Н. В. Сомин²

Аннотация: Предложен метод расчета положений, скоростей и эволюционных траекторий ключевых слов в векторном пространстве статической языковой модели. Семантическое расстояние между векторами слова в моменты времени t_1 и t_2 определяется как косинусное расстояние между этими векторами. Скорость семантических изменений вычисляется как семантическое расстояние, деленное на $t_2 - t_1$. Скорость семантических изменений выражает, насколько быстро меняется значение/семантика слова, его контекст, положение в векторном пространстве и семантически близкие слова. Метод позволяет рассчитывать скорости и эволюционные траектории тем, представляющих совокупность нескольких родственных ключевых слов. Для расчета скоростей и траекторий в анализируемый исходный текст вставляются специальные эволюционные метки рядом со словами из интересующей темы. Рассматривается случай скоростей и траекторий ключевых слов в области «машинного обучения», полученных из библиотеки PubMed. Векторы ключевых слов и их изменения во времени рассчитываются с помощью нейросети Word2Vec. Представлена семантическая карта, позволяющая визуальнo оценить эволюционные траектории и скорости. Она основана на алгоритме PCA (Principal Component Analysis), позволяющем получить проекцию траекторий на двумерную (2D) плоскость.

Ключевые слова: скорость семантических изменений; эволюционные траектории; векторное пространство; статическая языковая модель

DOI: 10.14357/08696527250405

EDN: ZFFAAA

1 Введение

Современные методы анализа текстов дают возможность значительно расширить возможности такой классической науки, как этимология, которая изучает происхождение и изменение смысла слов [1]. Ранее задачи этимологии решались путем визуального анализа текстов и создания обширных каталогов описаний слов [2], а также анализа совместного употребления слов [3], но на современном этапе с помощью нейросетевых лингвистических моделей [4–6] появилась

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, mc@keywen.com

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, chri-soc@yandex.ru

возможность для каждого слова быстро вычислять векторы, содержащие семантические расстояния между данным словом и остальными словами. Набор таких векторов представляет собой семантическое векторное пространство.

Опыт показывает, что для большинства слов семантические расстояния с течением времени изменяются, т. е. слово «движется» в этом пространстве. Это означает, что в определенной степени меняется семантика слова или его коннотация. Чтобы это движение слов оценить количественно, удобно ввести такую характеристику, как скорость слова. В настоящей статье предлагается методика определения этой характеристики.

Чтобы наглядно продемонстрировать семантические изменения терминов, была предложена методика визуализации изменений на двумерной плоскости, представляющей проекцию многомерного семантического векторного пространства. Чтобы разместить различные временные срезы в одном векторном пространстве, был предложен метод эволюционных меток, размещаемых перед ключевыми словами. Этот метод значительно упрощает построение траекторий ключевых слов.

Необходимо сказать несколько слов о работах, связанных со скоростью семантических изменений и эволюционными траекториями слов, концепций и тем. Ранние попытки построить эволюционную траекторию концепций основывались на анализе графов библиографических ссылок [7]. Позже были разработаны методы построения векторных пространств на основе нейросетевых языковых моделей [4–6], что сделало эволюционные траектории ключевых слов еще более наглядными. В работе [4] (2013 г.) авторы сосредоточились на распределенных представлениях слов, полученных с помощью нейронных сетей. Они предложили статическую языковую модель под названием Word2Vec для вычисления непрерывных векторных представлений слов из очень больших наборов текстовых данных. Векторы встраивания (embeddings) этой модели образуют векторное пространство и обеспечивают высокую производительность для измерения синтаксического и семантического сходства слов. Авторы использовали косинусное расстояние между векторами для определения семантического сходства соответствующих слов. Четыре года спустя появилась динамическая версия Word2Vec [8] для текстовых данных с метками времени, которая отслеживает семантическую эволюцию отдельных слов. Модель представляет слова и контексты как скрытые траектории в пространстве векторов встраивания. Эта динамическая модель строила траектории встраивания слов, которые были более интерпретируемы, чем конкурирующие методы, основанные на статических моделях, обученных отдельно на временных срезах.

В работе [9] (2018 г.) использовалась модель Dynamic Word2Vec для визуализации траекторий набора примеров слов во встроенном векторном пространстве. Авторы построили 2D-проекции временных векторов рассматриваемых слов с помощью метода визуализации t-SNE [10]. Они также построили ближайшие слова к целевому слову из каждого временного среза. Авторы отметили, что траектория слова во встроенном пространстве служит инструментом, полезным для понимания изменения значений слов с течением времени.

В работе [11] (2016 г.) авторы использовали модель Word2Vec и предложили два эмпирических количественных закона семантических изменений: (1) скорость семантических изменений обратно пропорциональна частоте употребления слов (закон соответствия); (2) слова с большей полисемичностью демонстрируют более высокую скорость семантических изменений (закон инноваций). В работе использовалось косинусное расстояние между векторами слова в разные моменты времени для количественной оценки скорости семантических изменений. Эти результаты были подтверждены и улучшены в 2020 г. в работе [12], где также использовались модель Word2Vec, косинусное расстояние и данные о частоте слов для улучшения качества обнаружения семантических сдвигов у популярных слов, которые меняются медленнее. При этом отмечалось, что одновременно с быстрыми семантическими изменениями векторов слов заметно растет относительная частота этих слов. В работе [13] (2022 г.) авторы обнаружили, что изменение значения более вероятно для слов, которые усваиваются в жизни позже и которые сложнее для восприятия, а слова, которые усваиваются раньше, имеют меньшую скорость семантических изменений.

Достижения в области искусственного интеллекта и нейронных сетей привели к появлению новых, более мощных лингвистических моделей [14–16]. В настоящее время большие языковые модели активно используются в системах искусственного интеллекта, разрабатываемых крупными компаниями. Так, в [5] в 2019 г. была описана модель BERT (Bidirectional Encoder Representations from Transformers), разработанная Google. Модель BERT была встроена в популярную текстовую поисковую систему компании, что позволило ей лучше понимать поисковые запросы. Для этого модель была обучена на огромном числе текстов (порядка 2500 млн слов), в ней доступны более 100 языков. Информация о динамической модели BERT была опубликована годом позже в [17] (2020 г.). Еще одна языковая модель — GPT (Generative Pre-trained Transformer) [6], выпущенная OpenAI в 2022 г., стала доступна многим пользователям известной среды ChatGPT.

Динамические версии языковых моделей обычно появляются позже статических. Не все статические модели имеют свои динамические версии. Предложенный в данной статье метод позволяет визуально сравнивать качество траекторий эволюции тем и траекторий эволюции ключевых слов, построенных с использованием разных статических моделей. Для такого сравнения можно использовать одни и те же текстовые данные, не изменяя программный код этих моделей.

2 Постановка задачи

Решаемая в статье задача заключается в анализе предложенной характеристики — скорости движения слов, а также в разработке методики ее вычисления. Для этого требуется:

- (1) определить векторы для выбранной группы ключевых слов или тем в различные моменты времени и получить траектории эволюционного движения этих ключевых слов в семантическом векторном пространстве;
- (2) определить семантические расстояния между полученными векторами и на этой основе выявить их семантическую динамику, а также вычислить скоростные характеристики их движения. Семантическое расстояние/сдвиг между векторами слова в моменты времени t_1 и t_2 определяется как косинусное расстояние между этими векторами, т. е. как единица минус косинус угла между векторами. Аналогичное определение семантического расстояния/сдвига на основе косинусного расстояния (cosine distance) используется в работах [4, 11, 12]. При этом скорость движения векторных представлений терминов (скорость семантических изменений, rate of semantic change) определяется как семантическое расстояние, деленное на $t_2 - t_1$. Смысл скорости заключается в том, чтобы измерить, насколько быстро меняются значение/семантика слова, его контекст, положение в векторном пространстве и семантически близкие слова, которые окружают слово в этом пространстве. Согласно известной в лингвистике дистрибутивной гипотезе вы узнаете слово по компании (контексту), которая его окружает (“you shall know a word by the company it keeps” [18]);
- (3) построить сравнительные диаграммы для скорости движения некоторых ключевых слов в различные моменты времени;
- (4) визуализировать на двумерной плоскости ключевые слова, их эволюционные траектории и скорости движения.

3 Описание метода и вычислительных экспериментов

3.1 Текстовый материал и отбор ключевых слов

Для экспериментов была выбрана коллекция PubMed [19] (названия, авторы, год публикации), содержащая информацию о более чем 36 млн статей (36 417 711) по медицине, биологии и смежным наукам по состоянию на начало 2024 г.

Сначала из коллекции PubMed проводилась выборка статей в области машинного обучения по ключу machine learning. Затем в отобранных статьях выявлялись наиболее часто встречающиеся трендовые ключевые слова (однословные или двухсловные). Для автоматизации этих процессов использовались методы, описанные в [20]. Среди ключевых слов находились трендовые (т. е. наиболее перспективные) ключевые слова. Они выявлялись с учетом долгосрочных прогнозов их трендов в PubMed.

3.2 Построение матрицы семантических расстояний

Семантическое сходство и семантические расстояния между выбранными ключевыми словами вычисляются с помощью модели (нейросети) Word2Vec.

Результатом является матрица семантических расстояний $N \times N$ (N — число выбранных ключевых слов), где расстояния выражены как относительные значения в диапазоне от 0 (семантическое совпадение) до 2 (максимальное различие). Строки в матрице расстояний, соответствующие ключевым словам, представляют собой векторы из некоторого семантического векторного пространства. Если семантические расстояния не вычислялись стабильно, то стабильность улучшалась путем многократного обучения языковой модели с последующим усреднением семантических расстояний.

Помимо ключевых слов в матрицу семантических расстояний еще включаются темы. Тема — это группа близких по смыслу ключевых слов. Темы выявляются после обработки текста с помощью нейросети Word2Vec. Для выявления тем используется кластеризация. При расчете траекторий тем эволюционные метки (см. ниже) ставятся перед каждым ключевым словом, относящимся к данной теме. В дальнейшем работа с темами осуществляется аналогично ключевым словам.

3.3 Эволюционные метки

Как указано выше, для объединения разных временных срезов в единое пространство перед выбранными ключевыми словами вставляются эволюционные метки. Каждая метка кодирует ключевое слово и год публикации статьи, содержащей ключевое слово. Семантические расстояния вычисляются для меток вместе с ключевыми словами, что позволяет выявлять эволюционные траектории ключевых слов в семантическом векторном пространстве.

В результате добавления эволюционных меток анализируемый исходный текст изменяется, а семантические расстояния между ключевыми словами в матрице расстояний $N \times N$ могут измениться и исказиться, что может привести к искажениям в расчете позиций/координат меток в семантическом векторном пространстве. На основании экспериментов выяснено, что чем больше эволюционных меток добавляется в текст одновременно, тем больше могут быть искажения. Поэтому, чтобы уменьшить эти искажения, каждая метка добавлялась в исходный текст отдельно и ее координаты вычислялись независимо от других меток.

3.4 Скоростные характеристики движения ключевых слов

Для расчета скорости движения ключевых слов использовались векторные представления ключевых слов, а также векторные представления их эволюционных меток в различные годы. Далее рассчитывалось семантическое расстояние между этими векторами. Полученное значение делилось на число лет. Результаты расчетов для терминов искусственный интеллект (artificial_intelligence), машинное обучение (machine-learning) и интерпретируемый (interpretable) отражены на рис. 1.

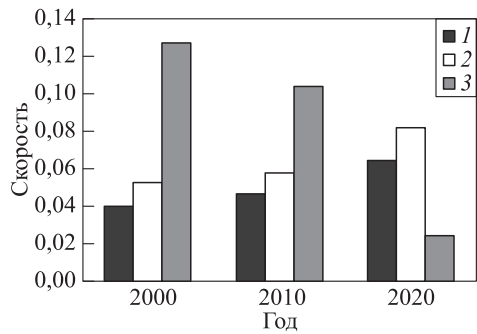


Рис. 1 Скорость движения представлений терминов искусственный интеллект (artificial_intelligence) (1), машинное обучение (machine-learning) (2) и интерпретируемый (interpretable) (3) в векторном пространстве нейросети Word2Vec, обученной по библиографическим данным PubMed

Близкие по семантике термины в разные годы

Год	Artificial intelligence	Machine-learning	Interpretable
2020	explainable XAI chatbots medical_artificial LLM ChatGpt	explainable multimodal_learning multi-classification deep-learning	explainable semisupervised graph-based transformer-based generalizable supervised_learning
2010	explainable logic-based ontology-aware physics-guided	knowledge-guided ontology-aware physics-guided explainable meta-learning	generalizable semisupervised explainable semi-supervised_learning graph-based
2000	explainable knowledge-guided logic-based ontology-driven	knowledge-guided ontology-aware physics-guided bayesian-based	explainable supervised_learning generalizable gaussian_process semisupervised random_forest graph-based

Для рассмотренных выше трех основных терминов и их меток были найдены близкие по семантике (соседние) другие термины, которые были близки в различные годы во время движения по эволюционной траектории. В таблице приведена информация о некоторых близких терминах, которые чаще других оказывались соседями во время движения по траектории.

В таблице представлены повторяемые и близкие по семантике термины в различные годы для терминов artificial intelligence, machine-learning и interpretable. Термины в клетках таблицы упорядочены по мере семантического сходства.

Из таблицы можно сделать следующие выводы. По отношению к термину *artificial intelligence* его близкие термины *knowledge-guided*, *logic-based* и *ontology-driven* уступили по значимости место терминам *chatbots*, *LLM* и *ChatGpt*. Для *machine-learning* термин *explainable* стал более значимым, а термины *knowledge-guided* и *ontology-aware* уступили место *multimodal learning* и *deep-learning*. Для *interpretable* метод контролируемого обучения (*supervised learning*) уступил по значимости место методу полуконтролируемого обучения (*semisupervised*), повысилась роль методов, основанных на графах (*graph-based*), а методы *gaussian process* и *random forest* уступили место методам, основанным на трансформерах (*transformer-based*).

3.5 Семантическая карта

Для визуализации выбранных ключевых слов и тем, а также их эволюционных траекторий на семантической карте необходимо рассчитать соответствующие координаты на 2D-плоскости. Расчет координат ключевых слов и эволюционных меток выполняется с помощью программы PCA, которая редуцирует многомерное семантическое векторное пространство к двум измерениям. Последующая визуализация в виде семантической карты выполняется с помощью библиотеки Python Matplotlib. Пример такой семантической карты показан на рис. 2.

Цвета ключевых слов на семантической карте соответствуют их долгосрочным прогнозам трендов в коллекции PubMed. Наиболее перспективные ключевые слова с самыми длительными трендами выделены красным цветом, средние тренды — синим, а минимальные тренды — черным.

Координаты ключевых слов, тем и меток на семантической карте определяют как проекции соответствующих точек многомерного векторного пространства на двумерную плоскость, определяемую главными направлениями (главными компонентами), вычисленными с помощью алгоритма PCA. Такая семантическая карта помогает визуально идентифицировать кластеры тесно связанных ключевых слов. На рис. 2 можно видеть следующие кластеры ключевых слов, близких по смыслу: слева — заболевания и медицинская терминология; справа — алгоритмы, методы и компьютерная терминология; вверху — ключевые слова с минимальными трендами (черные); внизу — трендовые ключевые слова (красные и синие).

Семантическая карта иллюстрирует скорость семантических изменений и значительные семантические сдвиги интересующих ключевых слов за 24 года с 2000 по 2024 гг. Чем выше скорость, тем больше сдвигается метка ключевого слова на карте и тем сильнее меняются окружающие семантически схожие слова. Например, на рис. 2 показано, что ключевое слово *machine learning* (машинное обучение) имело следующие семантически схожие слова в разные годы: *physics-guided* и *algorithm-generated* с 2000 по 2010 гг., *multiclassification* и *catboost* в 2020 г. и *deep learning* в 2024 г. Таким образом, рис. 2 показывает, что предлагаемый метод позволяет строить эволюционные траектории, которые мо-

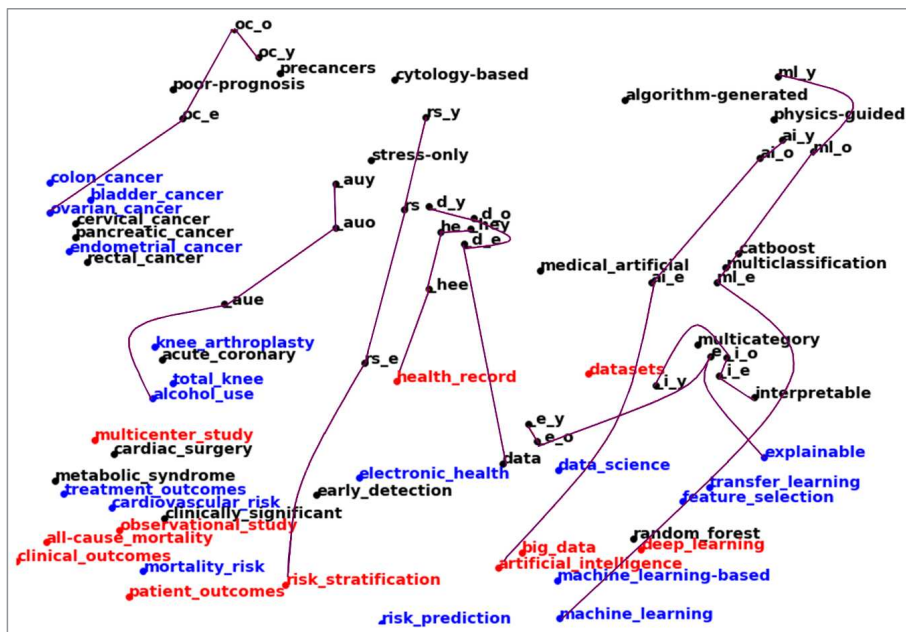


Рис. 2 Визуализация набора ключевых слов и тем с эволюционными траекториями. Здесь эволюционные метки, заканчивающиеся на «у», соответствуют годам публикации 2000 и ранее; заканчивающиеся на «о» соответствуют годам 2010 и ранее; заканчивающиеся на «е» соответствуют годам 2020 и ранее. Позиции самих ключевых слов показаны на 2024 г. Например, метка 'ai_y' заменяет ключевое слово artificial_intelligence, используемое в 2000 и ранее. Фиолетовые линии, соединяющие метки одного и того же ключевого слова, изображают его эволюционную траекторию в семантическом векторном пространстве

гут хорошо фиксировать семантические сдвиги и изменение скорости ключевых слов с течением времени. Также рис. 2 демонстрирует, что траектории разных ключевых слов имеют разные направления, длины и формы.

4 Заключение

Предлагаемый метод может быть полезным при решении целого ряда задач, таких как более точное понимание смысла текстов, выявление похожести между терминами вплоть до их взаимозаменяемости или, наоборот, альтернативности, определение авторства анонимных текстов, и других задач, для решения которых не находилось объективных методов в рамках классической этимологии, причем такой анализ может быть проделан для разных предметных областей

без какой-либо модификации программного обеспечения. Это дает возможность для более углубленного анализа изменений, происходящих в использовании тех или иных понятий в науке, технике, медицине, средствах массовой информации. Среди преимуществ предлагаемого метода картирования эволюционных траекторий можно выделить возможность использования различных современных языковых моделей (Word2Vec, BERT, GPT и др.) без модификации их программ, возможность использования различных алгоритмов маркировки понятий и тем в исходном тексте, а также возможность визуального сравнения качества различных языковых моделей.

Авторы считают, что, развивая предлагаемый подход, можно будет приближаться к автоматическому опознанию таких сложных семантических задач, как выявление окон Овертона [21]. Ведь смысл окон Овертона состоит в том, чтобы в течение достаточно короткого времени расширить значение термина или изменить его коннотацию. У таких терминов обычно быстро меняется семантическое окружение. Предлагаемый в работе метод позволяет отслеживать для терминов именно такие динамические изменения.

Физические понятия движение, скорость и траектория хорошо известны из опыта в материальном мире. В статье на конкретном примере с помощью точных измерений и визуализации продемонстрирована аналогия между соответствующими физическими и информационными процессами, что помогает понять такие сложные понятия, как скорость движения и траектории ключевых слов в векторном пространстве нейросети.

Литература

1. Бурлак С. А., Старостин С. А. Сравнительно-историческое языкознание. — Сер. Высшее профессиональное образование. Языкознание. — М.: Академия, 2005. 432 с. EDN: QRMHLV.
2. Blank A. Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen. — Berlin, Boston: Max Niemeyer Verlag, 1997. 533 p. doi: 10.1515/9783110931600.
3. Gulordava K., Baroni M. A Distributional similarity approach to the detection of semantic change in the Google Books Ngram Corpus // Workshop on Geometrical Models of Natural Language Semantics Proceedings. — Edinburgh, U.K.: Association for Computational Linguistics, 2011. P. 67–71.
4. Mikolov T. Efficient estimation of word representations in vector space. — Cornell University, 2013. arXiv:1301.3781 [cs.CL]. 12 p.
5. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding // Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies Proceedings. — Minneapolis, MN, USA: Association for Computational Linguistics, 2019. Vol. 1. P. 4171–4186.
6. Liu X., Zheng Y., Du Z., et al. GPT understands, too // AI Open, 2024. Vol. 5. P. 208–215. doi: 10.1016/j.aiopen.2023.08.012.

7. *Mina A., Ramlogan R., Tampubolon G., Metcalfe S.* Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge // *Res. Policy*, 2007. Vol. 36. No. 5. P. 789–806. doi: 10.1016/j.respol.2006.12.007.
8. *Bamler R., Mandt S.* Dynamic word embeddings // 34th Conference (International) on Machine Learning Proceedings. — Sydney, Australia: PMLR, 2017. Vol. 70. P. 380–389.
9. *Yao Z., Sun Y., Ding W., et al.* Dynamic word embeddings for evolving semantic discovery // 11th Conference (International) on Web Search and Data Mining Proceedings. — New York, NY, USA: Association for Computing Machinery, 2018. P. 673–681. doi: 10.1145/3159652.3159703.
10. *Maaten L., Hinton G.* Visualizing data using t-SNE // *J. Mach. Learn. Res.*, 2008. Vol. 9. No. 11. P. 2579–2605.
11. *Hamilton W. L., Leskovec J., Jurafsky D.* Diachronic word embeddings reveal statistical laws of semantic change. — Cornell University, 2016. 13 p. arXiv:1605.09096 [cs.CL].
12. *Englhard A., Willkomm J., Schäler M., et al.* Improving semantic change analysis by combining word embeddings and word frequencies // *Int. J. Digital Libraries*, 2020. Vol. 21. No. 3. P. 24–264. doi: 10.1007/s00799-019-00271-6.
13. *Li Y., Siew C. S.* Diachronic semantic change in language is constrained by how people use and learn language // *Mem. Cognition*, 2022. Vol. 50. No. 6. P. 1284–1298. doi: 10.3758/s13421-022-01331-0.
14. *Wu S., Koo M., Blum L., Black A., Kao L., Scalzo F., Kurtz I.* A comparative study of open-source large language models, GPT-4 and Claude 2: Multiple-choice test taking in nephrology. — Cornell University, 2023. 7 p. arXiv:2308.04709 [cs.CL].
15. *Rahman A., Mahir S. H., Tashrif Md T. An, et al.* Comparative analysis based on DeepSeek, ChatGPT, and Google Gemini: Features, techniques, performance, future prospects. — Cornell University, 2025. 20 p. arXiv:2503.04783 [cs.CL].
16. *Moshninov A., Komarov M.* Artificial intelligence in Russia: Evaluating Yandex GPT-5 and GigaChat 2.0 MAX in global context // Conference (International) on Artificial Intelligence Management and Trends Proceedings. — Abu Dhabi, UAE: ADSM, 2025. P. 38–44. doi: 10.63962/URBB4192.
17. *Hofmann V., Pierrehumbert J. B., Schütze H.* Dynamic contextualized word embeddings. — Cornell University, 2020. 15 p. arXiv:2010.12684 [cs.CL].
18. *Firth J.* A synopsis of linguistic theory 1930–1955 // *Studies in linguistic analysis*. — Oxford: Blackwell, 1957. P. 1–32.
19. Коллекция научной литературы PubMed. <https://pubmed.ncbi.nlm.nih.gov>.
20. *Charnine M., Klovov A., Kochiev L., Tishchenko A.* Research trending topic prediction as cognitive enhancement // Conference (International) on Cyberworlds Proceedings. — IEEE, 2021. P. 217–220. doi: 10.1109/CW52790.2021.00044.
21. *Шарнин М. М., Сомин Н. В.* Метод автоматизированной оценки достоверности альтернативных утверждений в коллекции научных статей на примере темы «окна Овертона» // *Искусственный интеллект и принятие решений*, 2024. № 1. С. 118–128. doi: 10.14357/20718594240110. EDN: YJICCC.

Поступила в редакцию 23.06.2025

Принята к публикации 15.10.2025

MOVEMENT, VELOCITY, AND TRAJECTORIES OF KEYWORD REPRESENTATIONS IN THE VECTOR SPACE OF THE LANGUAGE MODEL

M. M. Charnine and N. V. Somin

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: A method for calculating the positions, velocities, and evolutionary trajectories of keywords in the vector space of a static language model is proposed. The semantic distance between word vectors at times t_1 and t_2 is defined as the cosine distance between these vectors. The rate of semantic change is calculated as the semantic distance divided by time ($t_2 - t_1$). The rate of semantic change expresses how quickly the meaning/ semantics of a word, its context, position in the vector space, and semantically close words change. The method allows one to calculate the velocities and evolutionary trajectories of topics representing a set of several related keywords. To calculate the velocities and trajectories, special evolutionary labels are inserted into the analyzed source text next to the words from the topic of interest. The case of the velocities and trajectories of keywords in the field of “machine learning” obtained from the PubMed library is considered. Keyword vectors and their changes over time are calculated using the Word2Vec neural network. A semantic map is presented that allows one to visually assess evolutionary trajectories and velocities. It is based on the PCA (Principal Component Analysis) algorithm which allows one to obtain a projection of trajectories onto a two-dimensional plane.

Keywords: rate of semantic change; evolutionary trajectories; vector space; static language model

DOI: 10.14357/08696527250405

EDN: ZFFAAA

References

1. Burlak, S. A., and S. A. Starostin. 2005. *Sravnitel'no-istoricheskoe yazykoznanie* [Comparative historical linguistics]. Ser. Vysshee professional'noe obrazovanie. Yazykoznanie [Higher professional education ser. Linguistics]. Moscow: Academia. 432 p. EDN: QRMHLV.
2. Blank, A. 1997. *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*. Berlin, Boston: Max Niemeyer Verlag. 533 p. doi: 10.1515/9783110931600.
3. Gulordava, K., and M. Baroni. 2011. A distributional similarity approach to the detection of semantic change in the Google Books Ngram Corpus. *Workshop on Geometrical Models of Natural Language Semantics Proceedings*. Edinburgh, U.K.: Association for Computational Linguistics. 67–71.

4. Mikolov, T. 2013. Efficient estimation of word representations in vector space. Cornell University. 12 p. Available at: <https://arxiv.org/pdf/1301.3781> (accessed November 10, 2025).
5. Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies Proceedings*. Minneapolis, MN: Association for Computational Linguistics. 1:4171–4186.
6. Liu, X., Y. Zheng, Z. Du, *et al.* 2024. GPT understands, too. *AI Open* 5:208–215. doi: 10.1016/j.aiopen.2023.08.012.
7. Mina, A., R. Ramlogan, G. Tampubolon, and S. Metcalfe. 2007. Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Res. Policy* 36(5):789–806. doi: 10.1016/j.respol.2006.12.007.
8. Bamler, R., and S. Mandt. 2017. Dynamic word embeddings. *34th Conference (International) on Machine Learning Proceedings*. Sydney, Australia: PMLR. 70:380–389.
9. Yao, Z., Y. Sun, W. Ding, *et al.* 2018. Dynamic word embeddings for evolving semantic discovery. *11th ACM Conference (International) on Web Search and Data Mining Proceedings*. New York, NY: Association for Computing Machinery. 673–681. doi: 10.1145/3159652.3159703.
10. Maaten, L., and G. Hinton. 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9(11):2579–2605.
11. Hamilton, W. L., J. Leskovec, and D. Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. Cornell University. 13 p. Available at: <https://arxiv.org/pdf/1605.09096> (accessed November 10, 2025).
12. Enghard, A., J. Willkomm, M. Schäler, *et al.* 2020. Improving semantic change analysis by combining word embeddings and word frequencies. *Int. J. Digital Libraries* 21(3):247–264. doi: 10.1007/s00799-019-00271-6.
13. Li, Y., and C. S. Siew. 2022. Diachronic semantic change in language is constrained by how people use and learn language. *Mem. Cognition* 50(6):1284–1298. doi: 10.3758/s13421-022-01331-0.
14. Wu, S., M. Koo, L. Blum, A. Black, L. Kao, F. Scalzo, and I. Kurtz. 2023. A comparative study of open-source large language models, GPT-4 and Claude 2: Multiple-choice test taking in nephrology. Cornell University. 7 p. Available at: <https://arxiv.org/pdf/2308.04709> (accessed November 10, 2025).
15. Rahman, A., S. H. Mahir, Md T. An Tashrif, *et al.* 2025. Comparative analysis based on DeepSeek, ChatGPT, and Google Gemini: Features, techniques, performance, future prospects. Cornell University. 20 p. Available at: <https://arxiv.org/pdf/2503.04783> (accessed November 10, 2025).
16. Moshninov, A., and M. Komarov. 2025. Artificial intelligence in Russia: Evaluating Yandex GPT-5 and GigaChat 2.0 MAX in global context. *Conference (International) on Artificial Intelligence Management and Trends Proceedings*. Abu Dhabi, UAE: ADSM. 38–44. doi: 10.63962/URBB4192.
17. Hofmann, V., J. B. Pierrehumbert, and H. Schütze. 2020. Dynamic contextualized word embeddings. Cornell University. 15 p. Available at: <https://arxiv.org/pdf/2010.12684> (accessed November 10, 2025).

18. Firth, J. 1957. A synopsis of linguistic theory 1930–1955. *Studies in linguistic analysis*. Oxford: Blackwell. 1–32.
19. PubMed biomedical literature database. Available at: <https://pubmed.ncbi.nlm.nih.gov> (accessed November 10, 2025).
20. Charnine, M., A. Klovov, L. Kochiev, and A. Tishchenko. 2021. Research trending topic prediction as cognitive enhancement. *Conference (International) on Cyberworlds Proceedings*. IEEE. 217–220. doi: 10.1109/CW52790.2021.00044.
21. Charnine, M. M., and N. V. Somin. 2024. Metod avtomatizirovannoy otsenki dostovernosti al'ternativnykh utverzheniy v kollektzii nauchnykh statey na primere temy «okna Overtona» [A method for automated assessment of the reliability of alternative statements in a collection of scientific articles using the example of the topic “Overton windows”]. *Iskusstvennyy intellekt i prinyatie resheniy* [Artificial Intelligence and Decision Making] 1:118–128. doi: 10.14357/20718594240110. EDN: YJICCC.

Received June 23, 2025

Accepted October 15, 2025

Contributors

Charnine Mikhail M. (b. 1959) — Candidate of Science (PhD) in technology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; mc@keywen.com

Somin Nikoly V. (b. 1947) — Candidate of Science (PhD) in physics and mathematics, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; chri-soc@yandex.ru

МЕТАЭВРИСТИЧЕСКИЕ АЛГОРИТМЫ ДЛЯ ИДЕНТИФИКАЦИИ ПАРАМЕТРОВ ТРЕХМЕРНОЙ МОДЕЛИ СНИЖЕНИЯ УСТАЛОСТНОЙ ЖЕСТКОСТИ КОМПОЗИЦИОННЫХ МАТЕРИАЛОВ*

А. В. Пантелеев¹, Н. В. Турбин², И. С. Надоров³, Н. О. Кононов⁴

Аннотация: Статья посвящена применению современных метаэвристических алгоритмов оптимизации для определения параметров новых математических моделей, описывающих результаты экспериментов. Предметом исследований выбрано свойство композиционных материалов, используемых в авиастроении, терять жесткость в процессе эксплуатации. Для математического моделирования процесса деградации жесткости предложено использовать систему из трех обыкновенных дифференциальных уравнений, описывающих характерные этапы процесса. Для нахождения параметров правых частей уравнений решаются две задачи условной оптимизации, в которых минимизируются отклонения решений уравнений от известных результатов испытаний композиционных материалов. Решения уравнений находятся классическими явными методами численного интегрирования различного порядка точности. Для параметрической идентификации, т. е. нахождения значений параметров правых частей дифференциальных уравнений по результатам испытаний, предлагается использовать метаэвристические методы оптимизации: модифицированный метод, имитирующий поведение стаи мотыльков, а также метод случайного поиска с последовательной редукцией области исследования. Приведены численные результаты исследования конкретного композиционного материала, иллюстрирующие эффективность предложенного подхода.

Ключевые слова: метаэвристические алгоритмы оптимизации; метод, имитирующий поведение стаи мотыльков; модель поврежденности; композиционные материалы

DOI: 10.14357/08696527250406

EDN: PTRHPN

*Работа выполнена в рамках Соглашения о предоставлении из федерального бюджета гранта на проведение крупных научных проектов по приоритетным направлениям научно-технологического развития Минобрнауки России № 075-15-2024-552.

¹Московский авиационный институт (национальный исследовательский университет), Институт «Компьютерные науки и прикладная математика», avpanteleev@inbox.ru

²Московский авиационный институт (национальный исследовательский университет), Институт «Передовая инженерная школа», turbinnv@mai.ru

³Московский авиационный институт (национальный исследовательский университет), Институт «Компьютерные науки и прикладная математика», nnadorovivan@gmail.com

⁴Московский авиационный институт (национальный исследовательский университет), Институт «Авиационная техника», kononovno@mai.ru

1 Введение

Применение новых технологий в различных областях техники, в частности в проектировании современных летательных аппаратов из композиционных материалов, требует использования не только классических математических методов, но и новых подходов, хорошо зарекомендовавших себя при решении актуальных прикладных задач. Метаэвристические алгоритмы глобальной оптимизации, включающие в себя методы «роевого интеллекта», эволюционные, биоинспирированные, мультиагентные методы, а также методы, имитирующие физические, химические и социальные процессы, интенсивно развиваются в течение последнего десятилетия. Их использование вполне обоснованно, когда для вычисления значения целевой функции требуется реализовать процесс математического моделирования реального явления. В работе рассматривается трехмерная модель процесса роста поврежденности композиционного материала по мере наработки числа циклов нагрузки, который описывается решением системы нелинейных обыкновенных дифференциальных уравнений. Сформулирована и решена задача нахождения параметров правых частей этих уравнений по результатам испытаний композитного материала с целью наилучшего их описания. Для подсчета значений целевой функции, характеризующей точность идентификации параметров математической модели, используются классические явные методы интегрирования обыкновенных дифференциальных уравнений. С целью определения значений параметров правых частей предложено использовать два метаэвристических алгоритма оптимизации: модифицированный метод, имитирующий поведение стаи мотыльков, и метод с последовательной редукцией области исследования.

В качестве прикладной задачи рассматривается проблема определения характерных параметров модельного композиционного материала.

В настоящее время известно множество моделей для оценки усталостной прочности конструкций из полимерных композиционных материалов (ПКМ). Поскольку накопление усталостной поврежденности в ПКМ сопровождается потерей жесткости, то широкое распространение получили различные методики расчетной оценки долговечности по остаточной жесткости [1–4]. Определение рабочих характеристик ПКМ для подобного расчета ставилось задачей специального эксперимента [5]. В [1] предложена новая модель — прогрессирующее усталостное повреждение, расчетные уравнения которой позволяют аппроксимировать кривые остаточной прочности и жесткости. Для оценки разрушения материала рассматриваются феноменологические меры усталостного разрушения в нескольких направлениях с использованием критериев, основанных на напряжениях. В [6] разработана модель остаточной жесткости, которая хорошо согласуется с экспериментальными данными и не требует большого объема исходных данных. Для построения кривых остаточной жесткости необходимо регистрировать изменение модуля упругости конструкций из ПКМ в процессе циклических испытаний. В [5] разработан метод оценки изменения модуля упру-

гости в процессе циклических испытаний на основе регистрации полей деформаций, соответствующих разным циклам нагрузки. В [7, 8] описаны современные математические модели и методы анализа ПКМ.

При этом важной проблемой остается формирование и анализ математических моделей, описывающих процессы деградации жесткости в композиционных материалах. Основой таких моделей служат уравнения, аппроксимирующие кривые остаточной прочности или остаточной жесткости. Для использования моделей необходимо подбирать коэффициенты аппроксимации, входящие в расчетные уравнения, так, чтобы решения уравнений соответствовали данным, полученным в результате испытаний.

Одна из возможных моделей, предложенная в [2], содержит 5 неизвестных параметров, характеризующих усталостное снижение жесткости материала, но она описывает процессы только вдоль одного из направлений. Для этой модели задача идентификации параметров решена в [9]. Однако при исследовании ПКМ, используемых, в частности, в конструкциях самолетов, требуется анализировать процессы деградации жесткости по трем направлениям: продольному, поперечному и сдвиговому. В [2] предложена гипотеза описания такой модели.

В статье предлагается сформировать трехмерную модель деградации жесткости на примере модельного композиционного материала углепластика на основе однонаправленных лент с укладками $[0]n$, $[90]n$ и $[\pm 45]n$ с характерными для этого типа материала механическими и прочностными характеристиками и исследовать ее поведение путем математического моделирования, т. е. решить прямую задачу анализа. Одновременно предложить решение и обратной задачи, т. е. проблемы нахождения коэффициентов правой части трехмерной математической модели, описываемой системой из трех обыкновенных дифференциальных уравнений первого порядка, по результатам испытаний. При этом с целью верификации модели предполагается, что они соответствуют решению прямой задачи.

Задача нахождения параметров модели формулируется как задача поиска условного минимума целевой функции в условиях, когда на параметры наложены интервальные ограничения, характеризующие степень неопределенности их задания. Целевая функция характеризует меру отклонения поведения математической модели от результатов испытаний модельного материала в процессе циклического нагружения.

Для решения сформулированной задачи оптимизации могут быть применены как классические алгоритмы, так и относящиеся к метаэвристическим [10–12]. В качестве основного алгоритма выбрана модификация алгоритма, имитирующего поведение стаи мотыльков (Moth-Flame Optimization, MFO) [13], которая успешно применялась для решения задачи параметрической идентификации одномерной модели в [9]. Метод относится к группе биоинспирированных методов глобальной оптимизации [14, 15]. Результаты, полученные модифицированным методом мотыльков, улучшались последующим применением метода случайного поиска с последовательной редукцией области исследования (метода Luus–Jaakola) [16]. В итоге получены значения параметров трехмерной математической

модели, которые соответствуют заданным свойствам исследуемого композиционного материала.

2 Формирование трехмерной математической модели изменения усталостной жесткости

Будем использовать следующие основные обозначения.

Число циклов $[n]$, в котором один цикл: исходное состояние $\rightarrow$ напряженное состояние (растяжение) $\rightarrow$ исходное состояние, $n \in \{0, \dots, N\}$, где N — общее число выполненных испытаний. Напряжение $[\sigma, \text{МПа}]$ — физическая величина, характеризующая действие внутренних сил упругости в деформированном твердом теле. Напряжение разрушения $[X_T, \text{МПа}]$ — физическая величина, характеризующая напряжение, при котором происходит разрушение образца. Модуль жесткости или модуль упругости $[E, \text{МПа}]$ характеризует способность материала сопротивляться растяжению, сжатию при упругой деформации. Поврежденность $[D]$ — величина, характеризующая падение модуля упругости E образца.

Текущее значение поврежденности связано с текущим значением модуля упругости:

$$D(n) = 1 - \frac{E(n)}{E_0}, \quad (1)$$

где $E(n)$ — текущий модуль упругости, характеризующий способность материала сопротивляться растяжению/сжатию при упругой деформации после цикловой нагрузки; E_0 — исходный модуль упругости — константа, характеризующая способность материала сопротивляться растяжению, сжатию при упругой деформации; n — порядковый номер цикла нагружения. Поскольку $E(n) \in [0, E_0]$, то поврежденность $D(n) \in [0, 1]$.

Если известно значение поврежденности, то текущий модуль упругости находится по формуле, следующей из (1):

$$E(n) = E_0[1 - D(n)].$$

Индекс усталостного разрушения — переменная величина, характеризующая усталостное состояние образца [2]:

$$\Sigma(n) = \frac{\sigma}{[1 - D(n)]X_T}.$$

В [2] предложена одномерная дифференциальная модель величины скорости роста поврежденности образца:

$$\frac{dD(n)}{dn} = c_1 \Sigma(n) \exp \left[-\frac{c_2 D(n)}{\sqrt{\Sigma(n)}} \right] + c_3 D(n) \Sigma^2(n) [1 + \exp \{c_5 [\Sigma(n) - c_4]\}] \quad (2)$$

с начальным условием $D(0) = 0$. Она основана на опыте обработки результатов испытаний композитных материалов. Правая часть обыкновенного дифференциального уравнения (2) зависит от пяти параметров, определяющих процесс изменения поврежденности на характерных участках. Коэффициент c_1 характеризует скорость роста в режиме инициирования повреждений (на первом этапе), т. е. резкое начальное снижение кривой уменьшения модуля упругости. Коэффициент c_2 должен быть достаточно большим по величине, чтобы первое слагаемое по величине достаточно быстро убывало с ростом поврежденности. Коэффициент c_3 характеризует скорость роста в режиме распространения повреждений, т. е. когда происходит постепенное уменьшение модуля упругости. Коэффициенты c_4 и c_5 характеризуют этап взрывного роста поврежденности на третьем этапе.

По аналогии с (2) предлагается рассмотреть трехмерную модель на примере образцов с укладками $[0]n$, $[90]n$ и $[\pm 45]n$ из модельного ПКМ, армированного углеволокном, позволяющую исследовать процесс изменения поврежденности по трем направлениям: продольному, поперечному и сдвиговому.

Приведенные далее дифференциальные уравнения модели описывают скорость роста поврежденности в материале по соответствующему направлению. Характер кривой изменения поврежденности для каждого направления можно описать тремя этапами [2], стыкующимися друг с другом по непрерывности:

- (1) этап инициирования повреждения, на котором происходит уменьшение модуля упругости (по экспоненциальной зависимости);
- (2) этап постепенной деградации жесткости (по линейному закону);
- (3) этап с резким снижением модуля упругости, которое в итоге приводит к разрушению материала (согласно экспоненциальной зависимости, влияющей на процесс после проведения достаточно большого числа циклов нагружения).

Три слагаемых в правых частях уравнений описывают каждый из этапов, аппроксимируя скорость роста повреждаемости с помощью параметрической зависимости от величины коэффициентов $c_1, \dots, c_5$.

Направление 1–1. Дифференциальная модель имеет вид [2]:

$$\frac{dD_{11}(n)}{dn} = c_{1,11} \Sigma_{11}(n) \exp \left[-\frac{c_{2,11} D_{11}(n)}{\sqrt{\Sigma_{11}(n)}} \right] + c_{3,11} D_{11}(n) \Sigma_{11}^2(n) [1 + \exp \{c_{5,11} [\Sigma_{11}(n) - c_{4,11}]\}], \quad D_{11} = 0, \quad (3)$$

где $\Sigma_{11}(n) = \sigma_{11}/([1 - D_{11}(n)]X_T)$ при $\sigma_{11} \geq 0$.

Решение прямой задачи (задачи Коши (3)) находится с учетом значений параметров исследуемого модельного композитного материала [17]:

$E_0^{11} = 142\,000$ МПа — исходный модуль упругости в направлении 1–1;

$X_T = 2314$ МПа — напряжение разрушения в направлении 1–1;

$\sigma_{11} = 1619,8$ МПа — напряжение цикла (как правило, 70% от X_T);
 $c_{1,11} = 8,1 \cdot 10^{-5}$; $c_{2,11} = 673$; $c_{3,11} = 3 \cdot 10^{-6}$; $c_{4,11} = 0,856$; $c_{5,11} = 30$.

Направление 2–2. Дифференциальная модель имеет вид [2]:

$$\frac{dD_{22}(n)}{dn} = c_{1,22} [1 + D_{12}^2(n)] \Sigma_{22}(n) \exp \left[-\frac{c_{2,22} D_{22}(n)}{[1 + D_{12}^2(n)] \sqrt{\Sigma_{22}(n)}} \right] + \\ + c_{3,22} D_{22}(n) \Sigma_{22}^2(n) [1 + \exp \{c_{5,22} [\Sigma_{22}(n) - c_{4,22}]\}], \quad D_{22}(0) = 0, \quad (4)$$

где $\Sigma_{22}(n) = \sigma_{22}/([1 - D_{22}(n)]Y_T)$ при $\sigma_{22} \geq 0$.

Направление 1–2. Дифференциальная модель имеет вид [2]:

$$\frac{dD_{12}(n)}{dn} = c_{1,12} [1 + D_{12}^2(n)] \Sigma_{12}(n) \exp \left[-\frac{c_{2,12} D_{12}(n)}{[1 + D_{12}^2(n)] \sqrt{\Sigma_{12}(n)}} \right] + \\ + c_{3,12} D_{12}(n) \Sigma_{12}^2(n) [1 + \exp \{c_{5,12} [\Sigma_{12}(n) - c_{4,12}]\}], \quad D_{12}(0) = 0, \quad (5)$$

где $\Sigma_{12}(n) = \tau_{12}/([1 - D_{12}(n)]S_{12})$ при $\tau_{12} \geq 0$. Заметим, что традиционно модуль сдвига обозначается G_0^{12} , а не E_0^{12} . Поскольку в статье применяются формулы из [2], то используются нотации автора, разработавшего используемую в статье математическую модель. Кроме того, справедливы соотношения, аналогичные (1):

$$D_{11}(n) = 1 - \frac{E_{11}(n)}{E_0^{11}}; \quad D_{12}(n) = 1 - \frac{E_{12}(n)}{E_0^{12}}; \quad D_{22}(n) = 1 - \frac{E_{22}(n)}{E_0^{22}}.$$

Решение прямой задачи (задачи Коши (4)–(5)) находится с учетом значений параметров исследуемого модельного материала [17]:

$E_0^{22} = 8400$ МПа — исходный модуль упругости в направлении 2–2;
 $Y_T = 38$ МПа — напряжение разрушения в направлении 2–2;
 $\sigma_{22} = 19$ МПа — напряжение цикла (50% от Y_T);
 $c_{1,22} = 1,9 \cdot 10^{-5}$; $c_{2,22} = 492$; $c_{3,22} = 1,8 \cdot 10^{-6}$; $c_{4,22} = 0,765$; $c_{5,22} = 2,01$;
 $E_0^{12} = 4380$ МПа — исходный модуль упругости в направлении 1–2;
 $S_{12} = 148$ МПа — напряжение разрушения в направлении 1–2;
 $\tau_{12} = 74$ МПа — напряжение цикла (50% от S_{12});
 $c_{1,12} = 3 \cdot 10^{-6}$; $c_{2,12} = 1$; $c_{3,12} = 3 \cdot 10^{-5}$; $c_{4,12} = 0$; $c_{5,12} = 0$.

В результате имеем систему обыкновенных дифференциальных уравнений

$$\left. \begin{aligned}
 \frac{dD_{11}(n)}{dn} &= c_{1,11} \frac{\sigma_{11}}{[1 - D_{11}(n)] X_T} \exp \left[- \frac{c_{2,11} D_{11}(n)}{\sqrt{\sigma_{11}/([1 - D_{11}(n)] X_T)}} \right] + \\
 &+ c_{3,11} D_{11}(n) \left[\frac{\sigma_{11}}{[1 - D_{11}(n)] X_T} \right]^2 (n) \left[1 + \right. \\
 &+ \left. \exp \left\{ c_{5,11} \left[\frac{\sigma_{11}}{[1 - D_{11}(n)] X_T} - c_{4,11} \right] \right\} \right], \quad D_{11}(0) = 0; \\
 \frac{dD_{12}(n)}{dn} &= c_{1,12} [1 + D_{12}^2(n)] \frac{\tau_{12}}{[1 - D_{12}(n)] S_{12}} \times \\
 &\times \exp \left[- \frac{c_{2,12} D_{12}(n)}{[1 + D_{12}^2(n)] \sqrt{\tau_{12}/([1 - D_{12}(n)] S_{12})}} \right] + \\
 &+ c_{3,12} D_{12}(n) \left[\frac{\tau_{12}}{[1 - D_{12}(n)] S_{12}} \right]^2 \left[1 + \right. \\
 &+ \left. \exp \left\{ c_{5,12} \left[\frac{\tau_{12}}{[1 - D_{12}(n)] S_{12}} - c_{4,12} \right] \right\} \right], \quad D_{12}(0) = 0; \\
 \frac{dD_{22}(n)}{dn} &= c_{1,22} [1 + D_{12}^2(n)] \frac{\sigma_{22}}{[1 - D_{22}(n)] Y_T} \times \\
 &\times \exp \left[- \frac{c_{2,22} D_{22}(n)}{[1 + D_{12}^2(n)] \sqrt{\sigma_{22}/([1 - D_{22}(n)] Y_T)}} \right] + \\
 &+ c_{3,22} D_{22}(n) \left[\frac{\sigma_{22}}{[1 - D_{22}(n)] Y_T} \right]^2 \left[1 + \right. \\
 &+ \left. \exp \left\{ c_{5,22} \left[\frac{\sigma_{22}}{[1 - D_{22}(n)] Y_T} - c_{4,22} \right] \right\} \right], \quad D_{22}(0) = 0.
 \end{aligned} \right\} \quad (6)$$

Требуется найти законы изменения $D_{11}(n)$, $D_{12}(n)$ и $D_{22}(n)$ на промежутке $n = [0, \dots, N = 100\,000]$.

Для решения полученной задачи Коши (прямой задачи) можно применять как явные, так и неявные численные методы интегрирования обыкновенных дифференциальных уравнений различных порядков точности.

3 Параметрическая идентификация трехмерной модели по заданному поведению значений поврежденности

Эта задача — обратная по отношению к прямой.

Поскольку правая часть уравнения, описывающего изменение $D_{11}(n)$, не зависит от $D_{12}(n)$ и $D_{22}(n)$, то коэффициенты $c_{1,11}$, $c_{2,11}$, $c_{3,11}$, $c_{4,11}$ и $c_{5,11}$ могут быть найдены независимо в результате решения задачи оптимизации

$$I_1 = \sum_{n=0}^K \left(D_{11}(n) - \hat{D}_{11}(n) \right)^2 \rightarrow \min_{c_{1,11}, \dots, c_{5,11}}, \quad (7)$$

где $D_{11}(\cdot)$ — поврежденность на направлении 1–1, полученная в результате решения прямой задачи; $\hat{D}_{11}(\cdot)$ — поврежденность на направлении 1–1, полученная в результате интегрирования дифференциального уравнения (3) с параметрами $c_{1,11}$, $c_{2,11}$, $c_{3,11}$, $c_{4,11}$ и $c_{5,11}$, удовлетворяющими интервальным ограничениям.

Коэффициенты $c_{1,12}$, $c_{2,12}$, $c_{3,12}$, $c_{4,12}$, $c_{5,12}$ и $c_{1,22}$, $c_{2,22}$, $c_{3,22}$, $c_{4,22}$, $c_{5,22}$ находятся в результате решения задачи оптимизации

$$I_2 = \sum_{n=0}^K \left[\lambda_1 \left(D_{12}(n) - \hat{D}_{12}(n) \right)^2 + \lambda_2 \left(D_{22}(n) - \hat{D}_{22}(n) \right)^2 \right] \rightarrow \min_{c_{1,12}, \dots, c_{5,12}, c_{1,22}, \dots, c_{5,22}}, \quad (8)$$

где $D_{12}(\cdot)$ — поврежденность на направлении 1–2, полученная в результате решения прямой задачи; $\hat{D}_{12}(\cdot)$ — поврежденность на направлении 1–2, полученная в результате интегрирования дифференциального уравнения (4) с параметрами $c_{1,12}$, $c_{2,12}$, $c_{3,12}$, $c_{4,12}$ и $c_{5,12}$, удовлетворяющими интервальным ограничениям; $D_{22}(\cdot)$ — поврежденность на направлении 2–2, полученная в результате решения прямой задачи; $\hat{D}_{22}(\cdot)$ — поврежденность на направлении 2–2, полученная в результате интегрирования дифференциального уравнения (5) с параметрами $c_{1,22}$, $c_{2,22}$, $c_{3,22}$, $c_{4,22}$ и $c_{5,22}$, удовлетворяющими интервальным ограничениям; λ_1 и λ_2 — весовые коэффициенты (параметры линейной свертки критериев), задаваемые исходя из сравнительной важности первого и второго слагаемых.

Верхний предел суммирования находится из соотношения $K = \min\{K_{11}, K_{12}, K_{22}\}$, где K_{ij} — последний аргумент $D_{ij}(\cdot)$, при котором функция определена, т.е. процесс интегрирования и вычисления критериев заканчивается при достижении какой-либо из характеристик $D_{11}(n)$, $D_{12}(n)$, $D_{22}(n)$ окрестности значения, равного 1. При этом процесс интегрирования завершается и выдаются итоговые результаты.

Множество допустимых решений D задается интервальными ограничениями:

$$\begin{aligned} c_{1,11} &\in [0, 10^{-3}], \quad c_{2,11} \in [0, 1000], \quad c_{3,11} \in [-10^{-3}, 10^{-3}], \\ &\quad c_{4,11} \in [0, 1], \quad c_{5,11} \in [0, 50]; \\ c_{1,12} &\in [0, 10^{-5}], \quad c_{2,12} \in [0, 1000], \quad c_{3,12} \in [-10^{-4}, 10^{-4}], \\ &\quad c_{4,12} \in [0, 1], \quad c_{5,12} \in [0, 50]; \\ c_{1,22} &\in [0, 10^{-5}], \quad c_{2,22} \in [0, 1000], \quad c_{3,22} \in [-10^{-4}, 10^{-4}], \\ &\quad c_{4,22} \in [0, 1], \quad c_{5,22} \in [0, 50]. \end{aligned}$$

При выборе интервальных ограничений авторы использовали информацию о решении прямых задач (3) и (4)–(5), а также опорные значения из табл. 6.1 в [2]. В ходе решения задачи оптимизации значения левой и правой границ промежутков итерационно уточнялись с целью получения более точного ответа. При обработке реальных результатов эксперимента рекомендуется начинать с отрезков большей длины, а затем после получения результатов решения задач оптимизации последовательно уменьшать длины промежутков с целью достижения большей точности и сокращения затрат времени на поиск экстремума.

4 Модифицированный метод, имитирующий поведение стаи мотыльков

Рассматривается задача нахождения глобального условного минимума функции $f(x)$ на множестве допустимых решений D , т.е. такой точки $x^* \in D$, в которой $f(x^*) = \min_{x \in D} f(x)$, где $x = (x_1, x_2, \dots, x_n)^T$, $D = \{x | x_i \in [a_i, b_i], i = \overline{1, n}\}$.

В качестве целевых функций рассматриваются критерии (7), (8), характеризующие точность описания явления деградации жесткости с помощью описанных математических моделей.

Метод мотыльков и источников света (MFO) имитирует поведение мотыльков при их приближении к источникам света. Каждый мотылек из популяции мотыльков реализует движение по спирали к центру — источнику света [13, 18]. В начале создается случайный набор из N_p точек (положений мотыльков) в множестве допустимых решений, который можно представить следующей матрицей:

$$M(\text{Moth}) = \begin{pmatrix} m_1^1 & \dots & m_n^1 \\ \vdots & \ddots & \vdots \\ m_1^{N_p} & \dots & m_n^{N_p} \end{pmatrix},$$

где m_i^j — значение i -й координаты j -го мотылька, $j = \overline{1, N_p}$, $i = \overline{1, n}$.

Для всех текущих положений мотыльков $m^j = (m_1^j, \dots, m_n^j)^T$ вычисляются значения целевой функции $OM^j = f(m^j)$, $j = \overline{1, N_p}$. В результате формируется вектор $OM = (OM^1 \ OM^2 \ \dots \ OM^{N_p})^T$.

Помимо мотыльков в решении задачи играют роль источники света — лучшие положения мотыльков среди всех ранее занятых с точки зрения достижения наименьшего значения целевой функции. Поэтому число источников света тоже равно N_p . Аналогично мотылькам формируется матрица координат источников света и вектор значений целевой функции, упорядоченный по возрастанию (наилучший источник света имеет номер 1):

$$F(\text{Flame}) = \begin{pmatrix} f_1^1 & \dots & f_n^1 \\ \vdots & \ddots & \vdots \\ f_1^{\text{Np}} & \dots & f_n^{\text{Np}} \end{pmatrix}, \quad \text{OF} = \begin{pmatrix} \text{OF}^1 \\ \text{OF}^2 \\ \vdots \\ \text{OF}^{\text{Np}} \end{pmatrix},$$

где f_i^j — значение i -й координаты j -го источника света, $j = \overline{1, \text{Np}}$, $i = \overline{1, n}$; OF^j — значение целевой функции в точке $f^j = (f_1^j, \dots, f_n^j)^T$, $j = \overline{1, \text{Np}}$, соответствующей положению j -го источника света.

В процессе поиска мотыльки меняют свое положение: двигаются по спирали к источникам света по каждой координате. Спираль должна удовлетворять трем требованиям: начальная точка спирали начинается с текущего положения мотылька; в полюсе спирали расположен источник света; точки спирали не должны выходить за пределы множества допустимых решений D . В качестве такой спирали можно взять логарифмическую спираль:

$$S(m^j, f^j) = H^j e^{st} \cos(2\pi t) + f^j,$$

где m^j — положение j -го мотылька; f^j — положение j -го источника света; $H^j = |f^j - m^j|$ — вектор покоординатных расстояний между j -м источником света и j -м мотыльком; s — константа для определения формы спирали; $t \in [r-1, 0]$ — случайное число, определяющее положение мотыльков относительно спирали, задаваемое равномерным законом распределения. Число r — константа, значение которой уменьшается с ростом числа итераций по линейному закону от -1 до -2 .

Модификация метода MFO (Modified MFO, MMFO) заключается в рассмотрении сразу нескольких точек на спирали (их число τ задается пользователем), имитирующих передвижение мотылька к источнику света, в процессе которого реализуется стратегия поиска экстремума [18]. Если одна из таких точек окажется за пределами множества допустимых решений, вместо нее можно рассматривать случайную точку в пределах D . В качестве следующего положения мотылька следует выбирать то положение, в котором достигается наименьшее значение целевой функции.

Продвижение мотыльков обеспечивает поиск наилучшего решения среди генерированных на спирали допустимых решений, а источники света постоянно меняют свое положение. Это происходит следующим образом: рассмотрим положение первого мотылька, сравнивая значения целевой функции в точке его положения и в точке положения последнего источника света (в начале процесса поиска он имеет номер Np). Если выполняется условие $\text{OM}^1 < \text{OF}^{\text{last}}$, т. е. положение мотылька лучше, то за положение последнего источника света принимается текущее положение первого мотылька: $f^{\text{last}} = m^1$ и заменяется значение целевой функции последнего источника света на соответствующее значение для мотылька $\text{OF}^{\text{last}} = \text{OM}^1$. Затем следует упорядочить координаты вектора положений источников света по возрастанию значений целевой функции. При компьютерной

реализации процедуры упорядочения предлагается использовать быструю сортировку Хоара. Описанная последовательность действий повторяется для всех положений мотыльков.

Чтобы мотыльки не находили быстро локальные минимумы и не прекращали поиск, следует связать каждого мотылька ровно с одним источником света: первый летит по спирали к первому источнику, второй — ко второму, последний — к последнему. Поскольку источники постоянно меняются и сортируются, произойдет обеспечение тщательного исследования множества допустимых решений. Однако при длительном поиске этот процесс может ослабить целенаправленное движение к минимуму. Чтобы исключить такую возможность, число источников света сокращается постепенно по следующей формуле:

$$No_f = \text{round} \left(N_p - k \frac{N_p - 1}{T} \right),$$

где $\text{round}(x)$ — округление до ближайшего целого числа; k — номер текущей итерации; N_p — первоначальное число источников света; T — максимальное число итераций, задаваемое исходя из требований достижения баланса между достаточной точностью нахождения экстремума и используемым временем расчетов.

Сокращение числа источников происходит путем удаления из рассмотрения худшего на данный момент источника, т. е. последнего в векторе OF. Мотылек, ранее связанный с этим источником, теперь связывается с источником, бывшим ранее предпоследним. Таким образом достигается баланс между обширным исследованием множества допустимых решений и движением к минимуму.

Процесс поиска оканчивается при достижении максимального числа итераций или заданного предельного времени работы алгоритма.

5 Результаты моделирования процесса поврежденности при заданных значениях характерных параметров модели

Интегрирование уравнений системы (6) проводилось численно одним из следующих методов: методом Эйлера, методом Эйлера–Коши, методом Адамса второго и четвертого порядка с постоянной величиной шага, а также их модификациями с неравномерным шагом [9].

Установлено, что во многих случаях интегрирование не может быть до конца проведено (до $n = 100\,000$), так как значения поврежденности выходят за пределы допустимых значений. Это может происходить по нескольким причинам: использования слишком большого шага интегрирования, свойств численных методов, использования значений коэффициентов, которые порождают быстрое возрастание значения поврежденности. Кроме того, применение многошаговых методов интегрирования (Эйлера–Коши и Адамса) в силу необходимости предва-

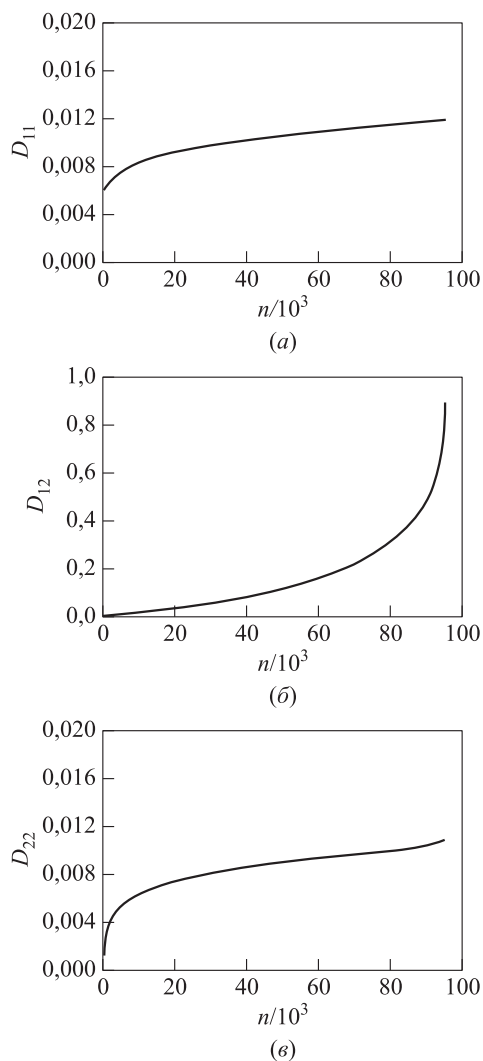


Рис. 1 Поврежденность на направлениях 1–1 (а), 1–2 (б) и 2–2 (в), полученная при решении прямой задачи методом Эйлера с шагом 100

помощи параметра η), а далее снова выполняется заданное число итераций. Процесс завершается после реализации заданного числа проходов, при выполнении условий неэффективности поиска (одновременного недостижения точности ε_1 по величине модуля приращения величины целевой функции и точности ε_2 велич-

рительного нахождения разгонных точек приводит к значительным колебаниям решений в начале процесса.

На рис. 1 приведен один из примеров решения прямой задачи (для получения непрерывной зависимости применялась линейная интерполяция вычисленных значений поврежденности).

6 Решение задачи параметрической идентификации трехмерной модели по заданному поведению значений поврежденности

При решении задачи оптимизации первым применялся метод ММФО, а его результат улучшался с использованием метода Luus–Jaakola [16].

Стратегия метода Luus–Jaakola заключается в выполнении следующих процедур: задается начальная точка на множестве допустимых решений; из нее генерируются R точек в случайных направлениях с учетом характерных размеров множества с проверкой принадлежности ему; среди полученных точек выбирается наилучшая, и из нее процесс продолжается. При этом размер множества поиска сокращается (при помощи параметра γ) от итерации к итерации вплоть до достижения ими заданного числа ITER; как только заданное число итераций выполнено, завершается «проход»; при переходе к следующему проходу размер множества поиска восстанавливается (при помощи параметра η), а далее снова выполняется заданное число итераций.

ны относительного приращения вектора решения), по истечении выделенного на решение времени.

Для решения задачи (7) на описанном выше множестве допустимых решений D применялся метод MMFO с параметрами: $N_p = 3000$, $s = 0,8$, $T = 1\,000\,000$, $\tau = 500$ при ограничении на время выполнения — 5 мин. Для уточнения результата применялся метод Luus-Jaakola с параметрами: $R = 100$, $P = 100$, $ITER = 100$, $\gamma = 0,8$, $\eta = 0,9$, $\varepsilon_1 = \varepsilon_2 = 10^{-50}$ при ограничении на время выполнения — 30 с. В качестве начальной точки для метода Luus-Jaakola использовался результат применения MMFO. Метод интегрирования — метод Эйлера с шагом 100. Результаты решения задачи (7) представлены на рис. 2.

Полученные в результате последовательного применения двух методов оптимизации значения коэффициентов имеют вид:

$$\begin{aligned} c_{1,11} &= 8,0955023112103 \cdot 10^{-5}; \\ c_{2,11} &= 672,50512114333; \\ c_{3,11} &= 1,15473098168472 \cdot 10^{-6}; \\ c_{4,11} &= 0,686408646360375; \\ c_{5,11} &= 22,9051597306067, \end{aligned}$$

а значение критерия (7):

$$I_1 = 4,15071314385573 \cdot 10^{-10}.$$

Заметим, что решения прямой и обратной задач практически не различимы между собой.

Для решения задачи (8) процесс начинается с задания весовых коэффициентов λ_1 и λ_2 равными 1. Далее значения изменяются в зависимости от точности полученных результатов. В данном случае задача идентификации параметров уравнения для переменной $D_{12}(n)$ решается сразу с достаточной точностью и есть проблемы с идентификацией параметров уравнения для переменной $D_{22}(n)$. Поэтому значение коэффициента λ_2 последовательно увеличивалось. Наилучшие результаты получены при $\lambda_1 = 1$ и $\lambda_2 = 10^4$.

При использовании весовых коэффициентов $\lambda_1 = 1$ и $\lambda_2 = 10^4$ применялся метод Luus-Jaakola [16] с параметрами: $R = 200$, $P = 200$, $ITER = 200$, $\gamma = 0,8$, $\eta = 0,9$, $\varepsilon_1 = \varepsilon_2 = 10^{-50}$ с ограничением на время выполнения: 60 с, метод интегрирования: метод Эйлера с шагом 100.

В качестве начального приближения для метода Luus-Jaakola использовался результат применения метода MMFO при использовании весовых коэффициентов $\lambda_1 = 1$ и $\lambda_2 = 100$:

$$\begin{aligned} c_{1,12} &= 3,06335104509683 \cdot 10^{-6}; \\ c_{2,12} &= 0,0360565134569079; \\ c_{3,12} &= 3,23801707906896 \cdot 10^{-5}; \end{aligned}$$

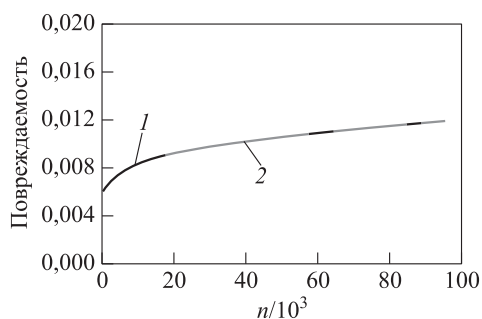


Рис. 2 Сравнение кривых поврежденности на направлении 1-1, полученных при решении прямой (черная, 1) и обратной (серая, 2) задач

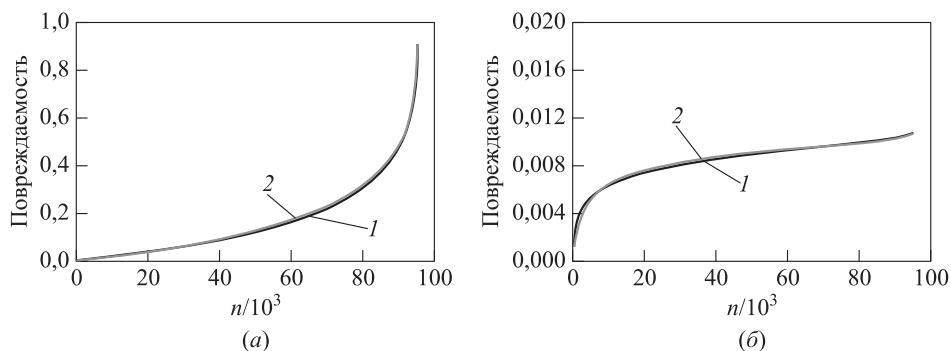


Рис. 3 Сравнение кривых повреждений на направлениях 1–2 (а) и 2–2 (б), полученных при решении прямой (черная, 1) и обратной (серая, 2) задач

$$\begin{aligned} c_{4,12} &= 0,99959425831806; \\ c_{5,12} &= 1,42837237268894; \\ c_{1,22} &= 2,200736913194343 \cdot 10^{-6}; \\ c_{2,22} &= 204,345098431823; \\ c_{3,22} &= -2,27037296384949 \cdot 10^{-5}; \\ c_{4,22} &= 0,798059542595711; \\ c_{5,22} &= 10,9375229807987. \end{aligned}$$

Полученные результаты последовательного применения двух методов оптимизации представлены на рис. 3. Найденные значения коэффициентов правых частей дифференциальных уравнений имеют вид:

$$\begin{aligned} c_{1,12} &= 3,15746410459782 \cdot 10^{-6}, \quad c_{2,12} = 0,0413824463669797, \\ c_{3,12} &= 2,72695039726558 \cdot 10^{-5}, \quad c_{4,12} = 0,9554012483105131, \\ & \quad c_{5,12} = 0,372390529265973; \\ c_{1,22} &= 8,51462220223207 \cdot 10^{-6}, \quad c_{2,22} = 365,709195625326, \\ c_{3,22} &= -1,01033043238019 \cdot 10^{-5}, \quad c_{4,22} = 0,760931609273087, \\ & \quad c_{5,22} = 7,36230735147148. \end{aligned}$$

Достигнутое значение критерия $I_2 = 0,257689079837312$.

Сравнительный анализ полученных решений прямой и обратной задач позволяет судить об их достаточной с практической точки зрения близости.

7 Анализ численных результатов

По результатам решения задачи (7) можно сделать вывод о том, что следует выбирать небольшой шаг интегрирования (например, $h = 100$). Метод Эйлера–

Коши может приводить к достаточно сильным колебаниям в решении. Результаты, полученные методом Адамса, визуально почти не отличаются от полученных методом Эйлера, только работают дольше, а результаты нахождения разгонных точек нарушают «гладкость» графика в начале процесса интегрирования.

В результате решения обратной задачи найдены коэффициенты, довольно близкие к заданным в прямой задаче. У большинства полученных коэффициентов тот же порядок, что и у заданных. К тому же графики поврежденности, вычисленной для каждого направления, почти совпадают с графиками поведения исходной модели. Стоит заметить, что найденный коэффициент $c_{3,22}$ — отрицательный, что свидетельствует о необходимости учета данного факта при задании множества допустимых решений. Заметим, что значения коэффициентов c_4 и c_5 играют роль только в случае возникновения этапа стремительного разрушения образца, а этот этап возникает не для всех направлений. Лучший результат решения задачи (8) получен при весовых коэффициентах $\lambda_1 = 1$ и $\lambda_2 = 10^4$, отражающих относительный приоритет слагаемых в критерии.

8 Заключение

Исследована возможность применения трехмерной математической модели для описания процесса деградации жесткости модельного композиционного материала в процессе испытаний. Параметры модели, входящие в правые части трех обыкновенных дифференциальных уравнений первого порядка, описывающих процесс деградации жесткости по трем направлениям, могут быть найдены в результате обработки данных с помощью процедуры параметрической идентификации. Для ее реализации предложено последовательно применить два метаэвристических алгоритма глобальной условной оптимизации. Для повышения эффективности численных методов оптимизации предложена модификация метода, имитирующего поведение стаи мотыльков. На примере модельного композиционного материала показано, что исследованная трехмерная модель позволяет получать результаты с приемлемой точностью.

Литература

1. *Shokrieh M. M., Lessard L. B.* Progressive fatigue damage modeling of composite materials. Part I: Modeling // *J. Compos. Mater.*, 2000. Vol. 34. No. 13. P. 1056–1080. doi: 10.1177/002199830003401301.
2. *Van Paeppegem W.* Development and finite element implementation of a damage model for fatigue of fibre-reinforced polymers. — Ghent, Belgium: Ghent University Architectural and Engineering Press, 2002. 395 p.
3. *Turbin N. V., Shelkov K. A.* Numerical study of fatigue damage accumulation in composite wing panels of prospective supersonic transport aircraft // *Aerospace Systems*, 2023. Vol. 6. P. 481–490. doi: 10.1007/s42401-023-00200-1.
4. *Shramko K. K., Kononov N. O., Lutoshkina A. E., Shadrinov A. V.* Computational estimate of the initial damage effect on the fatigue strength of composite materials // *J. Composites Science*, 2023. Vol. 7. Iss. 10. Art. 438. 15 p. doi: 10.3390/jcs7100438.

5. Brunbauer J., Arbeiter F., Stelzer S., Pinter G. Stiffness based fatigue characterisation of CFRP // *Adv. Mat. Res.*, 2014. Vol. 891-892. P. 166–171. doi: 10.4028/www.scientific.net/AMR.891-892.166.
6. Whitworth H. A. A stiffness degradation model for composite laminates under fatigue loading // *Compos. Struct.*, 1997. Vol. 40. No. 2. P. 95–101. doi: 10.1016/S0263-8223(97)00142-6.
7. Абгарян К. К., Журавлев А. А. Методы многомасштабного моделирования в задачах цифрового материаловедения. — М.: МАКС Пресс, 2022. 120 с. doi: 10.29003/m3138.978-5-317-06870-7. EDN: ASHFBK.
8. Абгарян К. К., Загордан Н. Л., Мочалова Ю. Д. Метод компьютерного моделирования упругих характеристик многослойных композиционных материалов // *Системы и средства информатики*, 2023. Т. 33. № 4. С. 92–101. doi: 10.14357/08696527230409. EDN: DPZJSX.
9. Пантелеев А. В., Турбин Н. В., Надоров И. С., Кононов Н. О. Параметрическая идентификация коэффициентов модели усталостной деградации жесткости композиционного материала // *Вестник Самарского университета. Аэрокосмическая техника, технологии и машиностроение*, 2024. Т. 23. № 3. С. 119–131. doi: 10.18287/2541-7533-2024-23-3-119-131. EDN: NQKXWU.
10. Encyclopedia of optimization / Eds. C. A. Floudas, P. M. Pardalos. — New York, NY, USA: Springer, 2009. 4646 p.
11. Handbook of metaheuristics / Eds. M. Gendreau, J.-Y. Potvin. — New York, NY, USA: Springer, 2019. 610 p. doi: 10.1007/978-3-319-91086-4.
12. Locatelli M., Schoen F. (Global) optimization: Historical notes and recent developments // *EURO J. Computational Optimization*, 2021. Vol. 9. Art. 100012. 15 p. doi: 10.1016/j.ejco.2021.100012.
13. Mirjalili S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm // *Knowl.-Based Syst.*, 2015. Vol. 89. P. 228–249. doi: 10.1016/i.knosys2015.07.006.
14. Del Ser J., Osaba E., Molina D., Yang X.-S., Salcedo-Sanz S., Camacho D., Das S., Suganthan P. N., Coello Coello C. A., Herrera F. Bio-inspired computation: Where we stand and what's next // *Swarm Evol. Comput.*, 2019. Vol. 48. P. 220–250. doi: 10.1016/j.swevo.2019.04.008.
15. Niculina Dragoi E., Dafinescu V. Review of metaheuristics inspired from the animal kingdom // *Mathematics*, 2021. Vol. 9. No. 18. Art. 2335. 52 p. doi: 10.3390/math9182335.
16. Luus R. Iterative dynamic programming. — London, U.K.: Chapman & Hall/CRC, 2000. 344 p.
17. «Росатом». Композитные технологии. <https://umatex.com/production/prepreg/?ysclid=mbi5471jgr741612511>.
18. Пантелеев А. В., Надоров И. С. Применение модификации метода, имитирующего поведение стаи мотыльков, для решения задачи оптимального программного управления мобильным роботом // *Моделирование и анализ данных*, 2025. Т. 15. № 1. С. 81–109. doi: 10.17759/mda.2025150105. EDN: PPGMJB.

Поступила в редакцию 18.08.2025

Принята к публикации 15.10.2025

METAHEURISTIC OPTIMIZATION ALGORITHMS OF PARAMETRIC IDENTIFICATION OF A THREE-DIMENSIONAL MODEL FOR REDUCING THE FATIGUE STIFFNESS OF COMPOSITE MATERIALS

A. V. Panteleev, N. V. Turbin, I. S. Nadorov, and N. O. Kononov

Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse,
Moscow 125933, Russian Federation

Abstract: The article explores the application of modern metaheuristic optimization algorithms to determine the parameters of new mathematical models describing experimental results. The subject of the study was the property of composite materials used in aircraft construction to lose fatigue stiffness during operation. For mathematical modeling of stiffness degradation, a system of three ordinary differential equations describing the characteristic stages of the process is proposed. To find the parameters of the right-hand sides of the equations, two constrained optimization problems are solved minimizing the deviations of the equation solutions from known composite material test results. The solutions to the equations are found using classical explicit numerical integration methods of varying orders of accuracy. For parametric identification, i. e., finding the values of the parameters of the right-hand sides of the differential equations based on test results, it is proposed to use metaheuristic optimization methods: a modified method simulating the behavior of a swarm of moths as well as a random search method with sequential reduction of the study area. Numerical results of the study of a specific composite material are presented illustrating the effectiveness of the proposed approach.

Keywords: metaheuristic optimization algorithms; moth swarm behavior simulating method; damage model; composite materials

DOI: 10.14357/08696527250406

EDN: PTPHPN

Acknowledgments

This study was carried out under the Agreement for the provision of grant funding from the federal budget for large scientific projects in priority areas of scientific and technological development of the Russian Ministry of Science and Higher Education No. 075-15-2024-552.

References

1. Shokrieh, M. M., and L. B. Lessard. 2000. Progressive fatigue damage modeling of composite materials. Part I: Modeling. *J. Compos. Mater.* 34(13):1056–1080. doi:10.1177/002199830003401301.

2. Van Paepegem, W. 2002. *Development and finite element implementation of a damage model for fatigue of fibre-reinforced polymers*. Ghent, Belgium: Ghent University Architectural and Engineering Press. 395 p.
3. Turbin, N. V., and K. A. Shelkov. 2023. Numerical study of fatigue damage accumulation in composite wing panels of prospective supersonic transport aircraft. *Aerospace Systems* 6:481–490. doi: 10.1007/s42401-023-00200-1.
4. Shramko, K. K., N. O. Kononov, A. E. Lutoshkina, and A. V. Shadrinov. 2023. Computational estimate of the initial damage effect on the fatigue strength of composite materials. *J. Composites Science* 7(10):438. 15 p. doi: 10.3390/jcs7100438.
5. Brunbauer, J., F. Arbeiter, S. Stelzer, and G. Pinter. 2014. Stiffness based fatigue characterisation of CFRP. *Adv. Mat. Res.* 891-892:166–171. doi: 10.4028/www.scientific.net/AMR.891-892.166.
6. Whitworth, H. A. 1997. A stiffness degradation model for composite laminates under fatigue loading. *Compos. Struct.* 40(2):95–101. doi: 10.1016/S0263-8223(97)00142-6.
7. Abgaryan, K. K., and A. A. Zhuravlev. 2022. *Metody mnogomasshtabnogo modelirovaniya v zadachakh tsifrovogo materialovedeniya* [Methods of multiscale modeling in the problems of digital materials science]. Moscow: MAX Press. 120 p. doi: 10.29003/m3138.978-5-317-06870-7. EDN: ASHFBK.
8. Abgaryan, K. K., N. L. Zagordan, and Y. D. Mochalova. 2023. Metod komp'yuternogo modelirovaniya uprugikh kharakteristik mnogoslownykh kompozitsionnykh materialov [Method for computer modeling of elastic characteristics of multilayer composite materials]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(4):92–101. doi: 10.14357/08696527230409. EDN: DPZJSX.
9. Panteleev, A. V., N. V. Turbin, I. S. Nadorov, and N. O. Kononov. 2024. Parametricheskaya identifikatsiya koeffitsientov modeli ustalostnoy degradatsii zhestkosti kompozitsionnogo materiala [Parametric identification of coefficients for a model of fatigue stiffness degradation of a composite material]. *Vestnik Samarskogo universiteta. Aerokosmicheskaya tekhnika, tekhnologii i mashinostroenie* [Vestnik of Samara University. Aerospace and Mechanical Engineering] 23(3):119–131. doi: 10.18287/2541-7533-2024-23-3-119-131. EDN: NQKXWU.
10. Floudas, C. A., and P. M. Pardalos, eds. 2009. *Encyclopedia of optimization*. New York, NY: Springer. 4646 p.
11. Gendreau, M., and J.-Y. Potvin, eds. 2019. *Handbook of metaheuristics*. New York, NY: Springer. 610 p. doi: 10.1007/978-3-319-91086-4.
12. Locatelli, M., and F. Schoen. 2021. (Global) optimization: Historical notes and recent developments. *EURO J. Computational Optimization* 9:100012. 15 p. doi: 10.1016/j.ejco.2021.100012.
13. Mirjalili, S. 2015. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowl.-Based Syst.* 89:228–249. doi: 10.1016/j.knsys.2015.07.006.
14. Del Ser, J., E. Osaba, D. Molina, X.-S. Yang, S. Salcedo-Sanz, D. Camacho, S. Das, P. N. Suganthan, C. A. Coello Coello, and F. Herrera. 2019. Bio-inspired computation: Where we stand and what's next. *Swarm Evol. Comput.* 48:220–250. doi: 10.1016/j.swevo.2019.04.008.
15. Niculina Dragoi, E., and V. Dafinescu. 2021. Review of metaheuristics inspired from the animal kingdom. *Mathematics* 9(18):2335. 52 p. doi: 10.3390/math9182335.

16. Luus, R. 2000. *Iterative dynamic programming*. London, U.K.: Chapman & Hall/CRC. 344 p.
17. Rosatom. Kompozitnye tekhnologii [Composite Technologies]. Available at: <https://umatex.com/production/prepreg/?ysclid=mbi5471jgr741612511> (accessed November 12, 2025).
18. Pantelev, A. V., and I. S. Nadorov. 2025. Primenenie modifikatsii metoda, imitiruyushchego povedenie stai motyl'kov, dlya resheniya zadachi optimal'nogo programmogo upravleniya mobil'nym robotom [Application of the modified method simulating the behavior of a flock of moths to solve the optimal open loop control problem of a mobile robot movement]. *Modelirovanie i analiz dannykh* [Modeling and Data Analysis] 15(1):81–109. doi: 10.17759/mda.2025150105. EDN: PPGMJJB.

Received August 18, 2025

Accepted October 15, 2025

Contributors

Pantelev Andrei V. (b. 1955) — Doctor of Science in physics and mathematics, professor, head of Department of Mathematical Cybernetics, Institute of Information Technology and Applied Mathematics, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; avpantelev@inbox.ru

Turbin Nikolay V. (b. 1989) — leading engineer, Advanced Engineering School, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; turbinnv@mai.ru

Nadorov Ivan S. (b. 2003) — assistant, Department of Mathematical Cybernetics, Institute of Information Technology and Applied Mathematics, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; nnadorovivan@gmail.com

Kononov Nikolay O. (b. 2000) — engineer, Aircraft Engineering School, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; kononovno@mai.ru

ПОСТРОЕНИЕ И АНАЛИЗ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ТЕХНИЧЕСКОГО СОСТОЯНИЯ БУКСОВЫХ УЗЛОВ ВАГОНОВ С ПРИМЕНЕНИЕМ ИНТЕЛЛЕКТУАЛЬНЫХ МЕТОДОВ ПРЕДИКТИВНОЙ АНАЛИТИКИ

*О. В. Дружинина¹, Э. Р. Корепанов², И. В. Макаренкова³, В. В. Максимова⁴,
А. А. Петров⁵*

Аннотация: Статья посвящена изучению проблемы построения и анализа модели прогнозирования технического состояния буксовых узлов железнодорожных вагонов на основе применения методов искусственного интеллекта. Актуальность указанной проблемы связана с необходимостью создания и совершенствования наукоемких и энергоэффективных инструментов анализа данных в диагностике технического состояния элементов и систем транспортной инфраструктуры. Для прогнозирования состояния при обработке последовательных данных (временных рядов) предлагается использовать нейросетевую архитектуру LSTM (Long Short-Term Memory). Синтетические датасеты для обучения нейросети сгенерированы с помощью разработанной имитационной стохастической модели термоконтроля буксовых узлов. Выполненное компьютерное моделирование в среде PyTorch позволило провести сравнительный анализ результатов вычислительных экспериментов и оценить эффективность обучения LSTM в рамках рассматриваемой задачи. Построенная модель предиктивной аналитики может служить основой модуля прогнозирования ABITech Thermal Forecast Module программного комплекса диагностики технического состояния буксовых узлов.

Ключевые слова: анализ данных; предиктивная аналитика; компьютерное моделирование; искусственные нейронные сети LSTM; временные ряды; алгоритмы машинного обучения; оценка технического состояния; интеллектуальные транспортные системы

DOI: 10.14357/08696527250407

EDN: RDTCFO

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, odruzhinina@frccsc.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ekorpanov@frccsc.ru

³Федеральный исследовательский центр «Информатика и управление» Российской академии наук, imakarenkova@frccsc.ru

⁴Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vmaksimova@frccsc.ru

⁵Елецкий государственный университет им. И. А. Бунина, xea191@yandex.ru

1 Введение

Современные технологии искусственного интеллекта, основанные на методах машинного обучения, позволяют эффективно обрабатывать большие объемы цифровых данных, собираемых из различных источников, для получения прогнозных рекомендаций [1]. Глубокие нейронные сети с архитектурой «трансформер» [2] в настоящее время успешно решают задачи обработки последовательных данных для выявления долгосрочных зависимостей, но очень требовательны к наличию существенных вычислительных ресурсов, значительных объемов памяти, качественных данных для обучения и развертывания. В то же время задачи анализа последовательных числовых данных (временных рядов) с получением прогнозных рекомендаций вполне успешно могут решаться методами машинного обучения с использованием класса рекуррентных нейронных сетей. При этом с помощью рекуррентных нейронных сетей типа LSTM [3, 4] можно сочетать достаточно простую вычислительную архитектуру и экономное использование памяти вычислителя с высокой эффективностью решения задач прогноза за счет использования специальных технологий отслеживания долгосрочных зависимостей во временных рядах. Применение LSTM-моделей для решения задач предиктивной аналитики, возникающих при изучении функционирования технических систем, рассмотрено в [5, 6] и в других работах.

Нейронные сети часто требуют достаточно большого объема вводной информации, которая получена в результате экспериментов или натурных измерений, однако на практике не всегда удается получить репрезентативные датасеты ввиду особенностей изучаемых процессов и систем, а также недостаточности требуемого оборудования и сложностей установки стационарных датчиков [7]. Для формирования таких данных в последнее время используется аналитический подход к созданию синтетических датасетов для последующей обработки в нейросетевых моделях. При данном подходе возможно использовать построение таких математических моделей с имитационными компонентами, реализация которых позволяет создавать наборы данных со свойствами, присущими реальным техническим системам. В частности, актуальной задачей стала разработка методов оценки технического состояния элементов и узлов транспортных систем, в том числе с применением методов нейросетевых алгоритмов и технологии цифровых двойников [8].

Системы температурного контроля буксовых узлов подвижного состава — одни из важнейших элементов обеспечения безопасности эксплуатации железнодорожного транспорта. Как известно [9], повышение температуры буксового узла служит значимым индикатором приближающегося отказа подшипников колесной пары задолго до того, как они выйдут из строя. В настоящее время основным методом обнаружения температурных аномалий буксовых узлов стало использование стационарных средств бесконтактного измерения температуры в составе автоматизированных многофункциональных комплексов технических средств диагностики, устанавливаемых на путях движения. При использовании

такого метода возникают сложности, связанные с необходимостью размещения на железной дороге большого числа дорогостоящих стационарных комплексов диагностики (например, КТСМ-02, КТСМ-03 [10, 11]). Стационарные комплексы размещаются на значительном расстоянии друг от друга и не имеют данных об изменениях температуры буксовых узлов на промежуточных участках, что может приводить к ошибочным решениям об отцеплении вагонов и необоснованным задержкам движения поездов. В связи с этим возникают задачи, направленные на реализацию возможности проведения расчетов в реальном времени, в том числе с использованием различных нейросетевых алгоритмов в рамках цифровых двойников [12].

В настоящей работе предлагается новый способ прогнозирования состояния буксовых узлов на основе построения модели предиктивной аналитики для анализа временных рядов в виде температурных данных, измеряемых непрерывно в ходе эксплуатации железнодорожных вагонов. Практическое использование разработанной модели может осуществляться с учетом внедрения современных систем бортового мониторинга состояния вагонов на основе современных микроконтроллеров и интеллектуальных датчиков. Предлагается решение задачи прогнозирования состояния буксовых узлов вагонов по температурным признакам с использованием методов машинного обучения для нейронных сетей архитектуры LSTM. Следует отметить, что предлагаемый подход к выбору архитектуры нейросети и к ее обучению разрабатывается с учетом особенностей обеспечения функциональной безопасности технических систем, в которых используются технологии искусственного интеллекта [13, 14].

В процессе решения задачи прогнозирования выполняется построение имитационной стохастической модели термоконтроля буксовых узлов колесной пары для генерации синтетических датасетов с последующим обучением предложенной нейросети. С учетом проведенных авторами вычислительных экспериментов дается оценка эффективности нейросети для прогноза температурных режимов с применением метрик производительности Precision, Recall и F1-score.

2 Особенности обработки натурных данных при термометрии буксовых узлов и постановка задачи исследования

Одним из перспективных направлений развития технологий мониторинга состояния подвижного состава железнодорожного транспорта считается разработка индивидуальных мобильных систем контроля, обеспечивающих непрерывный мониторинг маркеров возникновения аварийных ситуаций [15, 16]. Исследования в данном направлении связаны с применением современной микроэлектроники и технологии интернета вещей, которые позволяют создавать автономные интеллектуальные датчики со сверхнизким энергопотреблением, передающие данные по радиоканалу (см., например, [17] или [18]). Среди перспективных технических решений также можно отметить датчики, работающие за счет «сбора энергии»

из окружающей среды [19, 20] на протяжении всего срока эксплуатации по назначению. Реализация индивидуальных мобильных средств контроля значительно увеличивает объем получаемых данных о техническом состоянии состава, в связи с чем важной научной задачей становится разработка энергоэффективных методов обработки данных с использованием машинного обучения.

Повышение эффективности обработки температурных данных непосредственно на подвижном составе может быть достигнуто путем их реализации в виде искусственных нейронных сетей с использованием специализированных энергоэффективных чипов, в том числе с использованием 8/32-битных микроконтроллеров, энергоэффективных программируемых логических интегральных схем или специализированных заказных чипов под конкретную реализацию нейронной сети.

В предиктивной аналитике используются нейронные сети LSTM [2, 3]. Данная архитектура подтвердила свою эффективность в задачах прогноза новой информации (предсказание следующего элемента временного ряда) исходя из анализа и извлечения определенных закономерностей в «исторических данных» предыдущих временных периодов, в том числе и при обработке данных в режиме реального времени, получаемых от устройств интернета вещей в составе цифровых двойников [21, 22]. Специфика вычислительных операций, заложенная в LSTM, с учетом кусочно-линейной аппроксимации функций активации позволяет выполнить реализацию LSTM-модели на современных микропроцессорных архитектурах. На аппаратную платформу переносятся уже рассчитанные в процессе обучения весовые коэффициенты нейронной сети. При этом часто возникает проблема получения данных для обучения, решение которой при создании системы прогнозирования технического состояния буксовых узлов связано с необходимостью разработки методов формирования репрезентативных длинных временных рядов температурных признаков.

Целью настоящей статьи ставится построение и анализ модели прогнозирования технического состояния буксовых узлов вагонов с применением интеллектуальных методов предиктивной аналитики. Основными задачами ставятся: выбор и разработка архитектуры нейронной сети для машинного обучения данных температурных признаков; построение имитационной модели для генерации синтетических датасетов, необходимых для машинного обучения; разработка алгоритма обучения модели машинного обучения; оценка эффективности с использованием метрик Accuracy, F1-score и Recall. Схема исследования на основе построения интеллектуальной предиктивной модели дана на рис. 1.

Согласно рис. 1, поэтапное решение проблемы прогнозирования технического состояния буксовых узлов включает в себя изучение предметной области, постановку задачи предиктивной аналитики, а также экспериментальное получение данных термометрии с последующим построением предиктивной модели. Следует отметить, что указанная предиктивная модель может быть итеративно уточнена и верифицирована с использованием дополнительных данных. В работе получение данных осуществляется с помощью стохастического компьютерного

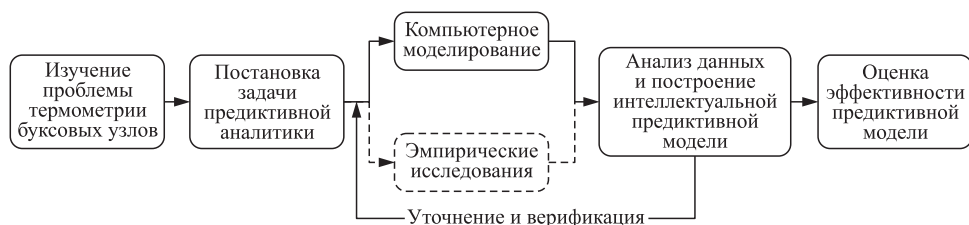


Рис. 1 Схема исследования на основе построения интеллектуальной предиктивной модели

моделирования, при этом одним из перспективных направлений развития исследования становится получение данных на основе натуральных испытаний.

3 Построение имитационной модели, разработка алгоритма машинного обучения и реализация в виде компьютерного прототипа

Для решения задачи оценки технического состояния буксовых узлов используется программное решение на основе алгоритма нейросети LSTM, реализованное с применением библиотеки PyTorch. Как известно [23], PyTorch считается одной из наиболее популярных библиотек глубокого обучения благодаря своей гибкости, простоте использования и высокой производительности. Среди ключевых особенностей указанной библиотеки можно отметить поддержку графовой структуры вычислений, высокую степень интеграции с программной инфраструктурой Python, поддержку гибридных вычислений на различных типах ускорителей (Nvidia, AMD, Intel), поддержку распределенных вычислений и вычислений на мобильных устройствах. Кроме того, следует отметить, что библиотека PyTorch может быть использована на отечественной программно-аппаратной платформе Эльбрус 801-PC [24].

Архитектура разработанной нейросети LSTM продемонстрирована на рис. 2.

Следует отметить, что архитектура гибридная, так как использует сочетание слоев LSTM и полносвязных слоев прямого распространения. Согласно рис. 2, сеть состоит из слоя входа размерности 2, двух скрытых слоев LSTM размерности 16 и выходного линейного слоя размерности 16. Выход нейросети обозначается предполагаемым единственным one-hot-вектором $N + 1$, возможные значения которого приведены в табл. 1.

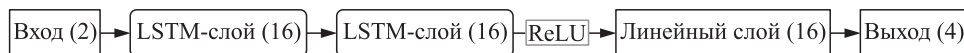


Рис. 2 Архитектура нейросети; ReLu — Rectified Linear Unit (выпрямительный линейный блок)

Таблица 1 Векторы технического состояния буксовых узлов

Вектор	Температурное состояние
[1, 0, 0, 0]	N — нормальное состояние
[0, 1, 0, 0]	E0 — тревога 1
[0, 0, 1, 0]	E1 — тревога 2
[0, 0, 0, 1]	E2 — тревога 3

Отметим, что классификация по четырем уровням состояния считается стандартной, принятой для анализа получаемых данных при температурном контроле буксовых узлов с помощью КТСМ. Однако получаемые данные на основе КТСМ в силу особенностей их применения не позволяют формировать репрезентативные длинные временные ряды температурных признаков.

Для возможности формирования таких временных рядов предлагается, во-первых, создать синтетические входные временные ряды на основе динамической вероятностной модели и, во-вторых, выполнить прогнозирование технического состояния на основе построения LSTM-модели с учетом сохранения принятой классификации состояний.

Идея построения алгоритма машинного обучения состоит в следующем. Общая последовательность обучающей выборки разделяется на батчи длиной $N + 1$, где N — длина одновременно обрабатываемых данных. Одновременно обрабатываемые данные указывают на возникновение временных корреляций, по которым можно отслеживать тренды на изменение температурных показателей. С учетом того что рассматриваются температурные показатели двух буксовых узлов колесной пары, на вход нейросети подается последовательность из N температурных параметров вида $(t_{11}, t_{21}), (t_{12}, t_{22}), \dots, (t_{1N}, t_{2N})$, где t_{in} и t_{jn} — результаты n -го температурного измерения для i -го и j -го буксовых узлов. Выбор параметрических «двоек» обусловлен числом измеряемых температурных показателей. Например, t_{12} — температура второго измерения для первого буксового узла.

В рамках решения задачи предиктивной аналитики осуществлена реализация блока программы, связанного с функционированием модели и предобработкой данных. Указанная предобработка соответствует построению соответствующих пар «последовательность измерений — термосостояние» (см. табл. 1). Пример предобработки датасета для нейросети LSTM представлен на рис. 3.

Параметры машинного обучения приведены во фрагменте программного кода на рис. 4. В качестве функции потерь используется категориальная кросс-энтропия (`nn.CrossEntropyLoss`), в качестве оптимизатора используется Adam со скоростью обучения (learning rate) 0,001. Приведенные на рис. 4 параметры позволяют достичь устойчивого снижения ошибки при обучении, что подтверждается вычислительными экспериментами.

```
tensor([[27., 35.],
        [28., 35.],
        [28., 34.],
        [28., 34.],
        ...,
        [25., 30.],
        [25., 30.],
        [26., 30.],
        [27., 30.]]) → tensor([0., 1., 0., 0.]
```

Рис. 3 Пример предобработки датасета с учетом стандартного типа данных tensor (значение [0, 1, 0, 0] соответствует тревоге «Е0»)

```
#model = LSTMPredictor(input_dim = 2, hidden_dim = 32, layer_dim = 2, output_dim = 4) ##создание модели
model = LSTMStatePredictor(input_dim = 2, hidden_dim = 16, layer_dim = 2, output_dim = 4)
criterion = nn.CrossEntropyLoss() ##определение функции потерь через категориальную кроссэнтропию
optimizer = torch.optim.Adam(model.parameters(), lr=0.001) ##промежуточный оптимизатор adam
```

Рис. 4 Фрагмент программы для реализации LSTM, содержащий определение модели

Для генерации синтетического датасета предлагается построение имитационной стохастической модели термоконтроля буксовых узлов колесной пары. В рамках указанной модели предполагается, что термоконтроль двух буксовых узлов колесной пары с учетом переходов из одного термосостояния в другое можно описать с помощью дискретного марковского процесса [25, 26]. Температурные показатели представлены в виде сетки состояний (рис. 5), где каждый узел

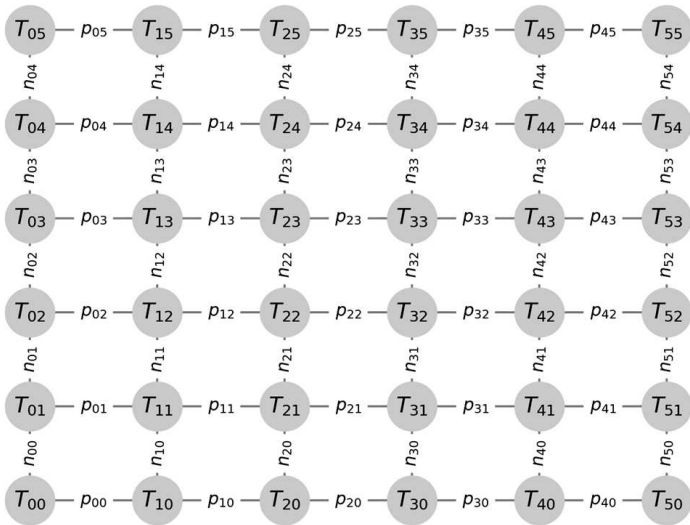


Рис. 5 Пример температурной сетки

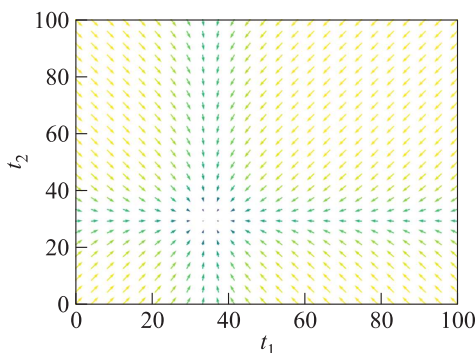


Рис. 6 Проекция векторного поля вероятностей перехода для совокупности $\langle a_1, a_2, a_3 \rangle$

представляет собой одну из возможных комбинаций температурных показателей t_i, t_j как T_{ij} для первого и второго буксового узла колесной пары соответственно, причем T_{ij} представляет собой узел сетки. Таким образом, множество всех узлов сетки задается как $T = (t_{11}, t_{12}, \dots, t_{1n}) \times (t_{21}, t_{22}, \dots, t_{2k})$, где n и k — размеры сетки; $\times$ — декартово произведение.

На рис. 5 приведен пример температурной сетки для 6 возможных температурных состояний для каждого буксового узла.

Вероятности переходов можно определить как векторное поле элементов S , т. е. $\forall T_{ij} \exists S_{ij} = (n_{ij}, p_{ij}, n_{ij-1}, p_{i-1,j}, s)$, где i и j соответствуют индексам состояний, через n и p обозначены вероятности переходов в соседние узлы, а через s — вероятность перехода в тот же самый узел. Поскольку процесс марковский, справедливо равенство $n_{ij} + p_{ij} + n_{ij-1} + p_{i-1,j} + s = 1$. Элементы рассматриваемого векторного поля S можно определить через совокупность $\langle a_1, a_2, a_3 \rangle$, для которой

$$a_1 = f(T_i) \times f(T_j); \quad a_2 = (\max(f), \max(f)) - a_1; \quad a_3 = 1 - \text{sum}(a_1 + a_2),$$

где f — скалярная функция распределения вероятностей роста температуры; $\max(f(T_i)) < 1$; $\min(f(T_i)) \geq 0$; sum — поэлементная сумма (аккумуляция) вектора. На рис. 6 приведена проекция векторного поля для значений $\text{sum}(a_1 + a_2)$, которая иллюстрирует наиболее вероятные направления роста температурных показателей.

Согласно рис. 6, цвет вектора описывает скалярное значение нормы, которое пропорционально вероятности изменения состояния системы (чем «светлее» вектор, тем больше вероятность перехода).

Предложенная имитационная модель позволяет выполнять генерацию температурных рядов для составления синтетического датасета для обучения искусственной нейронной сети. На рис. 7 и 8 представлены графики температурных

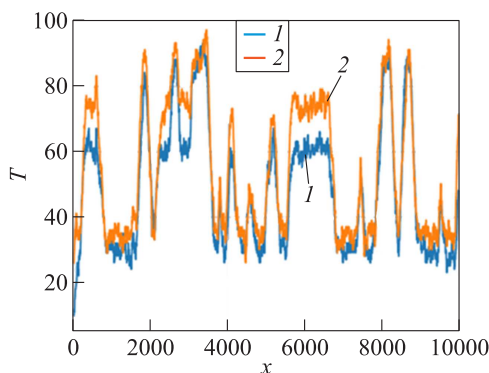


Рис. 7 Два сгенерированных температурных ряда для $T_{\text{БУ1}}$ (1) и $T_{\text{БУ2}}$ (2)

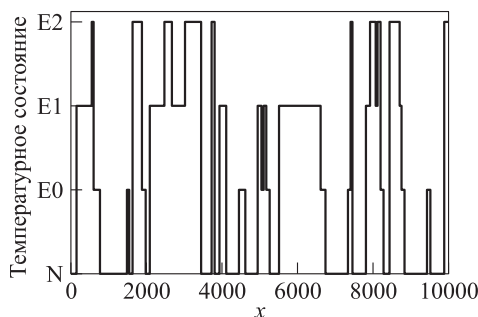


Рис. 8 Уровни тревоги для сгенерированных рядов

показателей и уровни тревоги соответственно для 10 000 сгенерированных измерений $T_{\text{БУ1}}$ и $T_{\text{БУ2}}$.

Для повышения репрезентативности и для реализации автоматической проверки сгенерированных датасетов предлагается учитывать наличие нескольких температурных сеток, соответствующих разным вероятным уровням тревоги. Переход между указанными температурными сетками также является дискретным марковским процессом [27]. На рис. 9 представлен пример тепловой карты вероятностей переходов от одной сетки к другой.

На рис. 9 по оси абсцисс обозначены векторы смещений узлов на графе, а по оси ординат порядковые номера узлов i, j . Для указанного примера предполагается наличие четырех возможных ожидаемых термических состояний, которые образуют полносвязную марковскую цепь. Цветовыми обозначениями представлены вероятности переходов из одного состояния в другое, согласно указанной цветовой шкале. Нулевые вероятности определяют «невозможные» тензоры смещения для несуществующих в графе узлов. Отметим, что если ни одно

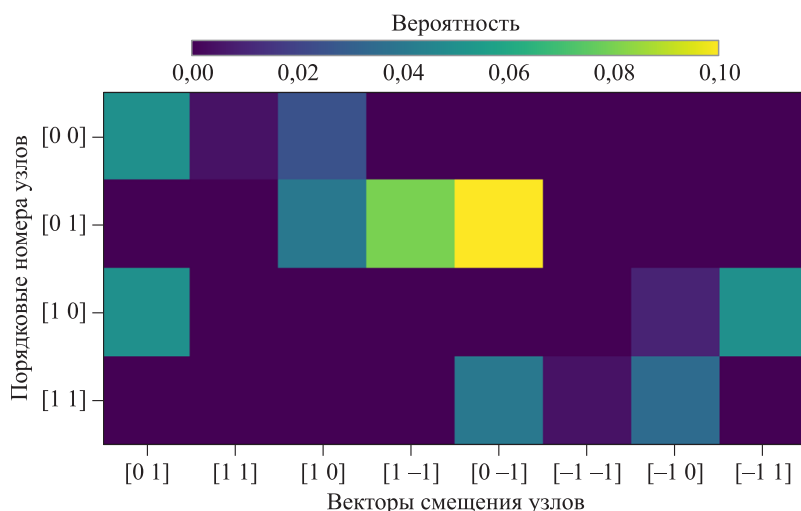


Рис. 9 Тепловая карта вероятностей переходов между сетками

из событий перехода не наступает, то состояние остается прежним, т. е. тензор смещения на текущей итерации алгоритма имеет вид $[0\ 0]$. Следует отметить, что рис. 9 представляет собой тепловую карту матричного представления марковской цепи, для которой вероятности переходов упорядочены по часовой стрелке, начиная от первого правого перехода (например, из $[0\ 0]$ в $[0\ 1]$ и т. д.).

Разработанное программное обеспечение на базе предложенной модели предиктивной аналитики и оригинальных алгоритмов входит в состав программного комплекса ABITechDiagnostics (Axle Box Intelligent Technical Diagnostics) и служит основой прогностического модуля ABITech Thermal Forecast Module. Ряд рабочих модулей комплекса описан в [28, 29].

4 Результаты компьютерного моделирования и оценка эффективности моделей

Для обучения используются синтетические последовательности из 1000 измерений параметров. Первая последовательность используется непосредственно для подбора весовых коэффициентов модели, а вторая применяется для валидации путем оценки метрик производительности модели. Таким образом, для разработанной модели выполняется кросс-валидация в пропорции 50/50.

Как известно, одними из наиболее часто используемых метрик производительности нейросетевых моделей служат Precision, Recall и F1-score [30]. Полученные метрики производительности модели приведены в табл. 2.

Отметим, что относительно невысокие метрики производительности модели на валидационном датасете обусловлены высокой сложностью задачи класси-

Таблица 2 Метрики производительности модели

Метрика	Значение на обучающем датасете	Значение на валидационном датасете
CrossEntropyLoss	0,7826	0,9965
Точность	0,9794	0,7626
F1-score	0,9659	0,7458
Recall	0,9527	0,7297

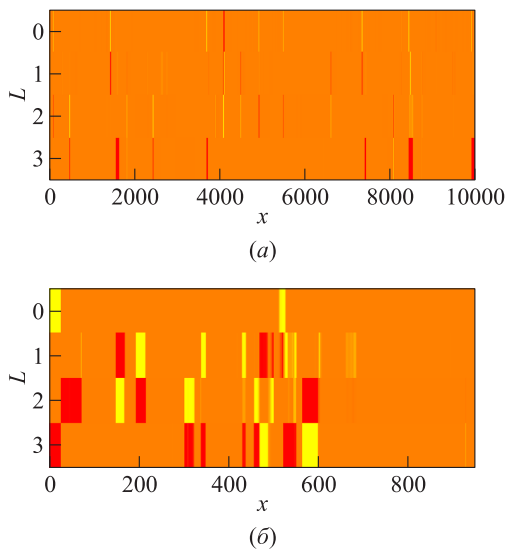


Рис. 10 Тепловые карты разностей предсказанных и истинных значений для обучающего (а) и валидационного (б) датасетов

фикации с использованием синтетического датасета, данные которого обладают большой энтропией. На рис. 10 приведены тепловые карты разностей предсказанных и истинных значений для обучающего и валидационного датасета соответственно. Однородный фон на указанных рисунках соответствует верным предсказаниям, симметричные красно-желтые полосы показывают систематические ошибки предсказаний, выборочные красные или желтые полосы отвечают неточной классификации.

Согласно рис. 10, а, на обучающем датасете ошибки рассеяны практически равномерно по всем четырем классам термосостояний. Высокие метрики производительности модели на обучающем датасете указывают на наличие незначительного переобучения.

Согласно рис. 10, б, модель дает некоторое число ошибочных предсказаний на валидационном датасете. Однако систематические ошибки локализованы

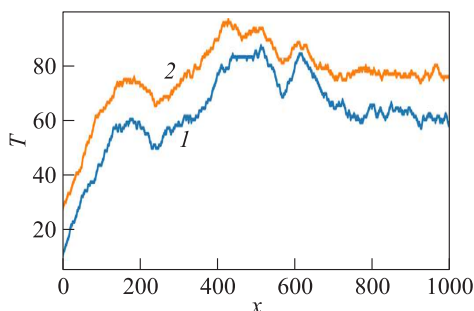


Рис. 11 Графики температур буксовых узлов $T_{БУ1}$ (1) и $T_{БУ2}$ (2) для тестового датасета из 1000 измерений

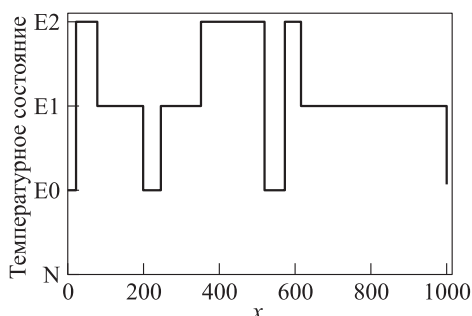


Рис. 12 Графики уровня температурной тревоги для тестового датасета

в основном по «соседним» уровням тревоги. В связи с указанным обстоятельством реализованная модель машинного обучения имеет среднюю производительность, что можно считать достаточным для целей обучения.

Далее рассмотрим графическое представление данных валидационного датасета. Графики температур буксовых узлов для тестового датасета представлены на рис. 11, графики уровня температурной тревоги представлены на рис. 12.

Согласно рис. 11 и 12, в тестовом датасете отсутствуют состояния системы «N», отвечающие нормальному температурному режиму. По этой причине невозможна правильная классификация на начальных этапах измерения, где температура буксового узла остается невысокой. В целом большая часть ошибок классификации между соседними классами отмечается в переходных режимах.

Следует отметить, что предложенная модель на базе LSTM с добавленным линейным полносвязным слоем использует результаты временного ряда температурных измерений для прогнозирования неисправностей, которые могут возникнуть в будущем. Представленное алгоритмическое и программное обеспечение разработано с учетом возможности реализации на основе встроенных микро-

процессорных измерительных комплексов с целью решения задач предиктивной аналитики.

5 Заключение

Рассмотрена постановка и изучены основные аспекты решения задачи прогнозирования технического состояния буксовых узлов железнодорожных вагонов с применением интеллектуальных методов предиктивной аналитики. Построенная модель предиктивной аналитики, включающая в себя имитационную модель термометрии буксовых узлов и модель машинного обучения, направлена на совершенствование методов сбора и обработки температурных данных буксовых узлов, на определение оптимальных параметров сигналов «тревога», на обеспечение баланса между риском возникновения аварийных ситуаций на перегоне и снижением задержек поездов из-за необходимости проведения проверок и отцепок вагонов.

Имитационная модель разработана на основе реализации дискретного марковского процесса. С использованием имитационной модели построены синтетические датасеты с учетом обучающей и валидационной компонент. Преимущество разработанной имитационной модели заключается в возможности генерации синтетических датасетов произвольного объема с сохранением репрезентативности данных, согласующихся с эмпирическими измерениями.

С использованием полученных данных реализована модель машинного обучения для многоклассовой классификации. В указанной модели используется нейросетевая архитектура LSTM для анализа временных рядов термометрии буксовых узлов. Классификация позволяет осуществлять предиктивную аналитику термосостояний, указывающих на уровень тревоги по четырехбалльной шкале.

Разработанное инструментальное обеспечение направлено на получение прогноза технического состояния железнодорожных составов с учетом возможностей получения термометрических данных на протяженных участках путей. В качестве перспектив исследования следует отметить предиктивный анализ с учетом увеличения количества температурных показателей и данных о вибрации, а также использование гибридных архитектур интеллектуальных классификаторов для улучшения метрик производительности нейросетевой модели. Результаты могут быть использованы при решении задач предиктивного анализа параметров транспортных систем, а также задач совершенствования средств технической диагностики.

Литература

1. Соколов И. А. Теория и практика применения методов искусственного интеллекта // Вестник Российской академии наук, 2019. Т. 89. № 4. С. 365–370. doi: 10.31857/S0869-5873894365-370. EDN: GQPYKZ.

2. *Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I.* Attention is all you need. — Cornell University, 2017. 15 p. arXiv:1706.03762v5 [cs.CL].
3. *Smagulova K., James A. P.* A survey on LSTM memristive neural network architectures and applications // *Eur. Phys. J. — Spec. Top.*, 2019. Vol. 228. No. 10. P. 2313–2324. doi: 10.1140/epjst/e2019-900046-x. EDN: HRKIKB.
4. *Ghojogh B., Ghodsi A.* Recurrent neural networks and long short-term memory networks: tutorial and survey. — Cornell University, 2023. 15 p. arXiv:2304.11461v1 [cs.LG].
5. *Федотов М. В., Грачев В. В.* Предиктивная аналитика технического состояния систем тепловозов с использованием нейросетевых прогнозных моделей // *Бюллетень результатов научных исследований*, 2021. № 3. С. 102–114. doi: 10.20295/2223-9987-2021-3-102-114. EDN: IKENIN.
6. *Чернухин А. В., Бозданова Е. А., Савицкая Т. В. и др.* Система предиктивной аналитики технического состояния эксгаустера агломашины с помощью методов искусственного интеллекта // *Искусственный интеллект и принятие решений*, 2024. № 3. С. 87–103. doi: 10.14357/20718594240307. EDN: HZKRKY.
7. *Дружинина О. В., Корепанов Э. Р., Петров А. А., Макаренкова И. В., Максимова В. В.* Построение модели генерации данных для решения задач классификации в диагностике неисправностей транспортных систем // *Нелинейный мир*, 2023. Т. 21. № 3. С. 16–26. doi: 10.18127/j20700970-202303-02. EDN: FOIGST.
8. *Белюсов В. В., Дружинина О. В., Корепанов Э. Р., Макаренкова И. В., Максимова В. В.* Подход к оценке технического состояния элементов и узлов транспортных систем с применением методов нейросетевого моделирования и технологии цифровых двойников // *Нейрокомпьютеры: разработка, применение*, 2021. Т. 23. № 5. С. 5–20. doi: 10.18127/j19998554-202105-01. EDN: EIAAOZ.
9. *Миронов А. А., Образцов В. Л., Павлюков А. Э.* Теория и практика бесконтактного теплового контроля буксовых узлов в поездах. — Екатеринбург: Ассорти, 2012. 396 с.
10. *Миронов А. А., Образцов В. Л., Митюшев В. С., Григорьев К. В.* Тепловой контроль буксовых узлов средствами КТСМ-02 // *Автоматика, связь, информатика*, 2011. № 12. С. 9–11. EDN: OMRLCN.
11. КТСМ-03 — Комплекс технических средств многофункциональный для диагностики ходовых частей железнодорожного подвижного состава. — Инфотэкс АТ, 2025. <https://infotecs-at.ru/products/7>.
12. *Белюсов В. В., Дружинина О. В., Корепанов Э. Р., Макаренкова И. В., Максимова В. В.* Применение нейронных сетей для решения задач классификации при выявлении неисправностей транспортных систем // *Нейрокомпьютеры: разработка, применение*, 2022. Т. 24. № 4. С. 18–27. doi: 10.18127/j19998554-202204-02. EDN: ZZAOWY.
13. *Гаврилов В. Е., Зацаринный А. А.* Проблемы и угрозы внедрения некоторых новых цифровых технологий // *Системы и средства информатики*, 2022. Т. 32. № 3. С. 15–25. doi: 10.14357/08696527220302. EDN: QKESJR.
14. *Гаврилов В. Е., Зацаринный А. А.* Особенности обеспечения функциональной безопасности автоматизированных систем с применением технологии искусственного интеллекта // *Системы и средства информатики*, 2024. Т. 34. № 3. С. 23–34. doi: 10.14357/08696527240303. EDN: RPQAIE.

15. WaggonTracker. Intelligent rail freight transport featured by WaggonTracker. — PJ Monitoring Gmbh (PJM), 2025. <https://pjm.co.at/en/waggontracker/>.
16. Даниленко Д. В., Стрельченко А. Н., Григорьев А. В. Система мониторинга состояния грузового железнодорожного вагона: Патент РФ № 2798214 С1 с приоритетом от 03.11.2022. Оpubл. 19.06.2023. EDN: DMKCVZ.
17. 802.15.4-2015 — IEEE Standard for Low-Rate Wireless Networks. — IEEEExplore, 2015. <https://ieeexplore.ieee.org/document/7460875>.
18. LoRaWAN Specification v1.1. — LoRa Alliance, 2017. 101 p. <https://resources.lora-alliance.org/technical-specifications/lorawan-specification-v1-1>.
19. Monagle D., Ponce E. A., Leeb S. B. Rule the Joule: An energy management design guide for self-powered sensors // IEEE Sens. J., 2024. Vol. 24. Iss.1. P. 6–15. doi: 10.1109/JSEN.2023.3336529.
20. Zewe A. Self-powered sensor automatically harvests magnetic energy. — MIT News, 2024. <https://news.mit.edu/2024/self-powered-sensor-harvests-magnetic-energy-0118>.
21. Прохоров А., Лysачев М. Цифровой двойник: анализ, тренды, мировой опыт / Под ред. А. Боровкова. — М.: Альянс Принт, 2020. 401 с.
22. Ковалев С. П. Применение нейронных сетей глубокого обучения в математическом обеспечении цифровых двойников электроэнергетических систем // Системы и средства информатики, 2021. Т. 31. № 1. С. 133–144. doi: 10.14357/08696527210111. EDN: NQPXR.
23. Programming with TensorFlow: Solution for edge computing applications / Eds. B. Kolla, G. R. Kanagachidambaresan. — Cham, Switzerland: Springer, 2021. 284 p. doi: 10.1007/978-3-030-57077-4.
24. Адаптированные нейросетевые алгоритмы на отечественных процессорах. — BrainTools.RU, 2025. <https://www.braintools.ru/article/13697>.
25. Ефимов И. Н., Морозов Е. А., Селиванов К. М. Компьютерное моделирование динамических систем. — Ижевск: Ин-т компьютерных исследований, 2014. 134 с.
26. Кузнецов С. В. Математические модели процессов и систем технической эксплуатации авионики как марковские цепи // Научный вестник МГТУ ГА, 2014. № 201. С. 56–64. EDN: RYFUTT.
27. Галажинская О. Н., Моисеева С. П. Теория случайных процессов. Ч. 2: Марковские процессы. — Томск: ТГУ, 2016. 126 с.
28. Дружинина О. В., Корепанов Э. Р., Петров А. А., Макаренкова И. В., Максимова В. В. Разработка инструментально-методического обеспечения для оценки технического состояния узлов транспортных систем с применением нейронных сетей, алгоритмов нечеткого и стохастического моделирования // Нейрокомпьютеры: разработка, применение, 2024. Т. 26. № 2. С. 13–22. doi: 10.18127/j19998554-202402-02. EDN: EPQPVE.
29. Дружинина О. В., Петров А. А., Макаренкова И. В., Максимова В. В. Программа синтеза данных и предобучения нейросети для цели мониторинга и диагностики буксовых узлов железнодорожных вагонов: Свидетельство о государственной регистрации программы для ЭВМ № 2024663875. Оpubл. 13.06.2024. EDN: DAQSRM.
30. Grandini M., Bagli E., Visani G. Metrics for multi-class classification: An overview. — Cornell University, 2020. 17 p. arXiv:2008.05756 [stat.ML].

Поступила в редакцию 20.06.2025

Принята к публикации 15.10.2025

CONSTRUCTION AND ANALYSIS OF MODEL TO PREDICT THE TECHNICAL STATE OF RAILWAY CAR AXLE BOXES USING INTELLIGENT PREDICTIVE ANALYTICS METHODS

*O. V. Druzhinina¹, E. R. Korepanov¹, I. V. Makarenkova¹, V. V. Maksimova¹,
and A. A. Petrov²*

¹Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Yelets State University named after I. A. Bunin, 28 Kommunarov Str., Yelets 399770, Lipetsk Region, Russian Federation

Abstract: The paper is devoted to the study of the problem of constructing and analyzing a model for predicting the technical state of axle boxes of railway cars based on the use of artificial intelligence methods. The relevance of this problem is related to the need to create and improve high-tech and energy-efficient data analysis tools for diagnosing the technical condition of elements and systems of transport infrastructure. It is proposed to use the LSTM (Long Short-Term Memory) neural network architecture to predict the state when processing sequential data (time series). Synthetic datasets for neural network training are generated using the developed simulation stochastic model of thermal control of axle boxes. The performed computer modeling in the PyTorch environment allowed to conduct a comparative analysis of the results of computational experiments and to evaluate the effectiveness of LSTM training in the framework of the problem under consideration. The constructed predictive analytics model can serve as the basis for the ABITech Thermal Forecast Module, a software package for diagnosing the technical state of axle boxes.

Keywords: data analysis; predictive analytics; computer modeling; neural networks LSTM; time series; machine learning algorithms; technical state assessment; intelligent transport systems

DOI: 10.14357/08696527250407

EDN: RDTFCFO

References

1. Sokolov, I. A. 2019. Theory and practice of application of artificial intelligence methods. *Her. Russ. Acad. Sci.* 89(2):115–119. doi: 10.1134/S1019331619020205. EDN: VYXJUC.
2. Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. 2017. Attention is all you need. Cornell University. 15 p. Available at: <https://arxiv.org/pdf/1706.03762v5> (accessed November 18, 2025).
3. Smagulova, K., and A. P. James. 2019. A survey on LSTM memristive neural network architectures and applications. *Eur. Phys. J. — Spec. Top.* 228(10):2313–2324. doi: 10.1140/epjst/e2019-900046-x. EDN: HRKIKB.

4. Ghojogh, B., and A. Ghodsi. 2023. Recurrent neural networks and long short-term memory networks: Tutorial and survey. Cornell University. 15 p. Available at: <https://arxiv.org/pdf/2304.11461v1> (accessed November 18, 2025).
5. Fedotov, M. V., and V. V. Grachev. 2021. Prediktivnaya analitika tekhnicheskogo sostoyaniya sistem teplovozov s ispol'zovaniem neyrosetevykh prognoznykh modeley [Predictive analytics of the technical condition of diesel locomotive systems using neural network predictive models]. *Byulleten' rezul'tatov nauchnykh issledovaniy* [Bulletin of Scientific Research Results] 3:102–114. doi: 10.20295/2223-9987-2021-3-102-114. EDN: IKEHIN.
6. Chernukhin, A. V., E. A. Bogdanova, T. V. Savitskaya, D. G. Kulakov, and I. R. Pavlov. 2024. Sistema prediktivnoy analitiki tekhnicheskogo sostoyaniya eks-gaustera aglomashiny s pomoshch'yu metodov iskusstvennogo intellekta [Predictive analytics system for the technical condition of a sinter extractor using artificial intelligence methods]. *Iskusstvennyy intellekt i prinyatie resheniy* [Artificial Intelligence and Decision Making] 3:87–103. doi: 10.14357/20718594240307. EDN: HZKRKY.
7. Druzhinina, O. V., E. R. Korepanov, A. A. Petrov, I. V. Makarenkova, and V. V. Maksimova. 2023. Postroenie modeli generatsii dannykh dlya resheniya zadach klassifikatsii v diagnostike neispravnostey transportnykh sistem [Design of a data generation model for solving classification problems in the diagnosis of transport system malfunctions]. *Nelineyny mir* [Nonlinear World] 21(3):16–26. doi: 10.18127/j20700970-202303-02. EDN: FOIGST.
8. Belousov, V. V., O. V. Druzhinina, E. R. Korepanov, I. V. Makarenkova, and V. V. Maksimova. 2021. Podkhod k otsenke tekhnicheskogo sostoyaniya elementov i uzlov transportnykh sistem s primeneniem metodov neyrosetevogo modelirovaniya i tekhnologii tsifrovyykh dvoynikov [An approach to assessing the technical condition of elements and nodes of transport systems using neural network modeling methods and digital twin technology]. *Neyrokomp'yutery: razrabotka, primeneniye* [Neurocomputers] 23(5):5–20. doi: 10.18127/j19998554-202105-01. EDN: EIAAOZ.
9. Mironov, A. A., V. L. Obraztsov, and A. E. Pavlyukov. 2012. *Teoriya i praktika beskontaktnogo teplovogo kontrolya buksovykh uzlov v poezdakh* [Theory and practice of noncontact thermal monitoring of axle boxes in trains]. Ekaterinburg: Assorti. 396 p.
10. Mironov, A. A., V. L. Obraztsov, V. S. Mityushev, and K. V. Grigor'ev. 2011. Teplovoy kontrol' buksovykh uzlov sredstvami KTSM-02 [Thermal control of axle boxes by means of KTSM-02]. *Avtomatika, svyaz', informatika* [Automation, Communications, Informatics] 12:9–11. EDN: OMRLCN.
11. KTSM-03 — kompleks tekhnicheskikh sredstv mnogofunktsional'nyy dlya diagnostiki khodovykh chastei zheleznodorozhnogo podvizhnogo sostava [KTSM-03: Multifunctional technical complex for diagnostics of running gear of railway rolling stock]. Available at: <https://infotecs-at.ru/products/7> (accessed November 18, 2025).
12. Belousov, V. V., O. V. Druzhinina, E. R. Korepanov, I. V. Makarenkova, and V. V. Maksimova. 2022. Primeneniye neyronnykh setey dlya resheniya zadach klassifikatsii pri vyyavlenii neispravnostey transportnykh sistem [The use of neural networks for solving classification problems in diagnosing malfunctions of transport systems]. *Neyrokomp'yutery: razrabotka, primeneniye* [Neurocomputers] 24(4):18–27. doi: 10.18127/j19998554-202204-02. EDN: ZZAOWY.
13. Gavrilov, V. E., and A. A. Zatsarinny. 2022. Problemy i ugrozy vnedreniya nekotorykh novykh tsifrovyykh tekhnologiy [Problems and threats of some new digital technologies]

- implementation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(3):15–25. doi: 10.14357/08696527220302. EDN: QKESJR.
14. Gavrilov, V. E., and A. A. Zatsarinny. 2024. Osobennosti obespecheniya funktsional'noy bezopasnosti avtomatizirovannykh sistem s primeneniem tekhnologii iskusstvennogo intellekta [Specific features of ensuring the functional safety of automated systems using artificial intelligence technology]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 34(3):23–34. doi: 10.14357/08696527240303. EDN: RPQAIE.
 15. WaggonTracker. 2025. Intelligent rail freight transport featured by WaggonTracker. PJ Monitoring GmbH. Available at: <https://pjm.co.at/en/waggontracker/> (accessed November 18, 2025).
 16. Danilenko, D. V., A. N. Strelchenko, and A. V. Grigoriev. 2023. Sistema monitoringa sostoyaniya gruzovogo zheleznodorozhnogo vagona [System for monitoring the condition of a freight railway car]. Patent RF No. 2798214. EDN: DMKCVZ.
 17. IEEE. 2015. Standard for low-rate wireless networks 802.15.4-2015. Available at: <https://ieeexplore.ieee.org/document/7460875> (accessed November 18, 2025).
 18. LoRaWAN specification v1.1. 2017. LoRa Alliance. 101 p. Available at: <https://resources.lora-alliance.org/technical-specifications/lorawan-specification-v1-1> (accessed November 18, 2025).
 19. Monagle, D., E. A. Ponce and S. B. Leeb. 2024. Rule the Joule: An energy management design guide for self-powered sensors. *IEEE Sens. J.* 24(1):6–15. doi: 10.1109/JSEN.2023.3336529.
 20. Zewe, A. 2024. Self-powered sensor automatically harvests magnetic energy. *MIT News*. Available at: <https://news.mit.edu/2024/self-powered-sensor-harvests-magnetic-energy-0118> (accessed November 18, 2025).
 21. Prokhorov, A., and M. Lysachev. 2020. *Tsifrovoy dvoynik: analiz, trendy, mirovoy opyt* [Digital twin: Analysis, trends, and global experience]. Ed. A. Borovkov. Moscow: Al'yans Print. 401 p.
 22. Kovalyov, S. P. 2021. Primenenie neyronnykh setey glubokogo obucheniya v matematicheskom obespechenii tsifrovyykh dvoynikov elektroenergeticheskikh sistem [Employing deep learning neural networks in mathematical basis of digital twins of electrical power systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(1):133–144. doi: 10.14357/08696527210111. EDN: NQPUXR.
 23. Kolla, B., and G. R. Kanagachidambaresan, eds. 2021. *Programming with TensorFlow. Solution for edge computing applications*. Cham, Switzerland: Springer. 284 p. doi: 10.1007/978-3-030-57077-4.
 24. Adaptirovannyye neyrosetevyye algoritmy na otechestvennykh protsessorakh [Adapted neural network algorithms on domestic processors]. 2025. Brain-Tools.RU. Available at: <https://www.braintools.ru/article/13697> (accessed November 18, 2025).
 25. Efimov, I. N., E. A. Morozov, and K. M. Selivanov. 2014. *Komp'yuternoe modelirovaniye dinamicheskikh sistem* [Computer modeling of dynamic systems]. Izhevsk: Institute of Computer Science. 134 p.
 26. Kuznetsov, S. 2014. Matematicheskie modeli protsessorov i sistem tekhnicheskoy ekspluatatsii avioniki kak markovskie tsepi [Processes and systems of avionics technical operation as Markov chains mathematical models]. *Nauchnyy vestnik MGTU GA* [Civil Aviation High Technologies] 201:56–64. EDN: RYFUTT.

27. Galazhinskaya, O.N., and S.P. Moiseeva. 2016. *Teoriya sluchaynykh protsessov. Ch. 2: Markovskie protsessy* [Theory of random processes. Part 2: Markov processes]. Tomsk: TGU. 126 p.
28. Druzhinina, O.V., E.R. Korepanov, A.A. Petrov, I.V. Makarenkova, and V.V. Maksimova. 2024. Razrabotka instrumental'no-metodicheskogo obespecheniya dlya otsenki tekhnicheskogo sostoyaniya uzlov transportnykh sistem s primeneniem neyronnykh setey, algoritmov nechetkogo i stokhasticheskogo modelirovaniya [Development of instrumental and methodological support for assessing the technical condition of transport system nodes using neural networks, fuzzy and stochastic modeling algorithms]. *Neyrokomp'yutery: razrabotka, primeneniye* [Neurocomputers] 26(2):13–22. doi: 10.18127/j19998554-202402-02. EDN: EPQPVE.
29. Druzhinina, O.V., A.A. Petrov, I.V. Makarenkova, and V.V. Maksimova. 2024. Programma sinteza dannykh i pereobucheniya neyroseti dlya tseli monitoringa i diagnostiki buksovykh uzlov zheleznodorozhnykh vagonov [The program of data synthesis and retraining of a neural network for the purpose of monitoring and diagnostics of axle boxes of railway wagons]. Certificate RF of State Registration of Computer Programs No. 2024663875. EDN: DAQSRM.
30. Grandini, M., E. Bagli, and G. Visani. 2020. Metrics for multi-class classification: An overview. Cornell University. 17 p. Available at: <https://arxiv.org/pdf/2008.05756> (accessed November 18, 2025).

Received June 20, 2025

Accepted October 15, 2025

Contributors

Druzhinina Olga V. (b. 1967) — Doctor of Science in physics and mathematics, professor, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; odruzhinina@frccsc.ru

Korepanov Eduard V. (b. 1966) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ekorepanov@frccsc.ru

Makarenkova Irina V. (b. 1964) — leading programmer, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; imakarenkova@frccsc.ru

Maksimova Vera V. (b. 1975) — leading programmer, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vmaksimova@frccsc.ru

Petrov Alexey A. (b. 1991) — Candidate of Science (PhD) in technology, associate professor, Department of Mathematical Modeling, Computer Technologies and Information Security, Yelets State University named after I. A. Bunin, 28 Kommunarov Str., Yelets 399770, Lipetsk Region, Russian Federation; xeal91@yandex.ru

КАРТОГРАФИРОВАНИЕ ПАХОТНЫХ ЗЕМЕЛЬ МУНИЦИПАЛЬНЫХ ОБРАЗОВАНИЙ ХАБАРОВСКОГО КРАЯ ПО ДАННЫМ СПУТНИКА «МЕТЕОР-М» № 2*

*А. С. Степанов¹, Л. В. Илларионова², К. Н. Дубровин³, Е. А. Фомина⁴,
А. Л. Верхотуров⁵*

Аннотация: Рассмотрена возможность применения еженедельных композитных изображений со спутника «Метеор-М» № 2 для классификации пахотных земель Хабаровского края. Для четырех классов растительности (соя, зерновые культуры, многолетние травы и залежь) в 2024 г. были построены средние ряды сезонного хода нормализованного вегетационного индекса (Normalized Difference Vegetation Index, NDVI) для муниципальных районов юга Хабаровского края и рассчитаны основные характеристики — значения максимума NDVI и дня наступления максимума. Выявлена статистическая значимость различий показателей для средних временных рядов NDVI разных классов растительности ($p < 0,0001$). На основе валидированных данных по Хабаровскому краю проведена классификация пахотных земель Бикинского, Вяземского районов и района им. Лазо на основе машинного обучения (алгоритм Random Forest (RF)). Средняя точность метода по результатам трехкратной кросс-валидации оказалась равной 87,6%. Для разных классов растительности значение F1-метрики находилось в диапазоне 0,61–0,93. Были построены карты пахотных земель для южных районов Хабаровского края. Установлено, что доля залежи от общей площади пахотных земель региона превысила 30%, а посевы сои в 2024 г. составили 48%. Результаты картографирования занесены в разработанную геоинформационную систему.

Ключевые слова: пахотные земли; машинное обучение; классификация; спутниковый мониторинг; ГИС

DOI: 10.14357/08696527250408

EDN: SGWRZO

*Исследование выполнено при поддержке РНФ (проект № 24-11-20030) и Правительства Хабаровского края (Соглашение № 106С/2024 от 31.07.2024).

¹Дальневосточный научно-исследовательский институт сельского хозяйства, stepanfx@mail.ru

²Вычислительный центр Дальневосточного отделения Российской академии наук, illarionova_l@list.ru

³Вычислительный центр Дальневосточного отделения Российской академии наук, k.dubrovin@ccfebras.ru

⁴Вычислительный центр Дальневосточного отделения Российской академии наук, eliz37@mail.ru

⁵Вычислительный центр Дальневосточного отделения Российской академии наук, andrey@ccfebras.ru

1 Введение

Дистанционное зондирование Земли (ДЗЗ) из космоса используется для решения различных сельскохозяйственных задач, таких как распознавание сельскохозяйственных культур, классификация растительного покрова и картографирование полей, благодаря его глобальному охвату и высокой доступности данных [1, 2]. Для решения задач спутникового мониторинга природных систем требуется анализ временных рядов данных (снимков, значений каналов и вегетационных индексов) [3], в том числе с использованием методов машинного обучения [4]. Применение геоинформационных систем (ГИС) и специализированных библиотек значительно упрощает использование геопространственных данных [5].

Спутниковые системы Landsat-8/9 и Sentinel-2 стали наиболее широко используемыми для мониторинга поверхности суши, в том числе для уточнения границ земель сельскохозяйственного назначения и оценке севооборота [2, 6]. Вместе с тем одним из приоритетов в развитии методов ДЗЗ, способствующих укреплению национального суверенитета в этой области, стало применение данных национальной спутниковой группировки для решения задач мониторинга [7]. Например, китайскими учеными активно исследуются возможности спутника Gaofen для идентификации сельскохозяйственных культур и оценки почвенных характеристик в сравнении с Sentinel-2 и Landsat [8, 9]. Точность классификации с использованием Sentinel-2 составила 97% при оценке F1 в 71%, в то время как точность классификации с использованием изображений Gaofen варьировалась от 75% до 97% при оценке F1 в 55%–73% [8]. Модели прогнозирования почвенных характеристик для спутников Landsat-8, Sentinel-2 и Gaofen-2 имели среднеквадратичные ошибки 0,34, 0,39 и 0,31 соответственно [9]. Возможности аргентинской группировки радиолокационных спутников SAOCOM-1A/1B достаточно широко используются для мониторинга пахотных земель Аргентины [10]. Точность классификации составляет от 62% до 90% в зависимости от фенологической фазы и вида сельскохозяйственной культуры.

Российская орбитальная группировка ДЗЗ также поддерживает задачи, связанные с мониторингом поверхности Земли. Спутники «Метеор-М» относятся к системам среднего пространственного разрешения. Спутники этой серии, находящиеся на полярной орбите, предназначены для предоставления оперативных данных Федеральной службе России по гидрометеорологии и мониторингу окружающей среды, научно-исследовательским институтам и другим организациям [11, 12]. Эти спутники используются для наблюдения за объектами и явлениями в атмосфере, гидросфере и криосфере в целях глобального мониторинга климата, реагирования на чрезвычайные ситуации и мониторинга окружающей среды [13]. В настоящее время на орбите функционируют четыре спутника «Метеор-М» — № 2, № 2.2, № 2.3 и № 2.4. Данные комплекса многозональной спутниковой съемки (КМСС) широко используются для дистанционной оценки

характеристик растительного покрова и оперативного мониторинга в России и прилегающих регионах [13, 14].

Ранее была проведена оценка возможности применения данных спутника «Метеор-М» № 2 в решении задач по идентификации сельскохозяйственных культур и классификации пахотных земель Хабаровского края [15]. По верифицированным данным о севообороте в период с 2021 по 2023 гг. сельскохозяйственных полей Хабаровского района была проведена попиксельная классификация на основе машинного обучения с использованием алгоритма Random Forest. По результатам кросс-валидации общая точность составила 0,95. В целом полученные результаты позволяют рассматривать изображения КМСС «Метеор-М» № 2 для решения задач классификации и картографирования пахотных земель, в том числе на территории Хабаровского края. На текущий момент информация о границах отдельных полей и произрастающих культурах южных районов Хабаровского края (Бикинский, Вяземский, им. Лазо), которая занесена в Единую федеральную информационную систему земель сельскохозяйственного назначения (ЕФИС ЗСН), требует уточнения и внесения корректировок. Основной целью исследования ставится построение карт пахотных земель трех южных районов Хабаровского края на основе спутниковых изображений КМСС «Метеор-М» № 2 за 2024 г. с использованием валидированных данных по Хабаровскому району. Важной сопутствующей целью выступает разработка и наполнение информационной системы, служащей технологической платформой для реализации данного проекта. Были поставлены следующие задачи.

1. Провести классификацию пахотных земель на основе машинного обучения с использованием алгоритма RF, оценить точность и сформировать карты пахотных земель.
2. Построить средние временные ряды сезонного хода значений NDVI для разных типов растительности на пахотных землях муниципальных районов Хабаровского края.
3. Интегрировать полученные результаты в информационную систему для их хранения, визуализации и дальнейшего анализа.

2 Материалы и методы

Исходные данные и объект исследования

Были рассмотрены 3155 сельскохозяйственных полей (2024 г.), которые относятся к одному из 4-х классов: соя, залежь, многолетние травы и зерновые (пшеница, овес, ячмень). Поля располагаются в четырех муниципальных районах южной части Хабаровского края: Хабаровском (1148 полей, 35 391 га), им. Лазо (686 полей, 21 411 га), Вяземском (555 полей, 22 796 га) и Бикинском (418 полей, 10 174 га). В этих муниципальных образованиях находятся 99% пахотных земель Хабаровского края. Общая площадь исследуемых полей составляет 89 772 га.

Спутниковые данные и предобработка

В качестве источника данных применялись ежедневные композитные изображения каналов ближнего инфракрасного (NIR) и красного (RED) диапазонов спектра среднего разрешения (60 м) прибора КМСС спутника «Метеор-М» № 2. Использовались данные за период вегетации 2024 г. — с 118-го (конец апреля) по 299-й (конец октября) календарные дни. Композиты были получены с использованием инфраструктуры Центра коллективного пользования системами архивации, обработки и анализа данных спутниковых наблюдений ИКИ РАН [16].

Расчет вегетационных индексов и формирование временных рядов

Для каждого пикселя в границах исследуемых полей рассчитывался индекс NDVI по формуле

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}},$$

где NIR — отражение в ближней инфракрасной области спектра; RED — отражение в красной области спектра.

Обработка снимков выполнялась с использованием разработанного программного обеспечения на базе библиотек Python Rasterio и Osgeo [17]. В качестве входных данных использовались наборы композитных изображений (формат .tiff) за выбранный временной отрезок, а также векторные контуры сельскохозяйственных полей. Действие программы заключается в поиске на растровых изображениях пикселей, относящихся к сельскохозяйственным полям, с последующей генерацией файлов в формате .csv для каждого поля, в которых сохраняются географические координаты и численные значения NDVI для каждой даты.

Основные возможности программного комплекса:

- попиксельный расчет значений вегетационных индексов, в том числе NDVI;
- совмещение снимков (корегистрация). Снимки преобразуются в систему координат EPSG:4326 (WGS 84), которая считается стандартом для спутниковых навигационных систем;
- изменение пространственного разрешения (при необходимости). Для передискретизации данных на единую сетку применяется метод билинейной интерполяции. Этот метод определяет значение пикселя выходного растра как взвешенное среднее значений четырех ближайших пикселей исходного изображения с учетом расстояния до их центров;
- объединение данных нескольких растровых изображений. Вместо того чтобы предварительно выполнять вычислительно затратную процедуру предварительной «склейки» (объединения) нескольких спутниковых сцен, покрывающих разные участки одного поля, в единый растровый файл, система

автоматически агрегирует все релевантные данные в единую таблицу для каждого конкретного поля. Данные из различных сцен, соответствующие границам одного поля, извлекаются и сохраняются в структурированном виде в итоговом файле формата .csv;

- автоматизированная обработка ошибок позволяет избежать предварительной «ручной» проверки изображений. Это не только исключает субъективный человеческий фактор, но и значительно повышает скорость и надежность процесса интеграции больших массивов спутниковой информации в систему.

Для снижения шума полученные попиксельные ежедневные композиты были преобразованы в еженедельные методом скользящего окна (значение NDVI для n -го дня рассчитывалось как среднее значение $NDVI_{n-3}, \dots, NDVI_{n+3}$).

На основе этих данных для каждого класса культур по районам были построены усредненные временные ряды NDVI, характеризующие сезонную динамику вегетации. Для средних временных рядов еженедельных композитов NDVI каждого класса растительности (со 121-го по 296-й календарные дни года (начало мая – конец октября, с шагом в 7 дней) рассчитывались значения максимума индекса $NDVI_{\max}$ и календарного дня наступления максимума $DOY_{NDVI_{\max}}$. Сформированные временные ряды использовались для классификации пахотных земель.

Классификация

Классификация осуществлялась с применением алгоритма RF («лес деревьев») библиотеки scikit-learn для Python при числе деревьев, равном 50. В качестве обучающей выборки использовались верифицированные данные о распределении выбранных классов растительности на пахотных землях Хабаровского района (предварительно была проведена экспертная оценка результатов работы классификатора) [15]. Построенная и валидированная модель применялась для автоматизированной классификации сельскохозяйственных полей муниципальных районов: Вяземского, Бикинского и им. Лазо.

Для оценки качества модели рассчитывались общая точность (OA, %) и F1-метрика по формулам

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \cdot 100\%;$$

$$F_1 = \frac{TP}{TP + (FP + FN)/2},$$

где TP — верно предсказанное отнесение пикселя к данному классу; TN — верно предсказанное отнесение пикселя к другому классу; FP — неверное отнесение пикселя к данному классу; FN — неверное отнесение пикселя к другому классу.

Архитектура системы мониторинга сельскохозяйственных земель юга Дальнего Востока

Фундаментальный принцип архитектуры программного комплекса, используемого для хранения и визуализации информации, — разделение данных и их визуального представления: хранение информации обеспечивается реляционной базой данных, а ее просмотр и аналитическая обработка осуществляются в ГИС. В качестве источников исходных данных используются снимки с различных спутников, а также композитные изображения на основе инфракрасного и красного каналов спутника «Метеор-М» за период с 2019 по 2024 гг. Совокупный объем архивных данных превышает 4 ТБ.

Организация базы данных реализована в виде кластерной структуры, где каждый элемент соответствует отдельному субъекту Российской Федерации. Логическая структура базы данных состоит из нескольких взаимосвязанных модулей [18]. Модуль «Атрибутивная информация о полях» аккумулирует условно-постоянные данные, включая классификатор сельскохозяйственных культур. Модуль «Информация о вегетационных индексах» предназначен для хранения еженедельных композитных значений индексов, вычисленных по различным алгоритмам. Центральным связующим звеном выступает модуль «Геоинформация», содержащий пространственные данные о границах полей, их площади, агрохимических характеристиках, показателях урожайности и иной агрономической информации. На текущий момент в базу данных загружены сведения о севообороте примерно 3 тыс. полей общей площадью свыше 300 тыс. га, что эквивалентно более чем 40 млн пикселей спутниковых снимков.

Дополнительно был разработан модуль загрузки данных, оснащенный графическим интерфейсом, который автоматически извлекает идентификаторы полей из файлов для корректной загрузки в базу данных. Его функционал включает предварительную проверку данных и пакетную обработку, что повышает скорость и надежность наполнения базы.

Архитектура ГИС построена по трехзвенной модели с применением микросервисного подхода. Система интегрирует базу данных на PostgreSQL с расширением PostGIS для работы с пространственными данными, бэкенд-сервер на Python с использованием фреймворка GeoDjango и полностью переработанное фронтенд-приложение на Vue.js 3 с библиотекой OpenLayers. Дополнительно в систему внедрен сервис GeoServer, предназначенный для эффективной публикации тайлов спутниковых снимков, что обеспечивает высокую производительность при работе с растровыми данными. Все компоненты системы контейнеризированы с помощью технологии Docker, что гарантирует их изолированность, масштабируемость и упрощает процесс развертывания в различных средах.

Геоинформационная система предоставляет пользователям функционал для просмотра картографических материалов, редактирования границ полей, экспорта данных с возможностью предварительной фильтрации, построения гра-

фиков временных рядов вегетационных индексов и визуализации результатов попиксельной классификации пахотных земель. Доступ к функциональным возможностям системы разграничен в соответствии с уровнем привилегий различных категорий пользователей — от сельскохозяйственных производителей до специалистов, обладающих расширенными правами.

3 Результаты

Общая точность классификации пахотных земель в 2024 г. на полях Вяземского, Бикинского районов и района им. Лазо оказалась равной 87%. Значения метрики F1 для разных типов растительности составили: соя — 0,93; зерновые — 0,87; многолетние травы — 0,61; залежь — 0,85. Для оценки устойчивости разработанного классификатора была проведена трехкратная кросс-валидация (табл. 1). Для каждой из трех итераций точность находилась на уровне 87%–88% (отсутствие вариации между итерациями показывает надежность классификатора), значение F1-метрики менялось от 0,77 до 0,83.

Дополнительно на примере Бикинского района проводилась экспертная верификация полученных результатов. Большая часть ошибок классификации была связана с отнесением полей с многолетними травами к неиспользуемым землям. В общем, несоответствие результатов классификации фактическому типу растительности было выявлено на 3 из 410 полей, что составляет 10,5% от общего числа и соответствует кросс-валидационной ОА. Для каждого поля были сформированы кривые сезонного хода NDVI, для каждого класса рассчитаны средние значения сезонных временных рядов NDVI.

На рис. 1 представлены кривые сезонной динамики средних значений индекса NDVI в 2024 г., отражающие фенологические особенности основных сельскохозяйственных культур Хабаровского края: сои, зерновых, многолетних трав, а также залежных земель.

Как видно из рис. 1, наиболее быстрый рост NDVI в мае–июне для всех 4-х классов растительности характерен для самых южных (Бикинского и Вяземского) районов. Для сельскохозяйственных культур это связано, в том числе, и с более ранними сроками посева в сравнении с Хабаровским районом (см. рис. 1, б и 1, в).

Таблица 1 Результаты кросс-валидации алгоритма классификации пахотных земель

Итерация	ОА, %	F1
1	87,2	0,77
2	87,4	0,79
3	88,2	0,83

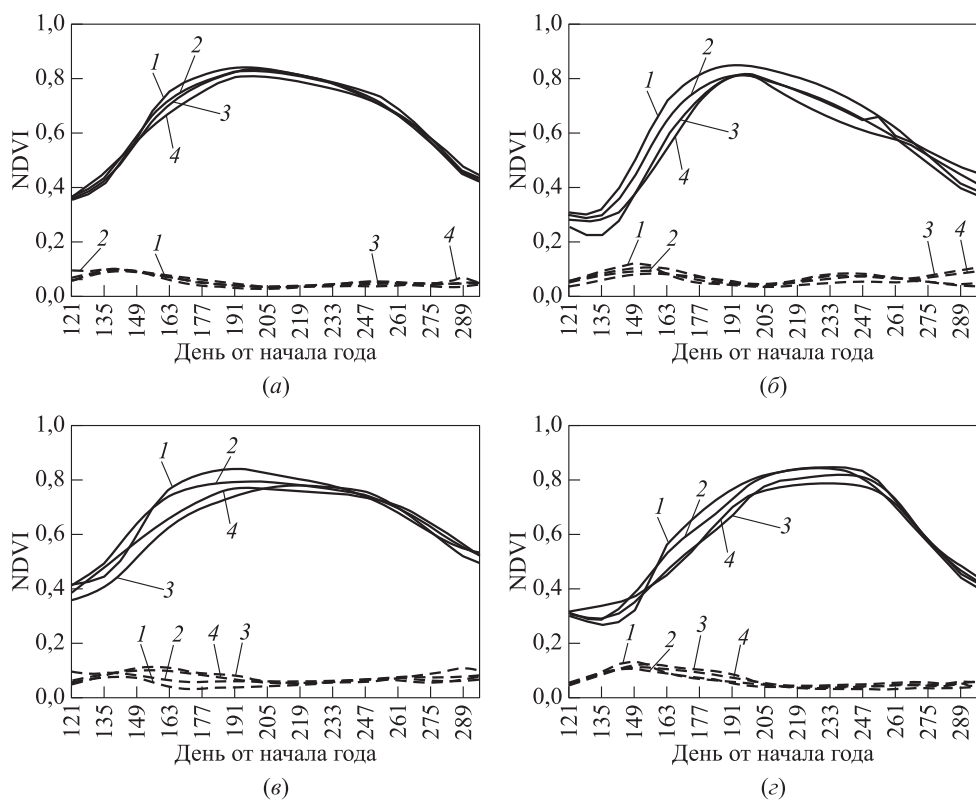


Рис. 1 Сезонный ход NDVI для разных классов растительности южной части Хабаровского края по данным «Метеор-М» № 2: (а) залежь, (б) зерновые, (в) многолетние травы, (г) соя; 1 — Бикинский район; 2 — Вяземский район; 3 — район им. Лазо; 4 — Хабаровский район; сплошные кривые — среднее; штриховые кривые — стандартное отклонение

Кривые сезонного хода NDVI имели характерный вид для каждого класса. Для сои значения временного ряда NDVI плавно нарастали с мая по август, максимум вегетации соответствовал 219–240 календарным дням, начиная с конца сентября наблюдалось резкое снижение величины NDVI (см. рис. 1, г). Многолетние травы характеризовались довольно быстрым ростом NDVI и достижением максимума NDVI в июне–июле (см. рис. 1, в). На полях с зерновыми культурами вследствие раннего сева также наблюдался быстрый рост значений NDVI, при этом максимум NDVI наблюдался в июле, после уборки в конце июля значения NDVI снижались достаточно плавно, что обусловлено практикой подсева многолетних трав на поля с зерновыми культурами (см. рис. 1, б). Для залежи наблюдался затяжной максимум, визуальные различия между кривыми сезонного хода NDVI для разных районов не были ярко выражены (см. рис. 1, а).

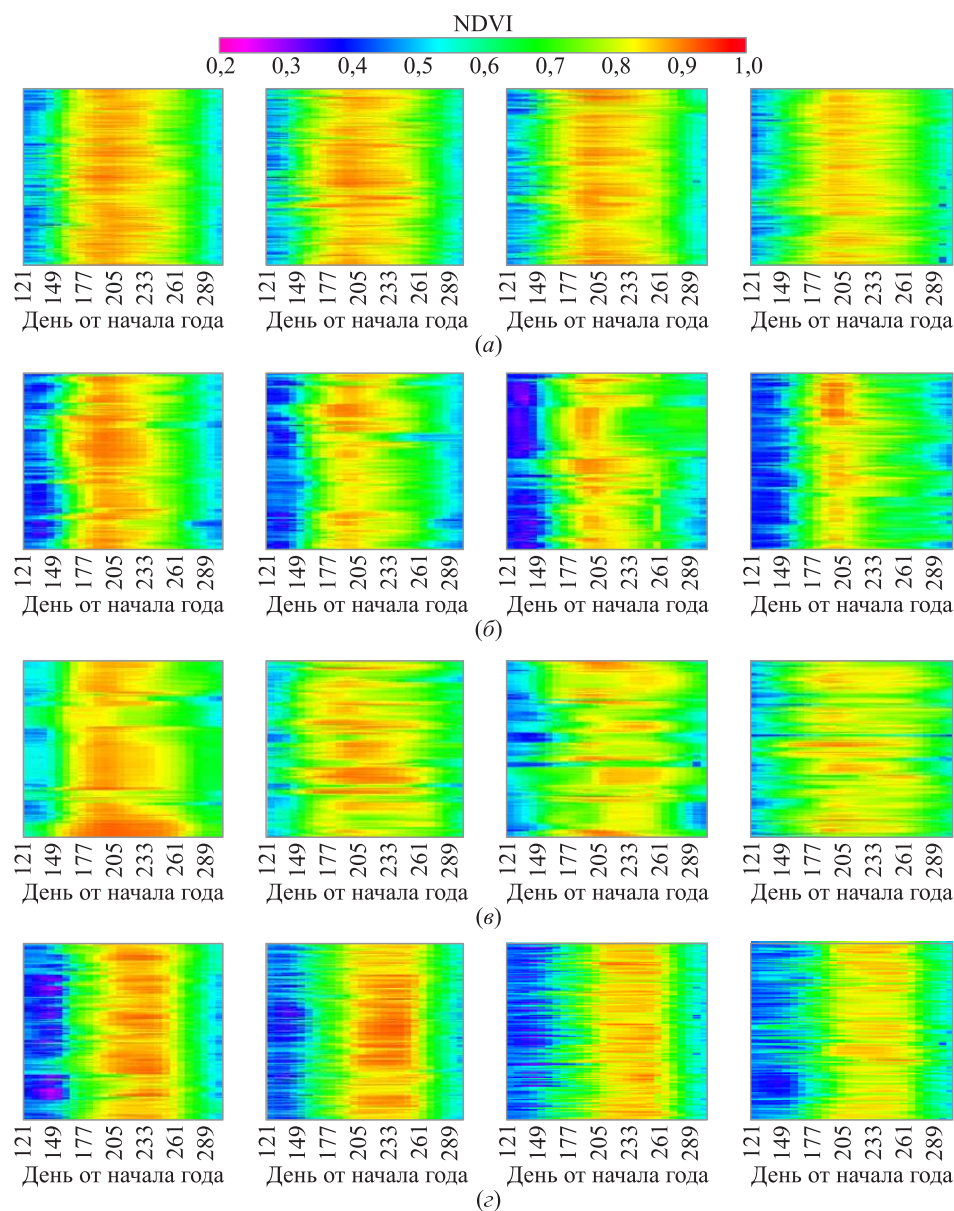


Рис. 2 Тепловые карты распределения NDVI для разных классов растительности ((a) залежь; (б) зерновые; (в) многолетние травы; (г) соя) на полях муниципальных районов Хабаровского края в 2024 г.: первая колонка — Бикинский район; вторая колонка — Вяземский район; третья колонка — район им. Лазо; последняя колонка — Хабаровский район

Тепловые карты, представленные на рис. 2, показали, что для зерновых культур в Хабаровском, Вяземском и районе им. Лазо значения максимума NDVI действительно оказались численно ниже, чем в Бикинском районе, и распределены на временной оси в большем интервале.

Для остальных культур и неиспользуемых земель тепловая карта Бикинского района также была более яркой в сравнении с тепловыми картами прочих территорий. Распределения $DOY_{NDVI_{max}}$ для сои, многолетних трав и залежи во всех четырех районах имели схожий характер.

В табл. 2 представлены основные показатели, которые характеризуют сезонные ряды значений NDVI для разных районов и классов растительности. В среднем максимальные значения NDVI для разных классов растительности находились в диапазоне от 0,79 до 0,85. День наступления максимума NDVI для зерновых и залежи соответствовал интервалам 193–197 и 196–200 календарных дней соответственно. Максимум NDVI для сои достигался существенно позже — в период с 223-го по 232-й календарный день с начала года.

Проведенный двухфакторный дисперсионный анализ выявил, что значения $NDVI_{max}$ и $DOY_{NDVI_{max}}$ значимо различались для разных районов и классов растительности ($p < 0,001$).

Таблица 2 Основные характеристики сезонных временных рядов NDVI пахотных земель в муниципальных образованиях Хабаровского края (по данным спутника «Метеор-М», 2024 г.)

Муниципальный район	$NDVI_{max} \pm \sigma$	$DOY_{NDVI_{max}} \pm \sigma$
Залежь		
Бикинский район	$0,85 \pm 0,02$	$196,0 \pm 7,7$
Вяземский район	$0,84 \pm 0,03$	$197,2 \pm 8,8$
Район им. Лазо	$0,84 \pm 0,03$	$198,3 \pm 8,1$
Хабаровский район	$0,82 \pm 0,03$	$199,6 \pm 9,0$
Зерновые		
Бикинский район	$0,85 \pm 0,03$	$193,2 \pm 6,5$
Вяземский район	$0,82 \pm 0,04$	$194,3 \pm 6,9$
Район им. Лазо	$0,82 \pm 0,03$	$197,1 \pm 6,7$
Хабаровский район	$0,81 \pm 0,05$	$194,9 \pm 6,9$
Многолетние травы		
Бикинский район	$0,84 \pm 0,03$	$192,2 \pm 8,1$
Вяземский район	$0,81 \pm 0,05$	$199,8 \pm 19,5$
Район им. Лазо	$0,80 \pm 0,05$	$214,1 \pm 18,7$
Хабаровский район	$0,79 \pm 0,05$	$207,3 \pm 22,2$
Соя		
Бикинский район	$0,85 \pm 0,03$	$223,3 \pm 11,2$
Вяземский район	$0,85 \pm 0,04$	$230,1 \pm 13,9$
Район им. Лазо	$0,83 \pm 0,04$	$231,6 \pm 15,6$
Хабаровский район	$0,80 \pm 0,04$	$226,3 \pm 16,7$

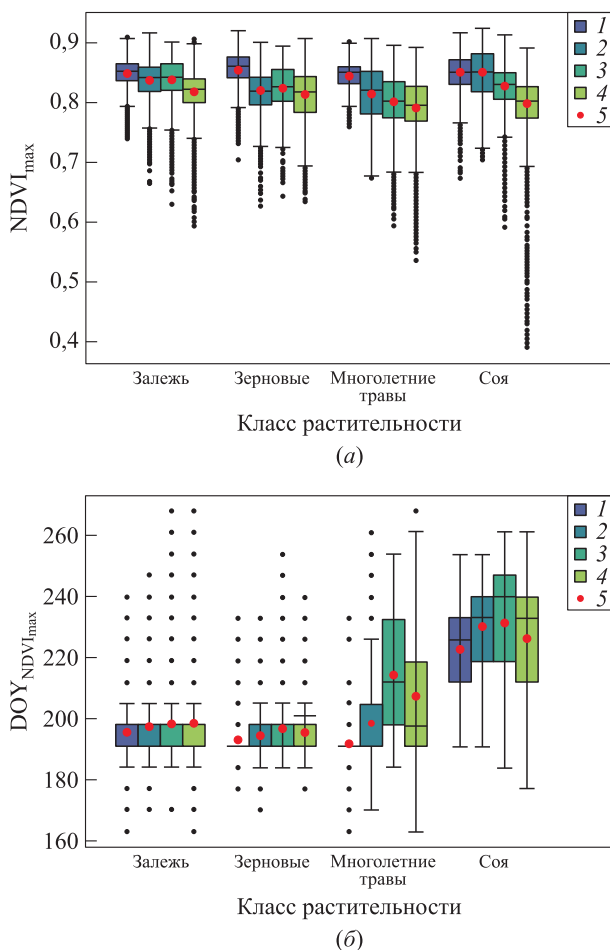


Рис. 3 Диаграмма размаха значений $NDVI_{\max}$ (а) и $DOY_{NDVI_{\max}}$ (б) по данным спутника «Метеор-М» на полях муниципальных образований Хабаровского края в 2024 г.: 1 — Бикинский район; 2 — Вяземский район; 3 — район им. Лазо; 4 — Хабаровский район; 5 — среднее значение

Как следует из рис. 3, вариативность максимума NDVI примерно одинакова для одного и того же класса растительности, но в разных муниципальных районах края. Вместе с тем разброс значений NDVI для разных полей минимален в Бикинском районе. Вариативность $DOY_{NDVI_{\max}}$ для залежи и зерновых была не очень высокой и примерно равной в четырех районах. Вариативность $DOY_{NDVI_{\max}}$ для сои и многолетних трав существенно выше, при этом в районе им. Лазо и в Хабаровском районе в 1,5–2 раза выше, чем в Бикинском районе. Высокая вариабельность дня наступления максимума у многолетних трав и сои

Таблица 3 Площадь пахотных земель в муниципальных образованиях Хабаровского края под идентифицированными классами растительности (по данным спутника «Метеор-М», 2024 г.), га

Класс растительности	Бикинский район	Вяземский район	Район им. Лазо	Хабаровский район (обучающая выборка)	Итого
Залежь	6 608	10 478	4 126	14 384	35 597
Зерновые	1 282	1 473	900	3 801	7 457
Многолетние травы	109	622	507	2 024	3 263
Соя	2 172	10 222	15 878	15 183	43 456
Итого	10 173	22 796	21 411	35 392	90 939

указывает на необходимость дифференцированного подхода к их классификации с учетом специфики сорта и используемых агротехнологий.

Различия временных рядов NDVI, полученных по снимкам «Метеор-М» для разных районов, обусловили возможность проведения попиксельной классификации на основе методов машинного обучения. Было проведено попиксельное определение классов растительности в границах полей (с точным соответствием контурам ЕФИС), уточнены площади посевов и сгенерированы тематические карты. Общая площадь идентифицированных классов в границах пахотных земель по муниципальным образованиям представлена в табл. 3.

Как видно из табл. 3, в целом в южных районах Хабаровского края площадь залежи в 2024 г. превышала 35 000 га, что составляло более 30% от пахотных земель, занесенных в ЕФИС ЗСН. Доля сои была чуть меньше 50%. Отметим, что идентифицировались только основные классы растительности, на части полей культивировались овощи, кукуруза и т. д. Однако доля этих культур была

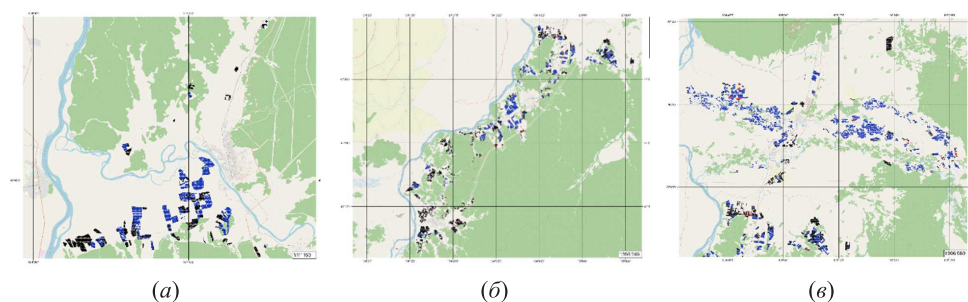


Рис. 4 Карты пахотных земель южных районов Хабаровского края, построенные по данным спутника «Метеор-М» в 2024 г. (синим цветом обозначена соя; красным — зерновые; зеленым — многолетние травы; черным — залежь): (а) Бикинский район, (б) Вяземский район, (в) район им. Лазо

достаточно низкой. На рис. 4 представлены карты пахотных земель Бикинского, Вяземского и района им. Лазо, построенные по результатам анализа спутниковых изображений «Метеор-М» в 2024 г.

4 Заключение

В ходе проведенного исследования было установлено, что сезонные ряды NDVI, полученные по данным спутника «Метеор-М», для пахотных земель южных районов Хабаровского края в 2024 г. для одних и тех же классов растительности в разных районах имеют схожий вид, при этом численные значения $NDVI_{max}$ и $DOY_{NDVI_{max}}$ заметно различаются для разных классов. На основе валидированных данных о севообороте в Хабаровском районе была проведена классификация и построены карты пахотных земель трех южных районов Хабаровского края. Средняя точность метода по результатам трехкратной кросс-валидации оказалась равной 87,6%, значение F1-метрики — 0,80. Для разных классов растительности значение F1 менялось от 0,61 до 0,93. Было установлено, что доля залежи от общей площади пахотных земель превысила 30%, а посевы сои в 2024 г. составили 48% для южных районов Хабаровского края. Полученные данные внесены в информационную систему.

Литература

1. Weiss M., Jacob F., Duveiller G. Remote sensing for agricultural applications: A meta-review // Remote Sens. Environ., 2020. Vol. 236. Art. 111402. doi: 10.1016/j.rse.2019.111402.
2. Karmakar P., Teng S. W., Murshed M., Pang S., Li Y., Lin H. Crop monitoring by multimodal remote sensing: A review // Remote Sensing Applications Society Environment, 2024. Vol. 33. Art. 101093. 15 p. doi: 10.1016/j.rsase.2023.101093.
3. Никишин Д. А. Моделирование динамических геообъектов на основе концепции фазовой вариантности // Системы и средства информатики, 2024. Т. 34. № 2. С. 95–106. doi: 10.14357/08696527240207. EDN: GCJQET.
4. Архипов П. О., Филиппских С. Л. Распознавание аномалий на разновременных панорамах с использованием нейросетевого метода консолидации моделей // Системы и средства информатики, 2023. Т. 33. № 2. С. 13–24. doi: 10.14357/08696527230202. EDN: TSMYNA.
5. Розенберг И. Н., Дулин С. К., Дулина Н. Г. Геоинформационная система — инструмент цифровой трансформации геоданных // Системы и средства информатики, 2022. Т. 32. № 1. С. 46–54. doi: 10.14357/08696527220104. EDN: SJGRML.
6. Прохорец И. О., Степанов А. С. Картографирование земель сельскохозяйственного назначения Хабаровского края методами машинного обучения с использованием изображений Sentinel-2 // Системы и средства информатики, 2024. Т. 34. № 1. С. 57–69. doi: 10.14357/08696527240105. EDN: YEZNXD.
7. Лысенко М. Н., Ложковой П. Н. Использование спутников для дистанционного зондирования Земли: проблемы и перспективы // Ж. Сибирского федерального

- университета. Сер. Гуманитарные науки, 2022. Т. 15. № 12. С. 1763–1767. doi: 10.17516/1997-1370-0950. EDN: EQTEVT.
8. Wang L., Wang J., Liu Z., Zhu J., Qin F. Evaluation of a deep-learning model for multispectral remote sensing of land use and crop classification // Crop J., 2022. Vol. 10. Iss. 5. P. 1435–1451. doi: 10.1016/j.cj.2022.01.009.
 9. Xia C., Zhang Y. Comparison of the use of Landsat 8, Sentinel-2, and Gaofen-2 images for mapping soil pH in Dehui, Northeastern China // Ecol. Inform., 2022. Vol. 70. Art. 101705. doi: 10.1016/j.ecoinf.2022.101705.
 10. Pascual I., Edrosa R., Machado F., Thiebault M. Operational Crop classification with Saocom L-Band Sar images in the Pampas Region, Argentina // Biennial Congress of Argentina Proceedings. — IEEE, 2024. Art. 10735849. 7 p. doi: 10.1109/ARGENCON62399.2024.10735849.
 11. Plotnikov D., Kolbudaev P., Matveev A., Proshin A., Polyanskiy I. Accuracy assessment of atmospheric correction of KMSS-2 Meteor-M #2.2 data over Northern Eurasia // Remote Sensing, 2023. Vol. 15. Art. 4395. 18 p. doi: 10.3390/rs15184395.
 12. Полянский И. В., Жуков Б. С., Кондратьева Т. В., Прохорова С. А., Сметанин П. С. Комплекс многозональной спутниковой съемки среднего разрешения для гидрометеорологических космических аппаратов // Современные проблемы дистанционного зондирования Земли из космоса, 2019. Т. 16. № 6. С. 83–92. doi: 10.21046/2070-7401-2019-16-6-83-92. EDN: DMKHCB.
 13. Кашицкий А. В., Луян Е. А., Плотников Д. Е., Толтин В. А. Анализ возможности использования данных различного пространственного разрешения при проведении мониторинга объектов // Современные проблемы дистанционного зондирования Земли из космоса, 2023. Т. 20. № 2. С. 60–74. doi: 10.21046/2070-7401-2023-20-2-60-74. EDN: CCHNYC.
 14. Kolbudaev P. A., Plotnikov D. E., Loupian E. A., Proshin A. A., Matveev A. M. The methods and automatic technology aimed at imagery georeferencing, cloud screening, atmospheric and radiometric correction of KMSS-M satellite data // E3S Web Conf., 2021. Vol. 333. Art. 01006. 5 p. doi: 10.1051/e3sconf/202133301006.
 15. Илларионова Л. В., Степанов А. С., Дубровин К. Н. и др. Опыт использования данных прибора КМСС КА «Метеор-М» № 2 для задач мониторинга сельскохозяйственных угодий юга Хабаровского края // Современные проблемы дистанционного зондирования Земли из космоса, 2025. Т. 22. № 1. С. 81–92. doi: 10.21046/2070-7401-2025-22-1-81-92. EDN: DHELYQ.
 16. Луян Е. А., Прошин А. А., Бурцев М. А. и др. Опыт эксплуатации и развития центра коллективного пользования системами архивации, обработки и анализа спутниковых данных (ЦКП «ИКИ Мониторинг») // Современные проблемы дистанционного зондирования Земли из космоса, 2019. Т. 16. № 3. С. 151–170. doi: 10.21046/2070-7401-2019-16-3-151-170. EDN: OVWKZV.
 17. Илларионова Л. В., Степанов А. С., Фомина Е. А. и др. Программный комплекс автоматизированной обработки композитных изображений спутника Метеор для получения сезонных временных рядов NDVI для сельхозугодий Хабаровского края: Свидетельство о государственной регистрации программы для ЭВМ № 2024684901 от 22.10.2024. EDN: NRNGOJ.
 18. Федосеев Д. В., Степанов А. С., Илларионова Л. В. и др. Разработка информационной системы для хранения и обработки результатов мониторинга сельскохозяй-

ственных угодий юга Дальнего Востока // Системы и средства информатики, 2024. Т. 34. № 1. С. 44–56. doi: 10.14357/08696527240104. EDN: XPPVVV.

Поступила в редакцию 04.09.2025

Принята к публикации 15.10.2025

MAPPING OF Khabarovsk Krai Arable Lands Using Meteor-M No. 2 Satellite Data

*A. S. Stepanov¹, L. V. Illarionova², K. N. Dubrovin², E. A. Fomina²,
and A. L. Verkhoturov²*

¹Far Eastern Research Institute of Agriculture, 13 Klubnaya Str., Vostochnoe 680521, Khabarovsk Krai, Russian Federation

²Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, 65 Kim Yu Chen Str., Khabarovsk 680000, Russian Federation

Abstract: The article considers the possibility of using weekly composite images from the Meteor-M No. 2 satellite to classify arable lands in Khabarovsk Krai. For four vegetation classes (soybeans, grain crops, perennial grasses, and fallow land), average Normalized Difference Vegetation Index (NDVI) seasonal variation series were constructed for municipal districts in the south of Khabarovsk Krai in 2024 and the main characteristics — the NDVI maximum values and the day of the maximum — were calculated. Statistically significant differences in the indicators for the average NDVI time series for different vegetation classes were revealed ($p < 0.0001$). Using validated data from Khabarovsk KRAI, a classification of arable lands in the Bikinsky, Vyazemsky, and Lazovsky Districts was conducted using machine learning (the Random Forest algorithm). The average accuracy of the method based on the results of three-fold cross-validation was equal to 87.6%. For different vegetation classes, the F1 metric value ranged from 0.61 to 0.93. Arable land maps were created for the southern regions of Khabarovsk Krai. It was found that fallow land accounts for over 30% of the region's total arable land area, while soybean crops accounted for 48% in 2024. The mapping results were entered into the developed geographic information system.

Keywords: crop land; machine learning; classification; satellite monitoring; GIS

DOI: 10.14357/08696527250408

EDN: SGWRZO

Acknowledgments

The research was supported by the Russian Science Foundation (project No. 24-11-20030) and the Government of Khabarovsk Krai (Agreement No. 106C/2024 dated July 31, 2024).

References

1. Weiss, M., F. Jacob, and G. Duveiller. 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* 236:111402. doi: 10.1016/j.rse.2019.111402.
2. Karmakar, P., S. W. Teng, M. Murshed, S. Pang, Y. Li, and H. Lin. 2024. Crop monitoring by multimodal remote sensing: A review. *Remote Sensing Applications Society Environment* 33:101093. 15 p. doi: 10.1016/j.rsase.2023.101093.
3. Nikishin, D. A. 2024. Modelirovanie dinamicheskikh geoob"ektov na osnove kontseptsii fazovoy variantnosti [An approach to modeling dynamic geoobjects based on the concept of phase variation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 34(2):95–106. doi: 10.14357/08696527240207. EDN: GCJQET.
4. Arkhipov, P. O., and S. L. Philippskih. 2023. Raspoznavanie anomalii na raznovremennykh panoramakh s ispol'zovaniem neyrosetevogo metoda konsolidatsii modeley [Recognition of anomalies on multitime panoramas using the neural network method of model amalgamation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(2):13–24. doi: 10.14357/08696527230202. EDN: TSMYNA.
5. Rozenberg, I. N., S. K. Dulin, and N. G. Dulina. 2022. Geoinformatsionnaya sistema — instrument tsifrovoy transformatsii geodannykh [Geographic information system — tool for digital geodata transformation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(1):46–54. doi: 10.14357/08696527220104. EDN: SJGRML.
6. Prokhorets, I. O., and A. S. Stepanov. 2024. Kartografirovaniye zemel' sel'skokhozyaystvennogo naznacheniya Khabarovskogo kraia metodami mashinnogo obucheniya s ispol'zovaniem izobrazheniy [Mapping of the Khabarovsk Region arable lands by machine learning using Sentinel-2 images]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 34(1):57–69. doi: 10.14357/08696527240105. EDN: YEZXND.
7. Lysenko, M. N., and P. N. Lozhkovej. 2022. Ispol'zovanie sputnikov dlya distantsionnogo zondirovaniya Zemli: problemy i perspektivy [Usage of satellites for the Earth remote sensing: Problems and prospects]. *Zh. Sibirskogo federal'nogo universiteta. Ser. Gumanitarnye nauki* [J. Siberian Federal University. Humanities and Social Sciences] 15(12):1763–1767. doi: 10.17516/1997-1370-0950. EDN: EQTEVT.
8. Wang, L., J. Wang, Z. Liu, J. Zhu, and F. Qin. 2022. Evaluation of a deep-learning model for multispectral remote sensing of land use and crop classification. *Crop J.* 10(5):1435–1451. doi: 10.1016/j.cj.2022.01.009.
9. Xia, C., and Y. Zhang. 2022. Comparison of the use of Landsat 8, Sentinel-2, and Gaofen-2 images for mapping soil pH in Dehui, Northeastern China. *Ecol. Inform.* 70:101705. doi: 10.1016/j.ecoinf.2022.101705.
10. Pascual, I., R. Edrosa, F. Machado, and M. Thiebault. 2024. Operational crop classification with Saocom L-band SAR images in the Pampas Region, Argentina. *IEEE Biennial Congress of Argentina Proceedings*. IEEE. Art. 10735849. 7 p. doi: 10.1109/ARGENCON62399.2024.10735849.
11. Plotnikov, D., P. Kolbudaev, A. Matveev, A. Proshin, and I. Polyanskiy. 2023. Accuracy assessment of atmospheric correction of KMSS-2 Meteor-M #2.2 data over Northern Eurasia. *Remote Sensing* 15(18):4395. 18 p. doi: 10.3390/rs15184395.
12. Polyanskiy, I. V., B. S. Zhukov, T. V. Kondratieva, S. A. Prokhorova, and P. S. Smetanin. 2019. Kompleks mnogozonal'noy sputnikovoy s"emki srednego raz-

- resheniya dlya gidrometeorologicheskikh kosmicheskikh apparatov [Medium-resolution multispectral satellite imaging system for hydrometeorological spacecraft]. *Sovremennye roblemy distantsionnogo zondirovaniya Zemli iz kosmosa* [Modern Problems of Remote Sensing of the Earth from Space] 16(6):83–92. doi: 10.21046/2070-7401-2019-16-6-83-92. EDN: DMKHCB.
13. Kashnitskii, A. V., E. A. Loupian, D. E. Plotnikov, and V. A. Tolpin. 2023. Analiz vozmozhnosti ispol'zovaniya dannykh razlichnogo prostranstvennogo razresheniya pri provedenii monitoringa ob"ektov [Analysis of the possibility of using different spatial resolution data for objects monitoring]. *Sovremennye roblemy distantsionnogo zondirovaniya Zemli iz kosmosa* [Modern Problems of Remote Sensing of the Earth from Space] 20(2):60–74. doi: 10.21046/2070-7401-2023-20-2-60-74. EDN: CCHNYC.
 14. Kolbudaev, P. A., D. E. Plotnikov, E. A. Loupian, A. A. Proshin, and A. M. Matveev. 2021. The methods and automatic technology aimed at imagery georeferencing, cloud screening, atmospheric and radiometric correction of KMSS-M satellite data. *E3S Web Conf.* 333:01006. 5 p. doi: 10.1051/e3sconf/202133301006.
 15. Illarionova, L. V., A. S. Stepanov, K. N., Dubrovin, *et al.* 2025. Opyt ispol'zovaniya dannykh pribora KMSS KA "Meteor-M" № 2 dlya zadach monitoringa sel'skokhozyaystvennykh ugodyy yuga Khabarovskogo kraya [Experience of using data from the Meteor-M No. 2 KMSS sensor for cropland monitoring in the south of Khabarovsk Krai]. *Sovremennye roblemy distantsionnogo zondirovaniya Zemli iz kosmosa* [Modern Problems of Remote Sensing of the Earth from Space] 22(1):81–92. doi: 10.21046/2070-7401-2025-22-1-81-92. EDN: DHELYQ.
 16. Loupian, E. A., A. A. Proshin, M. A. Bourtsev, *et al.* 2019. Opyt ekspluatatsii i razvitiya tsentra kollektivnogo pol'zovaniya sistemami arkhivatsii, obrabotki i analiza sputnikovyykh dannykh (TsKP "IKI Monitoring" [Experience of development and operation of the IKI-Monitoring center for collective use of systems for archiving, processing and analyzing satellite data]. *Sovremennye roblemy distantsionnogo zondirovaniya Zemli iz kosmosa* [Modern Problems of Remote Sensing of the Earth from Space] 16(3):151–170. doi: 10.21046/2070-7401-2019-16-3-151-170. EDN: OVWKZV.
 17. Illarionova, L., A. Stepanov, E. Fomina, K. Dubrovin, and A. Bordakov. 2024. Programmy kompleks avtomatizirovannoy obrabotki kompozitnykh izobrazheniy sputnika Meteor dlya polucheniya sezonnykh vremennykh ryadov NDVI dlya sel'khozugodyy Khabarovskogo kraya [A software package for automated processing of Meteor satellite composite images to obtain seasonal NDVI time series for agricultural lands in Khabarovsk Krai]. Certificate RF of State Registration of Computer Programs No. 2024684901. EDN: NRNGOJ.
 18. Fedoseev, D. V., A. S. Stepanov, L. V. Illarionova, V. A. Eliseev, I. O. Prokhorets, A. N. Frolov, and A. I. Shamurina. 2024. Razrabotka informatsionnoy sistemy dlya khraneniya i obrabotki rezul'tatov monitoringa sel'skokhozyaystvennykh ugodyy yuga Dal'nego Vostoka [Development of information system for storing and processing the results of agricultural land monitoring in the south of the Far East]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 34(1):44–56. doi: 10.14357/08696527240104. EDN: XPPVVV.

Received September 4, 2025

Accepted October 15, 2025

Contributors

Stepanov Alexey S. (b. 1974) — Doctor of Science in pharmacy, leading scientist, Far Eastern Research Institute of Agriculture, 13 Klubnaya Str., Vostochnoe 680521, Khabarovsk Krai, Russian Federation; stepanfx@mail.ru

Illarionova Lyubov V. (b. 1977) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, 65 Kim Yu Chen Str., Khabarovsk 680000, Russian Federation; illarionova_l@list.ru

Dubrovin Konstantin N. (b. 1993) — scientist, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, 65 Kim Yu Chen Str., Khabarovsk 680000, Russian Federation; k.dubrovin@ccfebras.ru

Fomina Elizaveta A. (b. 1998) — engineer, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, 65 Kim Yu Chen Str., Khabarovsk 680000, Russian Federation; eliz37@mail.ru

Verkhoturov Andrey L. (b. 1986) — scientist, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences, 65 Kim Yu Chen Str., Khabarovsk 680000, Russian Federation; andrey@ccfebras.ru

СОВРЕМЕННЫЕ МЕТОДЫ ИЗВЛЕЧЕНИЯ ГЕОГРАФИЧЕСКОГО ЗНАНИЯ ИЗ БОЛЬШИХ МАССИВОВ ПРОСТРАНСТВЕННЫХ ДАННЫХ

Д. А. Никишин¹

Аннотация: Рассмотрены основные концепции представления пространственных данных и ряд современных методов анализа и извлечения географических знаний из баз географических данных, включая современные тенденции развития географического искусственного интеллекта (ИИ). Поиск знаний в базах географических данных — нетривиальный процесс, который требует понимания фундаментальных географических концепций, учета особенностей пространственного или пространственно-временного представления геообъектов и сопряженных с этим специализированных алгоритмов и который сложно реализовать с помощью прямого переноса традиционных методов интеллектуального анализа данных. Географический ИИ (Geographical Artificial Intelligence, GeoAI) представляется перспективной технологией интеллектуальной обработки пространственных данных для решения различных задач (таких как кластеризация, классификация, сегментация, интерполяция и др.), особенно в контексте анализа пространственно-временных закономерностей. Полученные результаты должны послужить отправной точкой при разработке новых подходов к извлечению географического знания.

Ключевые слова: геоданные; базы географических данных; методы извлечения географических знаний; географический искусственный интеллект

DOI: 10.14357/08696527250409

EDN: CWPFFNM

1 Введение

Прогресс в области технологий ИИ, доступность высококачественных крупномасштабных географических данных, а также достижения в области высокопроизводительной обработки этих данных обусловили преобразования в ряде областей [1]. К настоящему моменту накоплены огромные объемы пространственных и пространственно-временных данных, собранных с помощью различных методов, включая системы глобального спутникового позиционирования (типа ГЛОНАСС), дистанционное зондирование Земли, геолокационные сервисы и некоторые другие. Обработка массивов этих данных, особенно высокого разрешения, открывает широкие возможности для получения новых знаний и лучшего понимания сложных географических явлений [2].

Вместе с тем многие традиционные методы пространственного анализа имеют ограничения обработки данных в условиях их большого размера, их высокой

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, dmnikishin@mail.ru

размерности (порядка сотен переменных), внутренней сложности структур данных, наличия явных или неявных пространственных отношений в структуре набора данных, неоднородности источников данных и др. Кроме того, появились некоторые новые типы данных, например траектории движущихся объектов¹, получаемые от устройств с определением местоположения, мультиспектральные изображения и т. п. Все это на фоне новых потребностей современных приложений обуславливает потребность в создании эффективных и действенных методов извлечения знаний из подобных наборов (массивов) пространственных данных [2].

Разработка теории, методологии и практики интеллектуального анализа пространственных данных и извлечения географических знаний из массивных и сложных пространственных баз данных в различных областях превратилась в настоящее время в активное направление исследований [3, 4].

Далее рассмотрены основные концепции и модели представления географических данных и ряд существующих методов анализа и извлечения из них географических знаний. Также представлен обзор современных тенденций развития географического ИИ.

2 Концептуальные модели пространственных данных

Концептуальные модели призваны отражать основные характеристики объектов и событий определенной предметной области с точки зрения пользователя и решаемой задачи. На концептуальном уровне можно выделить два основных подхода к моделированию.

В первом случае пространственная область рассматривается как фрагмент непрерывной поверхности, представляющей собой функцию местоположения в пространстве произвольной размерности (в частности, для географических приложений используются двух- или трехмерные модели), при этом на поверхности вводится регулярная или иррегулярная сеть (рис. 1), каждому из узлов

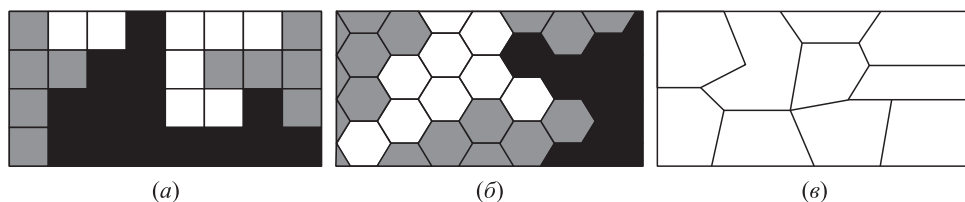


Рис. 1 Разновидности тесселяции (по [5]): (а) ячейки квадратной (прямоугольной) сети; (б) ячейки гексагональной сети; (в) ячейки иррегулярной сети

¹Описания траекторий движущихся объектов включают такие характеристики, как скорость и ускорение. Выявление закономерностей в этих данных является сложной задачей из-за темпоральной природы данных.

которой может быть сопоставлен однотипный одно- или многомерный вектор свойств (атрибутов) как пространственных, так и непространственных. Эта парадигма часто называется моделью поля или моделью тесселяции [5]. Модели данного класса хорошо подходят для моделирования однородных непрерывных изменений и позволяют реализовать эффективные структуры индексации. Характерным примером применения таких моделей служат данные дистанционного зондирования (съемки) местности, а также некоторые прикладные применения (геология, метеорология и др.).

Во втором случае концептуальная модель основана на парадигме моделирования отдельных географических объектов, выделяемых на местности и представляемых в виде конечных геометрических структур, каждой из которых может быть присвоен определенный набор непространственных свойств. В целом модель геообъекта можно охарактеризовать как «вектор» в общем, пространственно-семантическом контексте. Однако объектная модель более сложна для пространственного анализа, в частности при реализации оверлейных операций [5].

Объектная модель в принципе может использоваться и для представления концепции поверхности, в связи с чем модель нерегулярной тесселяции (см. рис. 1, в) может рассматриваться как переходная между этими двумя случаями.

Реализация объектной модели, в свою очередь, может иметь два основных варианта [5]. В первом случае пространственная информация о локализации представляется в явном («объектном») виде, что обеспечивает сравнительно простой доступ к геометрии отдельно взятого объекта. Однако информация о пространственных связях (так называемые «топологии») объектов в таких моделях зачастую представлена неявно, и ее извлечение представляет собой сложную задачу, требующую применения методов пространственного анализа данных. Во втором случае геометрия пространственных объектов реализована на основе связей агрегации отдельных пространственных элементов (так называемая «топологическая» модель), посредством чего становится возможным зафиксировать в явном виде топологические отношения между объектами.

3 Методы анализа и извлечения географических знаний на основе пространственных данных

Особенность *поиска знаний* в базах географических данных для их извлечения заключается в том, что, хотя многие алгоритмы анализа данных по своей сути могут быть распространены на многомерные пространственные объекты, не все они могут быть применены для моделирования географического пространства [5] ввиду наличия у последнего определенных ограничений. Извлечение пространственных характеристик географических объектов и установление пространственных (топологических) взаимосвязей совокупности объектов — весьма трудоемкий этап процесса обнаружения знаний [5]. Далее кратко рассмотрены основные современные подходы к анализу с использованием географических данных.

Методы кластеризации можно в целом разделить на две группы: кластеризация с разделением и иерархическая кластеризация. Методы первой группы делят набор данных на несколько неперекрывающихся кластеров, а текущий элемент данных относится к «ближайшему» кластеру на основе меры близости или различия. Во втором случае элементы данных организуются в последовательность вложенных разделов или группировок.

Используются три основных типа кластерного анализа [2]: пространственная кластеризация (определяется свойствами местоположения и расстояния), регионализация (кластеризация в пространственно смежные кластеры-«регионы») и обнаружение «горячих» точек (необычных концентраций событий в пространстве).

Вместе с тем предположение о выпуклых кластерах не подходит для многих наборов географических данных (рис. 2, а). Для решения этой проблемы в работе [6] был разработан алгоритм, основанный на плотности точечных данных, который позволяет находить кластеры произвольной формы (рис. 2, б).

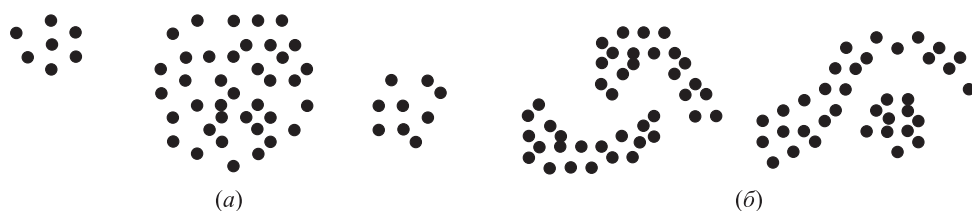


Рис. 2 Пространственные кластеры (по [5]): (а) выпуклой формы; (б) произвольной формы

Методы классификации и регрессионного анализа. Классификация как метод отнесения элементов данных к классам (категориям) в соответствии с их свойствами предполагает наличие некоторой модели классификации, которая создается, настраивается и верифицируется на обучающих, проверочных и тестовых наборах данных [2]. Методы пространственной классификации как расширения методов классификации общего назначения позволяют учитывать атрибуты как текущего, так и смежных объектов, а также их пространственные отношения [7]. Дистанционное зондирование стало одной из областей, в которых широко используются методы классификации пикселей изображения по маркированным категориям (см., например, [8]).

Пространственные регрессионные или прогнозные модели образуют особую группу анализа. К таким методам относятся, например, пространственные авторегрессионные модели (SAR) [2], однако методы, подобные SAR, предполагают манипулирование матрицей пространственных весов большой размерности, что требует больших вычислительных затрат. Поэтому более поздние исследования были направлены на разработку подходов к поиску приближенных решений для SAR в случаях обработки очень больших наборов данных [9].

Обстоятельством, объединившим методы классификации и регрессии, стало то, что неизвестное целевое значение некоторого объекта прогнозируется с учетом набора обучающих примеров [5]. Если целевая переменная дискретна, задача обучения называется классификацией. Если она непрерывна, она называется регрессией. Хорошо известен метод k ближайших соседей, который может быть применен как к задачам классификации, так и к задачам регрессии. Также применяются методы на основе деревьев пространственных моделей, географически взвешенная регрессия и кригинг.

Алгоритм k ближайших соседей (k -Nearest Neighbors, kNN) — метод обучения, который классифицирует новые объекты в соответствии с целевым значением из k наиболее похожих обучающих примеров [5]. Предполагается, что объекты со схожими характеристиками относятся к одному классу. В случае классификации текущему объекту присваивается значение, наиболее часто встречающееся среди схожих по свойствам объектов. В случае регрессии вычисляется (взвешенное) среднее значение. Для определения сходства между двумя объектами kNN требует измерение расстояния для каждого атрибута. Соответственно, географические координаты могут напрямую включаться в алгоритм определения расстояния между двумя местоположениями.

Деревья моделей (Model Trees) работают аналогично деревьям решений, но имеют листья, которые связаны не с фиксированными значениями, а с моделями (линейными функциями) [5]. При этом внутренние узлы дерева разделяют пространство выборок, а внешние узлы представляют конкретные модели для каждой части пространства выборок.

Географически взвешенная регрессия (Geographically Weighted Regression, GWR) — метод локальной регрессии, который использует преимущества положительной автокорреляции между соседними точками в пространстве [5]. Он расширяет традиционную структуру регрессии так, что все параметры оцениваются в локальном контексте. Модель GWR предполагает, что все параметры пространственно согласованы. Таким образом, параметры в текущем местоположении оцениваются на основе измерений, близких к этому местоположению. Это достигается введением диагональной весовой матрицы, которая определяет влияние каждого измерения на оценку параметров регрессии в текущей точке.

Кригинг (Kriging) — распространенный метод линейной интерполяции для оценки неизвестных значений в географических данных типа поля значений [5]. Для набора измерений кригинг оценивает неизвестные значения как взвешенную сумму данных соседних выборок и использует вариограмму для определения оптимальных значений неизвестных весов выборок (рис. 3). Вариограммы моделируют пространственную зависимость между местоположениями и служат функцией расстояния для любой пары объектов.

Поскольку географические объекты в общем случае изменяются с течением времени, важной проблемой становится извлечение знаний на **основе пространственно-временных данных**. При этом некоторые взаимосвязи с окружением (как в пространственном, так и в непространственном отношении) оста-

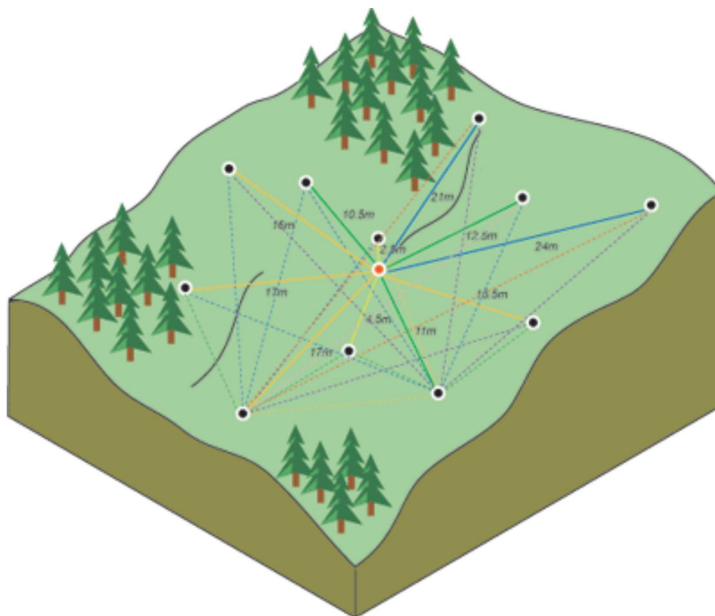


Рис. 3 Создание пар одной точки (красная точка) со всеми другими измеряемыми местоположениями (по [10]). Этот процесс повторяется для каждой измеряемой точки

ются постоянными, другие могут значительно изменяться с течением времени [11].

Важным становится вопрос о способах регистрации временной изменчивости. Так, при анализе движущегося объекта наиболее простым подходом будет проецирование его траектории в сугубо пространственное измерение, однако это приведет к значительной потере информации о динамике. Второй подход может рассматривать последовательность временных срезов, при которой пространственные закономерности обнаруживаются независимо в пределах отдельного временного среза и затем сопоставляются между собой. Оба подхода не оптимальны для широкого использования пространственно-временных структур, в связи с чем актуальна разработка специализированных алгоритмов интеллектуального анализа пространственно-временных геоданных.

4 Современные тенденции применения географического искусственного интеллекта

Географический ИИ стал результатом внедрения в традиционный функционал географических информационных систем ряда технологий ИИ, которые ранее нашли свое применение во многих специфических областях, таких как авто-

номные транспортные средства, распознавание лиц, перевод текста, спутниковая аналитика и многое другое. Актуальность этого подхода обусловлена рядом факторов [1]. Во-первых, это накопление *значительных объемов* весьма разнородных пространственных данных, получаемых со спутников, пилотируемых и беспилотных летательных аппаратов, осуществляющих наблюдения и дистанционное зондирование Земли оптическими, радарными, лидарными и другими способами. Во-вторых, появились *новые типы данных* и новые области их применения, такие как анализ местоположений и траекторий движущихся объектов, пространственных данных социальных сетей, пространственной информации в веб-документах, геокодированных мультимедиа и т. п. В-третьих, возросла роль слияния (fusion) данных [12–14] — концепции, заключающейся в объединении данных из различных источников в предположении, что это даст более точные, полные и достоверные результаты и позволит избежать проблем, связанных с недостаточным качеством или с односторонней репрезентативностью отдельных источников данных (см., например, [15]). Все это значительно расширило применение методов GeoAI для интеллектуального анализа пространственных данных и обнаружения новых типов закономерностей.

К важным задачам, решаемым посредством GeoAI, относятся: классификация изображений, обнаружение объектов (рис. 4), сегментация сцен, интерполяция, моделирование (simulation), прогнозирование (prediction), поиск и ответы на вопросы на основе естественного языка, оперативная интеграция данных, геообогащение и многие другие [1]. Существует широкий спектр решений GeoAI для конкретных задач.



Рис. 4 Результаты автоматического дешифрирования объектов гидрографии на спутниковом изображении (из [16])

Примеры исследовательских достижений в сфере GeoAI, характеризующих современное состояние дел, включают в числе прочих совершенствование методов быстрого и точного обнаружения объектов [1], обнаружения особенностей местности [17], мест плотного расположения строений [18], семантическую классификацию (например, лидарных облаков точек) [19]. Машинное обучение и обработка естественного языка облегчают извлечение географической информации из неструктурированных (текстовых) данных, а также сопоставление описаний природных объектов в альтернативных справочниках [20]. Технологии Semantic Web, онтологии и связанные данные используются для улучшения поиска географической информации и построения расширенных графов географических знаний [21–24]. Эти примеры показывают, что методы ИИ проникают во многие аспекты и виды деятельности и не ограничиваются конкретными задачами или типами данных.

Исследования GeoAI также способствуют разработке вопросно-ответных систем и, в более широком смысле, «умных» (интеллектуальных) цифровых помощников [1]. Доступность данных с установленными пространственными отношениями даст таким системам возможность эффективно решать ряд задач, таких как определение местоположения пользователя, расстояний до других объектов, рассуждения о топологических связях, понимание неопределенных когнитивных областей и т. п.

Методы машинного обучения и ИИ также играют важную роль в области «социального зондирования» [25], использующего создаваемый пользователями цифровой контент для решения ряда задач, от выявления моделей мобильности людей и изучения структуры социальных сетей до решений в области городского планирования [1, 26]. Процесс социального зондирования предполагает создание многомерных семантических сигнатур данных (пространственных, временных и тематических характеристик), извлекаемых из «цифрового следа» людей [25]. Такой «цифровой след» в основном регистрируется с помощью мобильных устройств и других устройств, оснащенных соответствующими сенсорами. Также важным источником таких данных служат датчики движения, компьютерное зрение и методы глубокого обучения изображения на уровне улиц. Получаемые данные включают в себя информацию о местоположении человека, температуру окружающей среды, освещенность, уровень шума и другие характеристики. Огромный объем данных, собранных с помощью этих устройств, а также неоднородность получаемого контента делают эти данные особенно подходящими для анализа с помощью методов GeoAI.

Для реализации исследований в области GeoAI требуются высококачественные наборы геопространственных данных, поскольку многие модели ИИ, особенно глубокие нейронные сети, необходимо обучать на большом наборе хорошо размеченных обучающих данных [1], а с точки зрения воспроизводимости и тиражируемости разрабатываемых моделей GeoAI необходима возможность связывания модели с набором исходных данных, на основе которых была разработана модель, в связи с чем обучающие данные, по сути, становятся неотъ-

емлемой частью моделей GeoAI. Аналогичным образом необходимо совместно пользоваться исходным кодом используемой архитектуры, поскольку на производительность глубоких нейронных сетей часто влияют детали реализации, такие как стратегии случайного заполнения и инициализации параметров.

5 Заключение

Резюмируя, следует отметить, что поиск знаний в базах географических данных — нетривиальный процесс, который невозможно решить с помощью прямого переноса традиционных методов интеллектуального анализа данных. Напротив, он требует понимания фундаментальных географических концепций, учета особенностей пространственного или пространственно-временного представления геообъектов и сопряженных с этим специализированных алгоритмов.

Географический ИИ представляется перспективной технологией интеллектуальной обработки пространственных данных для решения различных задач, таких как классификация изображений, обнаружение объектов, сегментация сцен, моделирование (*simulation*) и интерполяция, прогнозирование (*link prediction*), поиск и ответы на вопросы на основе естественного языка, оперативная интеграция данных, геообогащение и др., особенно в контексте анализа пространственно-временных закономерностей.

Результатом данной работы стал анализ основных пространственных концепций представления географических данных и ряда современных методов анализа и извлечения географических знаний из баз географических данных, включая современные тенденции развития GeoAI. Это должно послужить отправной точкой при разработке новых подходов к извлечению географического знания. Областью применения результата может стать проектирование географических информационных систем, предназначенных для извлечения и использования геознания.

Литература

1. Janowicz K., Gao S., McKenzie G., Hu Y., Bhaduri B. GeoAI: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond // *Int. J. Geogr. Inf. Sci.*, 2019. Vol. 34. Iss. 4. P. 625–636. doi: 10.1080/13658816.2019.1684500.
2. Mennis J., Guo D. Spatial data mining and geographic knowledge discovery — an introduction // *Comput. Environ. Urban.* 2009. Vol. 33. Iss. 6. P. 403–408. doi: 10.1016/j.compenvurbsys.2009.11.001.
3. Miller H., Han J. Geographic data mining and knowledge discovery: An overview // *Geographic data mining and knowledge discovery* / Eds. H. Miller, J. Han. — CRC Press, Taylor and Francis Group, 2009. P. 1–26.
4. Никишин Д. А. Направления развития методологической базы для работы с гео-данными в перспективных геоинформационных системах // *Системы и средства*

- информатики, 2023. Т. 33. № 2. С. 34–45. doi: 10.14357/08696527230204. EDN: YRDRJY.
5. *Rinzivillo S., Turini F., Bogorny V., Kórner C., Kuijpers B., May M.* Knowledge discovery from geographical data // *Mobility, data mining and privacy* / Eds. F. Giannotti, D. Pedreschi. — Berlin, Heidelberg: Springer, 2008. P. 243–265. doi: 10.1007/978-3-540-75177-9_10.
 6. *Ester M., Sander J., Kriegel H.-P., Xu X.* A density-based algorithm for discovering clusters in large spatial databases with noise // *2nd Conference (International) on Knowledge Discovery and Data Mining Proceedings*. — AAAI Press, 1996. P. 226–231. doi: 10.5555/3001460.3001507.
 7. *Ester M., Kriegel H.-P., Sander J.* Spatial data mining: A database approach // *Advances in spatial databases* / Eds. M. Scholl, A. Voisard. — Lecture notes in computer science ser. — Berlin: Springer-Verlag, 1997. Vol. 1262. P. 47–66. doi: 10.1007/3-540-63238-7_24.
 8. *Cleve C., Kelly M., Kearns F. R., Moritz M.* Classification of the wildland–urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography // *Comput. Environ. Urban*, 2008. Vol. 32. Iss. 4. P. 317–326. doi: 10.1016/j.compenvurbsys.2007.10.001.
 9. *Griffith D.* Faster maximum likelihood estimation of very large spatial autoregressive models: An extension of the Smirnov–Anselin result // *J. Stat. Comput. Sim.*, 2004. Vol. 74. Iss. 12. P. 855–866. doi: 10.1080/00949650410001650126.
 10. Как работает инструмент Кригинг. — ESRI. Программное обеспечение ArcMap 10.8: Справочник по инструментам, 2021. <https://desktop.arcgis.com/ru/arcmap/latest/tools/3d-analyst-toolbox/how-kriging-works.htm>.
 11. Нукишин Д. А. Обзор подходов к пространственно-временному моделированию и выявление основных тенденций развития T-GIS // *Системы и средства информатики*, 2022. Т. 32. № 3. С. 50–62. doi: 10.14357/08696527220305. EDN: TOWSDH.
 12. *Janowicz K., van Harmelen F., Hendler J. A., Hitzler P.* Why the data train needs semantic rails // *AI Mag.*, 2015. Vol. 36. Iss. 1. P. 5–14. doi: 10.1609/aimag.v36i1.2560.
 13. Дулин С. К., Дулина Н. Г., Кожунова О. С. Синтез геоданных в пространственных инфраструктурах на основе связанных данных // *Информатика и её применения*, 2019. Т. 13. Вып. 1. С. 82–90. doi: 10.14357/19922264190112. EDN: ZASZIT.
 14. Дулин С. К., Розенберг И. Н., Уманский В. И. О проблеме интеграции информационных ресурсов // *Системы и средства информатики*, 2019. Т. 29. № 3. С. 127–138. doi: 10.14357/08696527190311. EDN: PGJIFL.
 15. *Gao S., Janowicz K., Montello D. R., Hu Y., Yang J.-A., McKenzie G., Ju Y., Gong L., Adams B., Yan B.* A data-synthesis-driven method for detecting and extracting vague cognitive regions // *Int. J. Geogr. Inf. Sci.*, 2017. Vol. 31. Iss. 6. P. 1245–1271. doi: 10.1080/13658816.2016.1273357.
 16. Искусственный интеллект в ГИС. — КБ «Панорама». <https://www.gistoolkit.com/download/prezents/gisartintelligence.pdf>.
 17. *Li W., Hsu C.-Y.* Automated terrain feature identification from remote sensing imagery: A deep learning approach // *Int. J. Geogr. Inf. Sci.*, 2020. Vol. 34. Iss. 4. P. 637–660. doi: 10.1080/13658816.2018.1542697.

18. Xie Y., Cai J., Bhojwani R., Shekhar S., Knight J. A locally-constrained YOLO framework for detecting small and densely-distributed building footprints // *Int. J. Geogr. Inf. Sci.*, 2020. Vol. 34. Iss. 4. P. 777–801. doi: 10.1080/13658816.2019.1624761.
19. Guo Z., Feng C.-C. Using multi-scale and hierarchical deep convolutional features for 3D semantic classification of TLS point clouds // *Int. J. Geogr. Inf. Sci.*, 2020. Vol. 34. Iss. 4. P. 661–680. doi: 10.1080/13658816.2018.1552790.
20. Acheson E., Volpi M., Purves R. S. Machine learning for cross-gazetteer matching of natural features // *Int. J. Geogr. Inf. Sci.*, 2020. Vol. 34. Iss. 4. P. 708–734. doi: 10.1080/13658816.2019.1599123.
21. Ballatore A., Bertolotto M., Wilson D. C. Geographic knowledge extraction and semantic similarity in OpenStreetMap // *Knowl. Inf. Syst.*, 2013. Vol. 37. Iss. 1. P. 61–81. doi: 10.1007/s10115-012-0571-0.
22. Regalia B., McKenzie G., Gao S., Janowicz K. Crowdsensing smart ambient environments and services // *T. GIS*, 2016. Vol. 20. Iss. 3. P. 382–398. doi: 10.1111/tgis.12233.
23. Mai G., Janowicz K., Yan B., Scheider S. Deeply integrating linked data with geographic information systems // *T. GIS*, 2019. Vol. 23. Iss. 3. P. 579–600. doi: 10.1111/tgis.12538.
24. Yan B., Janowicz K., Mai G., Zhu R. A spatially explicit reinforcement learning model for geographic knowledge graph summarization // *T. GIS*, 2019. Vol. 23. Iss. 3. P. 620–640. doi: 10.1111/tgis.12547.
25. Janowicz K., McKenzie G., Hu Y., Zhu R., Gao S. Chapter 3: Using semantic signatures for social sensing in urban environments // *Mobility patterns, big data and transport analytics: Tools and applications for modeling* / Eds. C. Antoniou, L. Dimitriou, F. Pereira. — Elsevier, 2019. P. 31–54. doi: 10.1016/B978-0-12-812970-8.00003-8.
26. Resch B., Summa A., Sagl G., Zeile P., Exner J.-P. Urban emotions — geo-semantic emotion extraction from technical sensors, human sensors and crowdsourced data // *Progress in location-based services* / Eds. G. Gartner and H. Huang. — Springer, 2015. P. 199–212. doi: 10.1007/978-3-319-11879-6_14.

Поступила в редакцию 02.09.2025

Принята к публикации 15.10.2025

MODERN METHODS OF EXTRACTING GEOGRAPHICAL KNOWLEDGE IN LARGE ARRAYS OF SPATIAL DATA

D. A. Nikishin

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The main spatial concepts of spatial data representation and a number of modern methods for analyzing and extracting geographical knowledge from geographic databases including current trends in the development of geographic artificial intelligence (AI) are presented. Searching for knowledge in geographical

databases is a nontrivial process that requires understanding of fundamental geographical concepts taking into account the features of spatial or spatiotemporal representation of geobjects and related specialized algorithms and which is difficult to implement using direct transfer of traditional methods of data mining. Geographical AI is a promising technology for intelligent spatial data processing for solving various tasks (such as clustering, classification, segmentation, interpolation, etc.), especially in the context of the analysis of spatial and temporal patterns. The results obtained should serve as a starting point for the development of new approaches to the extraction of geographical knowledge.

Keywords: geodata; geographical databases; methods of extracting geographical knowledge; geographical artificial intelligence

DOI: 10.14357/08696527250409

EDN: CWPFFNM

References

1. Janowicz, K., S. Gao, G. McKenzie, Y. Hu, and B. Bhaduri. 2019. GeoAI: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *Int. J. Geogr. Inf. Sci.* 34(4):625–636. doi: 10.1080/13658816.2019.1684500.
2. Mennis, J., and D. Guo. 2009. Spatial data mining and geographic knowledge discovery — an introduction. *Comput. Environ. Urban* 33(6):403–408. doi: 10.1016/j.compenvurbsys.2009.11.001.
3. Miller, H., and J. Han. 2009. Geographic data mining and knowledge discovery: An overview. *Geographic data mining and knowledge discovery*. Eds. H. Miller and J. Han. CRC Press, Taylor and Francis Group. 1–26.
4. Nikishin, D. A. 2023. Napravleniya razvitiya metodologicheskoy bazy dlya raboty s geodannymi v perspektivnykh geoinformatsionnykh sistemakh [Directions of development of the methodological base for working with geodata in promising geoinformation systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(2):34–45. doi: 10.14357/08696527230204. EDN: YRDRJY.
5. Rinzivillo, S., F. Turini, V. Bogorny, C. Körner, B. Kuijpers, and M. May. 2008. Knowledge discovery from geographical data. *Mobility, data mining and privacy*. Eds. F. Giannotti and D. Pedreschi. Berlin, Heidelberg: Springer. 243–265. doi: 10.1007/978-3-540-75177-9_10.
6. Ester, M., J. Sander, H.-P. Kriegel, and X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *2nd Conference (International) on Knowledge Discovery and Data Mining Proceedings*. AAAI Press. 226–231. doi: 10.5555/3001460.3001507.
7. Ester, M., H.-P. Kriegel, and J. Sander. 1997. Spatial data mining: A database approach. *Advances in spatial databases*. Eds. M. Scholl and A. Voisard. Lecture notes in computer science ser. Berlin: Springer-Verlag. 1262:47–66. doi: 10.1007/3-540-63238-7_24.
8. Cleve, C., M. Kelly, F. R. Kearns, and M. Moritz. 2008. Classification of the wildland–urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography. *Comput. Environ. Urban* 32(4):317–326. doi: 10.1016/j.compenvurbsys.2007.10.001.

9. Griffith, D. 2004. Faster maximum likelihood estimation of very large spatial autoregressive models: An extension of the Smirnov–Anselin result. *J. Stat. Comput. Sim.* 74(12):855–866. doi: 10.1080/00949650410001650126.
10. How Kriging works. 2021. ESRI. ArcMap 10.8: Tool reference. Available at: <https://desktop.arcgis.com/en/arcmap/latest/tools/3d-analyst-toolbox/how-kriging-works.htm> (accessed November 18, 2025).
11. Nikishin, D. A. 2022. Obzor podkhodov k prostranstvenno-vremennomu modelirovaniyu i vyyavlenie osnovnykh tendentsiy razvitiya T-GIS [Overview of approaches to space–time modeling and the main trends in the development of T-GIS]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(3):50–62. doi: 10.14357/08696527220305. EDN: TOWSDH.
12. Janowicz, K., F. van Harmelen, J. A. Hendler, and P. Hitzler. 2015. Why the data train needs semantic rails. *AI Mag.* 36(1):5–14. doi: 10.1609/aimag.v36i1.2560.
13. Dulin, S. K., N. G. Dulina, and O. S. Kozhunova. 2019. Sintez geodannykh v prostranstvennykh infrastrukturakh na osnove svyazannykh dannykh [Synthesis of geodata in spatial infrastructures based on related data]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):82–90. doi: 10.14357/19922264190112. EDN: ZASZIT.
14. Dulin, S. K., I. N. Rozenberg, and V. I. Umanskiy. 2019. O probleme integratsii informatsionnykh resursov [About the problem of information resources integration]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(3):127–138. doi: 10.14357/08696527190311. EDN: PGJIFL.
15. Gao, S., K. Janowicz, D. R. Montello, Y. Hu, J.-A. Yang, G. McKenzie, Y. Ju, L. Gong, B. Adams, and B. Yan. 2017. A data-synthesis-driven method for detecting and extracting vague cognitive regions. *Int. J. Geogr. Inf. Sci.* 31(6):1245–1271. doi: 10.1080/13658816.2016.1273357.
16. Iskusstvennyy intellekt v GIS [Artificial Intelligence in GIS]. KB «Panorama» [KB Panorama]. Available at: <https://www.gistoolkit.com/download/prezents/gisartintelligence.pdf> (accessed November 18, 2025).
17. Li, W., and C.-Y. Hsu. 2020. Automated terrain feature identification from remote sensing imagery: A deep learning approach. *Int. J. Geogr. Inf. Sci.* 34(4):637–660. doi: 10.1080/13658816.2018.1542697.
18. Xie, Y., J. Cai, R. Bhojwani, S. Shekhar, and J. Knight. 2020. A locally-constrained YOLO framework for detecting small and densely-distributed building footprints. *Int. J. Geogr. Inf. Sci.* 34(4):777–801. doi: 10.1080/13658816.2019.1624761.
19. Guo, Z., and C.-C. Feng. 2020. Using multi-scale and hierarchical deep convolutional features for 3D semantic classification of TLS point clouds. *Int. J. Geogr. Inf. Sci.* 34(4):661–680. doi: 10.1080/13658816.2018.1552790.
20. Acheson, E., M. Volpi, and R. S. Purves. 2020. Machine learning for cross-gazetteer matching of natural features. *Int. J. Geogr. Inf. Sci.* 34(4):708–734. doi: 10.1080/13658816.2019.1599123.
21. Ballatore, A., M. Bertolotto, and D. C. Wilson. 2013. Geographic knowledge extraction and semantic similarity in OpenStreetMap. *Knowl. Inf. Syst.* 37(1):61–81. doi: 10.1007/s10115-012-0571-0.
22. Regalia, B., G. McKenzie, S. Gao, and K. Janowicz. 2016. Crowdsensing smart ambient environments and services. *T. GIS* 20(3):382–398. doi: 10.1111/tgis.12233.

23. Mai, G., K. Janowicz, B. Yan, and S. Scheider. 2019. Deeply integrating linked data with geographic information systems. *T. GIS* 23(3):579–600. doi: 10.1111/tgis.12538.
24. Yan, B., K. Janowicz, G. Mai, and R. Zhu. 2019. A spatially explicit reinforcement learning model for geographic knowledge graph summarization. *T. GIS* 23(3):620–640. doi: 10.1111/tgis.12547.
25. Janowicz, K., G. McKenzie, Y. Hu, R. Zhu, and S. Gao. 2019. Chapter 3: Using semantic signatures for social sensing in urban environments. *Mobility patterns, big data and transport analytics: Tools and applications for modeling*. Eds. C. Antoniou, L. Dimitriou, and F. Pereira. Elsevier. 31–54. doi: 10.1016/B978-0-12-812970-8.00003-8.
26. Resch, B., A. Summa, G. Sagl, P. Zeile, and J.-P. Exner. 2015. Urban emotions — geo-semantic emotion extraction from technical sensors, human sensors and crowdsourced data. *Progress in location-based services*. Eds. G. Gartner and H. Huang. Springer. 199–212. doi: 10.1007/978-3-319-11879-6_14.

Received September 2, 2025

Accepted October 15, 2025

Contributor

Nikishin Dmitry A. (b. 1976) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; dmnikishin@mail.ru

ПОДХОД К ИСПОЛЬЗОВАНИЮ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ ДЛЯ ФОРМИРОВАНИЯ ГИПОТЕЗ В КОНКРЕТНО-ИСТОРИЧЕСКОМ ИССЛЕДОВАНИИ

И. М. Адамович¹, О. И. Волков²

Аннотация: Статья посвящена развитию средств поддержки конкретно-исторических исследований. Предлагаемый подход позволяет определить приоритетные направления информационного поиска посредством автоматического формирования обоснованных гипотез. Особенность подхода заключается в его ориентированности не на изолированные исторические факты, а на динамику развития конкретно-исторических процессов. Подход опирается на методы кластерного анализа временных рядов. Описаны особенности временных рядов, отражающих доступную в исторических источниках информацию об изменении тех или иных объектов и явлений в прошлом. Определены алгоритмы, наиболее соответствующие данным особенностям. Особое внимание уделено определению специфической меры сходства для таких временных рядов. Подробно рассмотрен пример данных, отражающий особенности старообрядческого предпринимательства в Российской Империи, и показано, как с помощью описанного подхода можно сделать обоснованное предположение об отнесении объекта исследования к старообрядчеству.

Ключевые слова: конкретно-историческое исследование; формирование гипотез; кластеризация; временные ряды; мера сходства

DOI: 10.14357/08696527250410

EDN: EWERER

1 Введение

Как показано в [1], при проведении конкретно-исторического исследования выбор направления поиска недостающей информации остается полностью ручной неформальной процедурой, зависящей исключительно от опыта и квалификации исследователя. Неверный выбор направления поиска неизбежно приводит к значительному росту его объемов и к заметному увеличению сроков всего исследования. Поэтому механизмы, позволяющие предложить пользователю перспективные направления поиска, крайне востребованы. В [1] описан подход к формированию гипотез, основанный на процедуре импутации данных, т. е. автоматического формирования на базе неполной индукции гипотезы о значении лакуны в наборе исторических фактов. Ценность такого решения

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Adam@amsd.com

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Volkov@amsd.com

для конкретно-исторического исследования, как показано в [1], «заключается в важности обоснованной гипотезы как метода развития любого научного знания. Особенно это относится к историческому знанию, так как любое знание о прошлом несет в себе долю гипотетичности, поскольку историк не может наблюдать непосредственно исторические события как главный объект своего изучения [2]. Задачей, решаемой с помощью этого подхода, является не получение новой достоверной информации об объекте исследования, поскольку неполная индукция позволяет делать только вероятностные заключения, а создание обоснованной гипотезы, которая задаст новое направление поиска». Недостаток этого подхода — опора на сравнение статических данных без учета их динамики.

В конкретно-исторических исследованиях, сфокусированных на изучении конкретных событий, явлений или процессов в их исторической специфике, анализ временных рядов широко используется в следующих направлениях:

- экономическая история [3];
- демографическая история [4];
- социальная история [5];
- история климата и экологии [6].

Целью настоящей статьи ставится обоснование применения анализа временных рядов в задачах автоматизации формирования гипотез при проведении конкретно-исторических исследований и описание соответствующего подхода.

2 Описание и обоснование подхода

При проведении конкретно-исторических исследований встречаются ситуации, когда динамика некоторых процессов, связанных с историческими сущностями, обладающими неким общим свойством, имеет выраженную специфику. В таких случаях, если для определенной сущности динамика связанного процесса обладает той же спецификой, можно предположить, что эта сущность также обладает этим же свойством. Это предположение и будет искомой гипотезой.

Например, старообрядческое предпринимательство в России имело ряд специфических черт, отличавших его от практик нестарообрядцев. Эта специфика определялась следующими особенностями старообрядческого предпринимательства [7]:

- общинность, круговая порука и коллективная ответственность;
- отказ от демонстративной роскоши и приоритет вложений средств в развитие дела;
- строгая трудовая этика;
- сочетание традиционализма с технологическими новациями.

При этом старообрядческое предпринимательство нельзя считать однородным. Специфика старообрядческого бизнеса зависела от географии, конфессионального течения и стартовых условий. На этом примере можно проиллюстрировать предлагаемый подход.

В качестве сравниваемых процессов рассмотрим экономическую деятельность предпринимателей в Российской Империи. Эти процессы представлены набором зафиксированных в исторических документах (личные архивы предпринимателей, архивы предприятий, историко-статистические сведения, исторические акты и уложения, отчеты и т. д.) показателей. Каждый такой набор представляет собой неполный аperiодический временной ряд, поскольку исторические данные обладают свойством фрагментарности. Более того, первая точка временного ряда может не совпадать с началом процесса, а последняя может не совпадать с моментом завершения экономической деятельности, т. е. с концом процесса. Представляет интерес совокупность данных S , для которых известно, что субъект, осуществлявший деятельность, был старообрядцем.

Поскольку старообрядческое предпринимательство неоднородно, всю совокупность следует разбить на однотипные группы (кластеры) K_i : $\cup K_i = S_i$; $\forall i, j$ $K_i \cap K_j = \emptyset$. Группы, имеющие значительное число элементов, будем называть представительными и далее рассматривать только их, поскольку отнесение процесса, для субъекта которого делается предположение о принадлежности к старообрядчеству, к той или иной группе однотипных процессов из совокупности S (отнесение соответствующего временного ряда к тому или иному кластеру) должно подтверждаться статистически.

Для этого должны быть определены:

- критерий значительности для представительных кластеров;
- алгоритм кластеризации;
- способ сравнения процессов, представленных временными рядами (мера сходства).

Наиболее естественным критерием значительности для представительных кластеров выглядит превышение отношения объема кластера к общему объему совокупности данных некоторого порога: $|K_i|/|S| = \alpha$.

Среди множества алгоритмов кластеризации наиболее популярными [8] остаются методы К-средних (k-Means), агломеративной иерархической кластеризации, а также алгоритм DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Метод k-Means неспособен выделить кластеры несферической формы, и число кластеров должно быть задано как параметр изначально, что делает этот алгоритм непригодным для решения рассматриваемой задачи. Этих недостатков лишены алгоритмы агломеративной иерархической кластеризации и DBSCAN, что позволяет использовать их для кластеризации временных рядов, характеризующих экономическую деятельность предпринимателей-старообрядцев. Для этого следует определить меру близости для этих экономических процессов, представленных в виде временных рядов.

3 Специфика задачи

Для выбора наиболее адекватной для описанной задачи меры сходства следует учесть следующие особенности рассматриваемых процессов.

- 3.1. Сходные процессы могут начинаться в разные моменты времени. Следовательно, мера сходства не должна быть чувствительной к сдвигу любого из сравниваемых процессов по оси времени.
- 3.2. Динамика развития процесса на каждом этапе служит важным показателем экономической деятельности. Следовательно, мера сходства должна быть чувствительна к сжатию или растяжению отдельных участков процесса по оси времени.
- 3.3. Временные ряды, описывающие процесс, неравномерны. Следовательно, мера сходства должна учитывать несовпадение временных меток сравниваемых временных рядов.
- 3.4. Начало и конец процесса могут не совпадать с началом и концом описывающего его временного ряда. Поэтому мера сходства должна учитывать тот факт, что временной ряд может отражать не весь процесс, а только некоторую его часть, соответствующую непрерывному интервалу наблюдений в пределах всего времени существования процесса, и сходные процессы могут быть представлены несходными временными рядами, поскольку относятся к разным интервалам наблюдений.
- 3.5. Вся зафиксированная в исторических документах экономическая информация отражает реальные изменения в ходе процесса и считается значимой. Следовательно, допустимо использование меры сходства, чувствительной к выбросам и шумам.

4 Метрика для временных рядов

Для определения меры сходства процессов, представленных временными рядами, следует опереться на некоторый способ сравнения этих рядов. Выбор оптимальной метрики из множества созданных для временных рядов определяется спецификой задачи, а именно: особенностями 3.2 (чувствительность к искажениям временной шкалы) и 3.3 (неравномерность временных рядов). Эти особенности делают невозможным использование многих метрик, основанных на сравнении формы временных рядов, в том числе и очень популярное DTW-расстояние, основанное на алгоритме динамической трансформации временной шкалы (Dynamic Time Warping). Поэтому следует опереться на широко используемое в кластерном анализе и многих других приложениях для описания различий между объектами евклидово расстояние [9]. В случае с двумя временными рядами X и Y , имеющими одинаковый интервал из n наблюдений, оно вычисляется следующим образом:

$$m(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}. \quad (1)$$

Евклидова метрика подвержена влиянию выбросов и шумов, что в соответствии с особенностью 3.5 (допустимость чувствительности к выбросам и шумам) не считается существенным недостатком. Проблема заключается в требовании одинаковой длины сравниваемых рядов и совпадении их временных меток, в то время как в рассматриваемой задаче временные ряды нерегулярны и могут иметь разную длину. Наиболее распространенная процедура при обработке нерегулярных временных рядов — предварительная обработка данных с помощью линейной интерполяции [10]. Этот метод имеет существенный недостаток, поскольку вводит дополнительную информацию во временные ряды с гораздо большей неопределенностью, чем реальные наблюдения. Но, как показано выше, целью ставится не получение новой достоверной информации об объекте исследования, а только создание обоснованной гипотезы, которая задаст новое направление поиска, что позволяет использовать данный подход, несмотря на высокий уровень неопределенности в данных. В результате сравнение неравномерных временных рядов X и Y , определенных на одном временном интервале T , будет осуществляться в три этапа:

- (1) построение непрерывных на интервале T функций $x(t)$ и $y(t)$ линейной интерполяцией временных рядов X и Y соответственно;
- (2) создание равномерной последовательности $\{t_i\}$ временных меток на интервале T и построение новых временных рядов $\tilde{X} = \{x(t_i)\}$ и $\tilde{Y} = \{y(t_i)\}$;
- (3) вычисление евклидова расстояния между рядами $\tilde{X}$ и $\tilde{Y}$ в соответствии с формулой (1).

Евклидова метрика чувствительна к сдвигам по времени, что входит в противоречие с особенностью 3.1. Это противоречие будет устранено далее при определении меры сходства процессов.

5 Мера сходства процессов

При определении меры сходства рассматриваемых в данной статье процессов с опорой на выбранную выше метрику следует учесть, что:

- мера сходства не должна быть чувствительной к сдвигу любого из сравниваемых процессов по оси времени (особенность 3.1);
- мера сходства должна быть определена с учетом того, что временной ряд может отражать не весь процесс, а только некоторую его часть (особенность 3.4).

Определим метрику M , отражающую степень сходства процессов S_i и S_j , которым соответствуют временные ряды $X_i(t)$ и $X_j(t)$ с интервалами наблюдений $T(X_i)$ и $T(X_j)$ соответственно. Пусть m_T — выбранная выше метрика для временных рядов, определенных на одном временном интервале T . Тогда определим метрику $\tilde{m}$ для временных рядов с пересекающимися периодами наблюдений как

$$\tilde{m}(X, Y) = \left(\frac{|T(X) \cup T(Y)|}{|T(X) \cap T(Y)|} \right)^\beta m_{T(X) \cap T(Y)}(X, Y),$$

т. е. метрика m , определенная на пересечении интервалов наблюдений временных рядов X и Y с поправочным коэффициентом, представляющим собой штраф за несовпадение интервалов наблюдений. Константа β определяет значимость этого штрафа. Метрика $\tilde{m}$ учитывает особенность 3.4.

Определим для временного ряда $X = \{x_{t_1}, \dots, x_{t_n}\}$, определенного на интервале $T(X) = [t_1, t_n]$, ряд X^Δ , представляющий собой ряд X , сдвинутый на величину Δ по оси времени:

$$X^\Delta = \{y_{p_1}, \dots, y_{p_n} | p_i = t_i + \Delta; y_{p_i} = x_{t_i}\};$$

$$T(X^\Delta) = [t_1 + \Delta, t_n + \Delta].$$

Тогда определим метрику M в описанных выше обозначениях следующим образом:

$$M(S_i, S_j) = \min_{\Delta: T(X_i^\Delta) \cap T(X_j) \neq \emptyset} \tilde{m}(X_i^\Delta, X_j).$$

Иными словами, если для пары временных рядов сдвинуть один из них по оси времени на некоторую величину Δ и посчитать расстояние между ними в соответствии с метрикой $\tilde{m}$, то это расстояние можно считать функцией от одной переменной Δ . Минимум этой функции на интервале значений, для которых интервалы наблюдений имеют непустое пересечение, и принимается за значение метрики M . Вычисление минимума функции на некотором интервале — обычная оптимизационная задача нелинейного программирования [11] и может решаться многими методами, например методом сканирования с переменным шагом.

6 Экспериментальная проверка подхода

Для проверки правомерности предложенного подхода были взяты данные из доступных источников. В качестве источника информации о состоянии и динамике промышленных предприятий России были выбраны «Указатель фабрик и заводов Европейской России с Царством Польским и Вел. Кн. Финляндским» за 1881–1894 гг. и «Список фабрик и заводов России» за 1910 г., что позволило сформировать временные ряды экономических показателей отдельных предприятий. Для каждого предприятия в источниках приведены:

- владелец;
- местонахождение;
- число паровых машин;
- количество выработанной продукции в год;

- сумма производства за год в рублях;
- численность рабочих.

Сведения сгруппированы по отраслям и по губерниям.

Поскольку сведения о числе паровых машин и количестве выработанной продукции несопоставимы для разных отраслей, в состав экономических показателей в исследуемых временных рядах были включены:

- сумма производства в рублях;
- численность рабочих.

Поскольку в указанных справочниках не приводится вероисповедание владельцев, в качестве дополнительного источника информации был использован «Список купеческих старообрядческих фамилий Москвы (XIX – начало XX в.)». В результате была сформирована выборка со следующими свойствами:

- (1) рассматриваются только промышленные предприятия Москвы (685 шт.) в связи с отсутствием информации о конфессиональной принадлежности владельцев предприятий других регионов;
- (2) в выборке отсутствуют особо малые, с объемом производства менее 2 тыс. руб. в год, и кустарные предприятия, поскольку информация о таких предприятиях не приведена в используемых справочниках;
- (3) список предпринимателей-старообрядцев содержит всего 178 фамилий, что свидетельствует о неполноте этого списка, поскольку, согласно исследованиям, к началу XX в. старообрядцам принадлежали около 60%–70% промышленных предприятий Москвы, а в некоторых отраслях, например в текстильной, их доля достигала 75%–80%;
- (4) в используемом списке предпринимателей-старообрядцев представлены большей частью «раскольники поповщинского толка» и преимущественно представители текстильной промышленности;
- (5) в списке успешные предприниматели, владеющие крупными предприятиями, представлены более полно, чем предприниматели, владеющие небольшими производствами.

Из приведенных свойства выборки следует невозможность ее использования для получения достоверных исторических выводов, но возможная достаточность для проверки работоспособности алгоритма. Кластеризация данной выборки в соответствии с описанным выше алгоритмом дала следующие результаты:

- были выявлены 6 мелких кластеров, совместно содержащих 21 предприятие, которые явно следует считать в описанных выше терминах неrepresentативными;
- были выявлены также 6 кластеров, представленных в таблице, предприятия в которых явно сгруппированы по объему производства: 1 кластер крупных предприятий, 3 кластера средних предприятий и 2 кластера небольших предприятий.

Представительные кластеры

№	Тип кластера	Среднее число работников	Размер кластера	Число старообрядческих предприятий	Доля старообрядческих предприятий
1	Крупные	811	13	11	85%
2	Средние	185	76	51	67%
3			111	29	26%
4			49	1	2%
5	Небольшие	17	194	37	19%
6			221	46	21%

Единственный кластер крупных предприятий подтверждает тот факт, что крупный бизнес, особенно в текстильной промышленности, большей частью принадлежал старообрядцам, и каких-то дополнительных выводов к этому общеизвестному факту рассматриваемый алгоритм добавить не в состоянии.

Два примерно равных по объему кластера небольших предприятий имеют практически равное представительство старообрядцев — около 20%. Это означает, что на данной выборке в части малого бизнеса алгоритму не удалось выявить специфику старообрядческого предпринимательства. Это может объясняться неполнотой выборки именно в этой области. Известно, что реальное представительство старообрядцев значительно превышало 20%.

Наиболее интересные результаты получены для средних предприятий. Три кластера различны по объему и имеют очень разное представительство старообрядцев:

- (1) большое для кластера 2 — 67%;
- (2) среднее для кластера 3 — 26%;
- (3) малое для кластера 4 — 2%.

Это означает, что алгоритму удалось выявить специфику старообрядческого предпринимательства для средних предприятий и для предпринимателей из кластера 2 предположение об их принадлежности к старообрядчеству, а для предпринимателей из кластера 4 предположение об их непринадлежности к старообрядчеству оказалось вполне обоснованным.

7 Выводы

В статье предложен подход, расширяющий возможности применения анализа временных рядов в конкретно-исторических исследованиях. Предложен новый способ автоматизации формирования гипотез в целях определения перспективных направлений информационного поиска для случаев, когда у исследователя имеется информация о динамике тех или иных процессов, связанных с объектом исследования. Подход проиллюстрирован примером, связанным с динамикой экономических процессов, но также применим и к процессам другой природы.

Приведено описание применения подхода и указаны соответствующие алгоритмы, учитывающие специфику конкретно-исторических данных. Численный эксперимент на реальных исторических данных подтвердил принципиальную применимость подхода для заявленных целей.

Литература

1. *Адамович И. М., Волков О. И.* Механизм формирования гипотез в технологии поддержки конкретно-исторических исследований // Системы и средства информатики, 2023. Т. 33. № 1. С. 135–145. doi: 10.14357/08696527230113. EDN: JFFUKV.
2. *Бочаров А. В.* Алгоритмы использования основных научных методов в конкретно-историческом исследовании. — Томск: ТГУ, 2007. 140 с. EDN: VJKTBT.
3. *Миронов Б. Н.* Хлебные цены в России за два столетия (XVIII–XIX вв.) — Л.: Наука, 1985. 301 с. EDN: VIKLNV.
4. *Сарафанов Д. Е.* Смертность населения в Барнауле в XIX–начале XX вв. (анализ демографических временных рядов) // Информационный бюллетень ассоциации «История и компьютер», 2018. № 47. С. 98–99. EDN: YWKJKP.
5. *Андрианова Е. Г., Головин С. А., Зыков С. В., Лесько С. А., Чукалина Е. Р.* Обзор современных моделей и методов анализа временных рядов динамики процессов в социальных, экономических и социотехнических системах // Российский технологический ж., 2020. Т. 8. № 4. С. 7–45. doi: 10.32362/2500-316X-2020-8-4-7-45. EDN: ZDZRVТ.
6. *Слепцов А. М., Клименко В. В.* Обобщение палеоклиматических данных и реконструкция климата Восточной Европы за последние 2000 лет // История и современность, 2005. № 1. С. 118–135. EDN: PFFIUУ.
7. *Керов В. В.* Хозяйственная культура и хозяйственное поведение старообрядцев в России XVIII–XIX вв.: ценностно-институциональный подход // Вестник Российского университета дружбы народов. Сер. История России, 2011. № 1. С. 40–57. EDN: NCQMEH.
8. *Ковалёв С. П.* Использование алгоритма кластеризации DBSCAN для фильтрации выбросов в данных // Компьютерные системы и сети: 55-я Юбилейная научная конференция аспирантов, магистрантов и студентов. — Минск: БГУИР, 2019. С. 198–200.
9. *Маслицкий С. Э.* Анализ временных рядов с помощью R, 2020. <https://ranalytics.github.io/tsa-with-r>.
10. *Adorf H. M.* Interpolation of irregularly sampled data series — a survey // Astr. Soc. P., 1995. Vol. 77. P. 460–463.???
11. *Прокопенко Н. Ю.* Методы оптимизации. — Нижний Новгород: ННГАСУ, 2018. 118 с.

Поступила в редакцию 17.07.2025

Принята к публикации 15.10.2025

AN APPROACH TO USING TIME SERIES ANALYSIS IN HYPOTHESES FORMING FOR CONCRETE HISTORICAL INVESTIGATION

I. M. Adamovich and O. I. Volkov

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The article is devoted to the development of means of concrete historical investigation supporting. The proposed approach allows one to determine the priority directions of information search by automatically forming the reasonable hypotheses. A special feature of the approach is its focus not on isolated historical facts but on the dynamics of the development of concrete historical processes. The approach is based on the methods of cluster analysis of time series. The article describes the features of time series reflecting information available in historical sources on changes in certain objects and phenomena in the past. The algorithms that best match these features have been identified. Special attention is paid to the definition of a specific similarity measure for such time series. An example of data reflecting the peculiarities of Old Believer entrepreneurship in the Russian Empire is considered in detail and it is shown how, using the described approach, it is possible to make an informed assumption about the assignment of the object of research to Old Believers.

Keywords: concrete historical investigation; hypothesis formation; clustering; time series; similarity measure

DOI: 10.14357/08696527250410

EDN: EWERER

References

1. Adamovich, I. M., and O. I. Volkov. 2023. Mekhanizm formirovaniya gipotez v tekhnologii podderzhki konkretno-istoricheskikh issledovaniy [Hypothesis formation mechanism in the technology of concrete historical investigation support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(1):135–145. doi: 10.14357/08696527230113. EDN: JFFUKV.
2. Bocharov, A. V. 2007. *Algoritmy ispol'zovaniya osnovnykh nauchnykh metodov v konkretno-istoricheskom issledovanii* [Algorithms for using basic scientific methods in concrete historical investigation]. Tomsk: Publishing House of the Tomsk State University. 140 p. EDN: VJKTBT.
3. Mironov, B. N. 1985. *Khlebnye tseny v Rossii za dva stoletiya (XVIII–XIX vv.)* [Bread prices in Russia over two centuries (18th–19th centuries)]. Leningrad: Nauka. 301 p. EDN: VIKLNV.
4. Sarafanov, D. E. 2018. Smernost' naseleniya v Barnaule v XIX–nachale XX vv. (analiz demograficheskikh vremennykh ryadov) [Population mortality in Barnaul in the 19th–early 20th century (analysis of demographic time series)]. *Informatsionnyy*

- byulleten' assotsiatsii «Istoriya i komp'yuter»* [Information Bulletin of Association "History and Computer"] 47:98–99. EDN: YWKJKP.
5. Andrianova, E. G., S. A. Golovin, S. V. Zykov, S. A. Lesko, and E. R. Chukalina. 2020. Obzor sovremennykh modeley i metodov analiza vremennykh ryadov dinamiki protsessov v sotsial'nykh, ekonomicheskikh i sotsiotekhnicheskikh sistemakh [Review of modern models and methods of analysis of time series of dynamics of processes in social, economic and socio-technical systems]. *Russian Technological J.* 8(4):7–45. doi: 10.32362/2500-316X-2020-8-4-7-45. EDN: ZDZRVV.
 6. Sleptsov, A. M., and V. V. Klimenko. 2005. Obobshchenie paleoklimaticheskikh dannykh i rekonstruktsiya klimata Vostochnoy Evropy za poslednie 2000 let [Generalization of the paleoclimate data and reconstruction of the climatic conditions of Eastern Europe during the last 2000 years]. *Istoriya i sovremennost'* [History and Modernity] 1:118–135. EDN: PFFIUW.
 7. Kerov, V. V. 2011. Khozyaystvennaya kul'tura i khozyaystvennoe povedenie staroobryadtsev v Rossii XVIII–XIX vv.: tsennostno-institutsional'nyy podkhod [Economic culture and economic behavior of Russian Old Believers in the 18th–19th centuries: Value and institutional approach]. *Vestnik Rossiyskogo universiteta družby narodov. Ser. Istoriya Rossii* [RUDN Journal of Russian History] 1:40–57. EDN: NCQMEH.
 8. Kovalev, S. P. 2019. Ispol'zovanie algoritma klasterizatsii DBSCAN dlya fil'tratsii vybrosov v dannykh [Using the DBSCAN clustering algorithm for filtering outliers in data]. *Komp'yuternye sistemy i seti: 55-ya Yubileynaya nauchnaya konferentsiya aspirantov, magistrantov i studentov* [55th Anniversary Scientific Conference of Graduate Students, Undergraduates and Students "Computer Systems and Networks" Proceedings]. Minsk: BSUIR. 198–200.
 9. Mastitskiy, S. E. 2020. *Analiz vremennykh ryadov s pomoshch'yu R* [Time series analysis with R]. Available at: <https://ranalytics.github.io/tsa-with-r> (accessed October 17, 2025).
 10. Adorf, H. M. 1995. Interpolation of irregularly sampled data series — a survey. *Astr. Soc. P.* 77:460–463.
 11. Prokopenko, N. Yu. 2018. *Metody optimizatsii* [Optimization methods]. Nizhny Novgorod: NNGASU. 118 p.

Received July 17, 2025

Accepted October 15, 2025

Contributors

Adamovich Igor M. (b. 1934) — Candidate of Science (PhD) in technology, senior scientist, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; adam@amsd.com

Volkov Oleg I. (b. 1964) — leading programmer, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; volkov@amsd.com

СЛОЖНОСТЬ АЛГОРИТМА ПОИСКА СОВПАДЕНИЙ В НЕСКОЛЬКИХ ПОСЛЕДОВАТЕЛЬНОСТЯХ

А. А. Грушо¹, М. И. Забежайло², В. В. Кульченков³, Е. Е. Тимонина⁴

Аннотация: Анализ данных на предмет наличия в них следствий неизвестной, но присутствующей причины возникает во многих задачах, и в работе приведен только один из примеров, связанный с поиском признаков мошенничества среди множества получателей потребительского кредита в банке. Для построения исходных данных был выбран метод, в котором признаки мошенничества проявляются в транзакционной активности после получения кредита, а именно: признаки основываются на том, каким образом выводятся полученные средства. Приведенный пример — частный случай ситуаций, когда в ограниченном множестве прецедентов данных, имеющих большую размерность, присутствуют и повторяются следствия одной причины. В этих условиях большое значение имеет задача нахождения повторения следствий. Построен алгоритм такого поиска, имеющий сложность меньше квадратичной. Сложность построенного алгоритма поиска всех повторений в m упорядоченных последовательностях прецедентов не превосходит mN , где N — длины все прецедентов. С учетом сложности упорядочения каждого прецедента и при исходной упорядоченности всего множества характеристик сложность решения задачи не превосходит $mN \log_2 N$.

Ключевые слова: сложность задачи классификации; машинное обучение; причинно-следственные связи; поиск повторений в наборе последовательностей

DOI: 10.14357/08696527250411

EDN: GAAUSE

1 Введение

Анализ данных на предмет наличия в них следствий неизвестной, но присутствующей причины возникает во многих задачах, и далее приведен только один из примеров. Рассматриваемый класс задач объединяет то, что в ограниченном числе прецедентов данных, имеющих большую размерность, присутствуют и повторяются следствия одной причины [1].

Пример того, как возникают потребности в таких алгоритмах, рассмотрим на задаче, связанной с поиском признаков мошенничества [2–4] среди множества

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, grusho@yandex.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук, m.zabezhailo@yandex.ru

³Банк ВТБ (ПАО), vlad.kulchenkov@gmail.com

⁴Федеральный исследовательский центр «Информатика и управление» Российской академии наук, eltimon@yandex.ru

получателей потребительского кредита в банке. Ряд примеров из других областей можно найти в работах [1, 5–7].

Для исследования был выбран метод, в котором признаки мошенничества проявляются в транзакционной активности после получения кредита, а именно: признаки строятся на том, каким образом выводятся полученные средства.

Метод основан на том, что получатель кредита выводит деньги небольшими суммами в течение ряда дней с помощью системы быстрых платежей. Получателей платежей много, и их отслеживание на первом этапе — это трудоемкий процесс. Поэтому на первом этапе используется последовательность характеристик вида (отправитель платежа, комментариев к платежу). Если наблюдения ведутся несколько дней, то на фоне больших данных различных честных и нечестных отправителей платежей проявляются идентичные комментарии у различных отправителей платежей. Выявление таких повторений служит первичным признаком мошенничества. При этом выявляются потенциальные люди, относительно которых необходимо проводить глубокий анализ более трудоемкими методами. Кроме того, отправитель платежей может в системе быстрых платежей отправлять деньги, не связанные с мошеннической схемой, но целью ставится выявление признаков следствий причины, которая состоит хотя бы в частичном участии отправителя платежей в мошеннической схеме.

Особенность задач такого поиска в том, что большая длина последовательностей, в которых надо организовать поиск (платежи в течение суток). Этот и другие примеры подобных поисковых задач объединяет то, что в ограниченном наборе прецедентов данных, имеющих большую размерность, присутствуют и повторяются следствия одной причины [1]. Точное описание причины появится только после выявления и анализа тех прецедентов, которые обладают общими свойствами на фоне других несущественных для поиска повторений данных. Исследования причинно-следственных связей на фоне шума присутствуют также в ряде других работ [8–11].

В данной работе в рамках упрощенной модели примеров указанных задач при допущении обучения с учителем рассматривается проблема сложности поиска признаков проявления неизвестной причины через повторяющиеся следствия. Простейший алгоритм имеет квадратичную сложность. Построенный в статье алгоритм выявления повторений признаков существования причины в наборе прецедентов имеет сложность $O(N \ln N)$.

2 Модель данных

Пусть данные представляют собой последовательности в алфавите большого размера, элементы которого назовем характеристиками. Рассматривается задача классификации на два класса последовательностей большой длины при наличии ограниченного набора из m обучающих образцов данных одного класса (обучение с учителем). В этих условиях можно использовать причинно-следственные связи [1], основанные на том, что в данных обучающей выборки возникают

детерминированные повторения отдельных характеристик, которые отвечают существованию причины P , порождающей эти повторения как следствия. Кроме повторений, оставшаяся часть данных считается случайной. Альтернативой считаются целиком случайные данные, полученные по равновероятной полиномиальной схеме (второй класс).

В тех случаях когда причина известна и, следовательно, известны ее следствия и описания близких причин, такой метод называется контрфактуальным [12]. Когда причина неизвестна, но она есть, используется обучение на прецедентах, имеющих эту причину [1]. При этом имеется возможность обобщить задачу на случай обучения без учителя.

В работе [1] найдены условия успешного решения поставленной задачи классификации на основе причинно-следственных связей. Пусть M — число характеристик (мощность исходного алфавита), из которых строятся данные, и все характеристики различные. Если часть данных порождена следствиями причины P (сама причина неизвестна), а N — число характеристик в каждом образце обучающей выборки размера m , то с вероятностью, стремящейся к 1, классификатор успешно решает задачу при $m = \text{const}$, $N \rightarrow \infty$, $M \rightarrow \infty$ и $N^m/M^{m-1} \rightarrow 0$. В этих условиях классификатор использует повторения следствий причины P в данных обучающей выборки и данных, которые необходимо классифицировать. Однако следствия причины P могут быть разбросаны в данных в произвольном порядке. Необходимо искать в образцах характеристики, встречающиеся во всех образцах. Такие наборы будем называть повторениями. Простейший алгоритм поиска повторения в обучающей выборке размера m , содержащей по N характеристик в каждом образце, требует $(m-1)N^2$ сравнений элементов выборок. В данной статье рассматривается актуальная задача построения алгоритма поиска повторений меньшей сложности.

3 Описание алгоритма поиска повторений

Предположим, что все M характеристик упорядочены, например перенумерованы натуральными числами. Тогда можно считать, что характеристики — это значения порядковых номеров исходных характеристик. Число характеристик в любом образце равно N . Будем считать, что все характеристики в каждом из m образцов обучающей выборки упорядочены по убыванию. Известно [13, 14], что упорядочение одного образца требует $N \log_2 N$ операций перестановок характеристик. Тогда сложность упорядочения всех образцов равна $mN \log_2 N$.

Пусть характеристики в каждом из m образцов, упорядоченных по убыванию, образуют вектор

$$X^0(j) = (X^0(1, j), X^0(2, j), \dots, X^0(N, j)), \quad j = \overline{1, m},$$

координаты которого пронумерованы по возрастанию от 1 до N слева направо. Здесь через $X^0(i, j)$ обозначены характеристики образцов, где $i = \overline{1, N}$ —

номер места в образце, $j = \overline{1, m}$ — номер образца. Алгоритм поиска всех повторений характеристик в различных образцах состоит из следующих шагов.

Шаг 0. Запишем все векторы $X^0(j)$, $j = \overline{1, m}$, обучающих образцов в прямоугольную таблицу размера $m \times 2N$, начиная с $(N + 1)$ -го столбца, предполагая в дальнейшем, что в результате работы алгоритма строки таблицы будут сдвигаться влево не более чем на N позиций.

Шаг 1. Рассмотрим в построенной таблице $(N + 1)$ -й столбец характеристик $\alpha^0 = (X^0(1, 1), X^0(1, 2), \dots, X^0(1, m))$, который будем называть *окном*. Обозначим через X^0 характеристику, имеющую наибольшее значение в окне α^0 , а Y^0 — характеристику, имеющую наименьшее значение в этом окне α^0 .

Если $X^0 > Y^0$ и $X^0 = X^0(1, k_1) = \dots = X^0(1, k_s)$ для некоторых $1 \leq k_1 < \dots < k_s \leq m$, то сдвигаем строки с номерами $k_1, \dots, k_s$ на одну позицию влево. Получим на том же месте вместо $(N + 1)$ -го столбца α^0 новый столбец (окно) $\alpha^1 = (X^1(1, 1), X^1(1, 2), \dots, X^1(1, m))$, полученный после сдвига строк с одинаковыми X^0 влево на одну позицию. В окне α^1 имеем $X^1(1, j) = X^0(1, j)$ для $j \neq k_1, \dots, k_s$ и $X^1(1, j) = X^0(2, j)$ для $j = k_1, \dots, k_s$.

Если $X^0 = Y^0$, то найдено первое повторение, равное X^0 . Тогда сдвигаем все строки таблицы на одну позицию влево и получим новое окно $\alpha^1 = (X^1(1, 1), X^1(1, 2), \dots, X^1(1, m))$, где $X^1(1, j) = X^0(2, j)$ для всех $j = \overline{1, m}$.

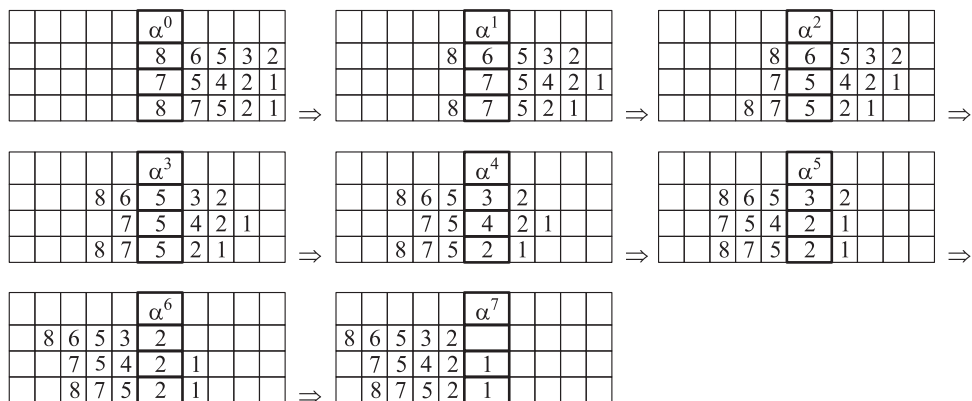
Шаг 2. В новой таблице рассмотрим окно α^1 . Опять обозначим через X^1 характеристики, имеющие наибольшее значение в этом окне значений характеристик, а через Y^1 — характеристики, имеющие наименьшее значение в этом окне. Если $X^1 > Y^1$, то сдвигаем все последовательности в таблице, имеющие в окне α^1 значение X^1 , на одну позицию влево. Если $X^1 = Y^1$, то найдено повторение, равное X^1 .

Шаг 3. Повторяем шаги 1 и 2 до тех пор, пока в окне α^t впервые возникнет пустое место (места), что означает конец поиска повторений.

Результатом работы алгоритма станут все значения характеристик X^l , $l = \overline{0, t-1}$, для которых $X^l = Y^l$ и t — момент остановки алгоритма.

Пример. Рассмотрим пример работы алгоритма для $M = \{1, 2, \dots, 8\}$, $m = 3$, $N = 5$ и образцов $X^0(1) = (8, 6, 5, 3, 2)$, $X^0(2) = (7, 5, 4, 2, 1)$, $X^0(3) = (8, 7, 5, 2, 1)$. Запишем векторы обучающих образцов в таблицу и рассмотрим последовательность действий с таблицей в соответствии с описанным алгоритмом, представленную на рисунке.

Из рисунка следует, что работа алгоритма закончилась в момент $t = 7$, а результатом работы алгоритма стали повторения $X^3 = 5$ и $X^6 = 2$.



Последовательность работы алгоритма

4 Обоснование алгоритма

Обозначим через $n = 0, 1, \dots$ число всех сдвигов строк таблицы в шагах 1 и 2 в работе алгоритма. Тогда имеет место следующая лемма.

Лемма 1. Если на шаге n алгоритма в получившемся векторе α^n имеет место неравенство $X^n > Y^n$, то значение X^n не может войти в число повторившихся характеристик, присутствующих во всех t образцах обучающей выборки.

Доказательство. Для $n = 0$ значение X^0 будет всегда больше всех значений в последовательности образца, в которой первый номер характеристики равен Y^0 . Последовательность, которой принадлежит Y^0 , упорядочена по убыванию, поэтому X^0 не может встретиться в этой последовательности. Аналогично, если $X^n > Y^n$, то значение X^n не может войти в число повторившихся, если рассматривать только участки последовательностей справа от окна α^n .

Пусть для простоты $X^n = X^n(1, 1)$ — максимум в окне α^n и $X^n > Y^n$, где $Y^n = X^n(1, m)$. Тогда X^n не может встретиться в m -й строке для всех дальнейших членов этой последовательности. В следующем окне α^{n+1} максимум может быть достигнут в другой строке за счет снижения значений первой последовательности при ее сдвигах. Также возможно равенство $X^{n+1} = Y^{n+1}$, где $Y^{n+1} = Y^n$. В этом случае возникает повторение.

Докажем, что в m -й строке не могло встретиться значение X^n до того, как оно появилось в первой строке. Будем рассуждать от противного и предположим, что $X^{n-k}(1, m) = X^n$ для некоторого $k > 0$. Пусть последовательность в первой строке в окне α^{n-k} принимает значение X' . Тогда $X' \geq X^n$. Для того чтобы m -я строка была сдвинута влево на k позиций в ходе работы алгоритма, она должна была принимать в соответствующих окнах максимальное значение. Эти значения в последовательности m -й строки не могут быть больше, чем X^n (из-за убывания последовательности и сделанного предположения). Это возможно только тогда,

когда $X' = X^n = X^{n-j}(1, m)$ при $j = \overline{1, k}$. Это противоречит условию, что в окне α^n выполнено условие $X^n > Y^n$. Если $X^n = Y^n$, то это повторение. Лемма 1 доказана.

Следствие. При $X^n > Y^n$ повторения могут возникнуть только при снижении значений характеристик при сдвигах строк, имеющих максимальные значения до уровней меньших значений в других строках.

Лемма 2. До своего останова алгоритм восстановит все полные повторения.

Доказательство. Доказательство будем вести от противного. Предположим, что существует характеристика X^* , которая встретилась во всех образцах обучающей выборки, но не выявлена алгоритмом.

На нулевом шаге алгоритма все последовательности образцов записываются в строки таблицы размера $m \times 2N$ друг под другом начиная с $(N+1)$ -го столбца. По предположению, в каждой строке встречается X^* . Пусть для простоты X^* в каждой из строк таблицы находится в столбце с номером $N+1+d_i$, $i = \overline{1, m}$, где $d_1 < d_2 < \dots < d_m$. Появление X^* при работе алгоритма возможно только в порядке от первой строки до m -й строки. В самом деле, если впервые значение характеристики X^* в соответствующем окне появилось во второй строке, то появление X^* из первой строки в одном и том же окне невозможно. Для того чтобы X^* из первой и второй строки попали в одно окно, необходимо, чтобы произошел сдвиг первой строки влево на $d_2 - d_1$ позиций. Но в промежутке между $(N+1+d_1)$ -м и $(N+1+d_2)$ -м столбцами значения характеристик второй строки больше значений характеристик первой строки. Следовательно, сдвиги первой строки не смогут обеспечить попадание X^* из первой строки в одно и то же окно с X^* второй строки. Этот вывод распространяется на все пары появления X^* .

Отметим, что значения характеристик в m -й строке на промежутке между $(N+1+d_{m-1})$ -м и $(N+1+d_m)$ -м столбцами таблицы больше, чем значения характеристик в $(m-1)$ -й строке. Поэтому в паре m -й и $(m-1)$ -й строк значения характеристик в m -й строке будут доминировать на всем промежутке между $(N+1+d_1)$ -м и $(N+1+d_2)$ -м столбцами, что позволит путем сдвигов создать частичное повторение. Для остальных пар можно провести аналогичные рассуждения. Легко видеть, что такая последовательность действий — часть последовательности сдвигов образцов и работа алгоритма приводит к получению повторения X^* . Полученное противоречие доказывает лемму 2.

Лемма 3. Если до появления пустого значения в окне алгоритм не нашел повторений, то с вероятностью, стремящейся к единице, верно утверждение, что причины P в построенной обучающей выборке нет.

Доказательство следует из леммы 2 и теоремы [1] об отсутствии случайных повторений при выполнении условий этой теоремы.

Лемма 4. Сложность алгоритма поиска всех повторений в упорядоченных образцах обучающей выборки не превосходит mN .

Доказательство. Каждый шаг алгоритма в худшем варианте соответствует сдвигу одной строки таблицы влево на одну позицию от очередного окна α . Поэтому число сдвигов не превосходит mN .

Теорема. Сложность алгоритма поиска всех повторений с учетом операций упорядочения не превосходит $mN \log_2 N$.

Доказательство. Согласно лемме 4, основная сложность поиска повторений связана с упорядочением всех образцов обучающей выборки. Теорема доказана.

Доказанные результаты дают обоснование эффективности построенного алгоритма поиска повторений в обучающей выборке.

5 Заключение

В работе построен алгоритм поиска повторений характеристик из большого алфавита, встречающихся в различных последовательностях большой длины. Приведен пример появления задач подобного типа из практики поиска мошеннических действий в множестве получателей потребительских кредитов в банках. Задача построения обучающей выборки для поиска таких действий с использованием причинно-следственных связей сводится к поиску повторений одинаковых характеристик, присутствующих во всех образцах обучающей выборки. При этом используется ансамблевый подход к поиску используемых методов мошеннических схем.

Для m образцов обучающей выборки проблема поиска повторения существенна для практической реализации всего подхода к классификации на основе причинно-следственных связей. Поэтому доказанная теорема служит основой реализуемости рассматриваемого подхода.

Литература

1. Грушо А. А., Грушо Н. А., Забейайло М. И., Кульченков В. В., Тимонина Е. Е. Классификация малых наборов данных большой размерности // Информатика и её применения, 2025. Т. 19. Вып. 3. С. 67–72. doi: 10.14357/19922264250308. EDN: SUSRRC.
2. Fahd E., Fahad A., Amal A. Enhancing loan fraud detection process in the banking sector using data mining techniques // Indonesian J. Electrical Engineering Computer Science, 2023. Vol. 32. No. 2. P. 956–967. doi: 10.11591/ijeecs.v32.i2.pp956-967.
3. Akinjole A., Shobayo O., Popoola J., Okoyeigbo O., Ogunleye B. Ensemble-based machine learning algorithm for loan default risk prediction // Mathematics, 2024. Vol. 12. No. 21. Art. 3423. 31 p. doi: 10.3390/math12213423.
4. Воробьев И. А. Исследования по разработке методов противодействия мошенничеству в финансовых организациях с применением машинного обучения: Дис. . . . канд. техн. наук. — М.: НИУ ВШЭ, 2024. 103 с.
5. Glymour C., Zhang K., Spirtes P. Review of causal discovery Methods based on graphical models // Frontiers Genetics, 2019. Vol. 10. Art. 524. 15 p. doi: 10.3389/fgene.2019.00524.

6. *Richens J. G., Lee C. M., Johri S.* Improving the accuracy of medical diagnosis with causal machine learning // *Nat. Commun.*, 2020. Vol. 11. Art. 3923. 9 p. doi: 10.1038/s41467-020-17419-7.
7. *Grusho A. A., Grusho N. A., Zabezhalo M. I., Timonina E. E.* Localization of the root cause of the anomaly // *Autom. Control Comp. S.*, 2021. Vol. 55. Iss. 8. P. 978–983. doi: 10.3103/S0146411621080137. EDN: YALOOQ.
8. *Грушо А. А., Грушо Н. А., Забейхайло М. И., Кульченков В. В., Тимонина Е. Е., Шоргин С. Я.* Причинно-следственные связи в задачах классификации // *Информатика и её применения*, 2023. Т. 17. Вып. 1. С. 43–49. doi: 10.14357/19922264230106. EDN: DTQZPK.
9. *Грушо А. А., Забейхайло М. И., Кульченков В. В., Смирнов Д. В., Тимонина Е. Е., Шоргин С. Я.* Причинно-следственные связи в задачах анализа ненаблюдаемых свойств процессов // *Системы и средства информатики*, 2023. Т. 33. № 2. С. 71–78. doi: 10.14357/08696527230207. EDN: HMOFGP.
10. *Грушо А. А., Грушо Н. А., Забейхайло М. И., Тимонина Е. Е., Шоргин С. Я.* Сложные причинно-следственные связи // *Информатика и её применения*, 2023. Т. 17. Вып. 2. С. 84–89. doi: 10.14357/19922264230212. EDN: TGXQIW.
11. *Грушо А. А., Грушо Н. А., Забейхайло М. И., Смирнов Д. В., Тимонина Е. Е.* Классификация с помощью причинно-следственных связей // *Информатика и её применения*, 2023. Т. 17. Вып. 3. С. 71–75. doi: 10.14357/19922264230310. EDN: AKWBZD.
12. *Höfler M.* Causal inference based on counterfactuals // *BMC Med. Res. Methodol.*, 2005. Vol. 5. Art. 28. 12 p. doi: 10.1186/1471-2288-5-28.
13. *Нивергельт Ю., Фаррар Дж., Рейнголд Э.* Машинный подход к решению математических задач / Пер. с англ. — М.: Мир, 1977. 352 с. (*Nievergelt J., Farrar J. C., Reingold E. M.* Computer approaches to mathematical problems. — Englewood Cliffs, NJ, USA: Prentice-Hall, 1974. 257 p.)
14. *Ахо А., Хопкрофт Дж., Ульман Дж. Д.* Структуры данных и алгоритмы / Пер. с англ. — М.: Вильямс, 2016. 400 с. (*Aho A. V., Hopcroft J. E., Ullman J. D.* Data structures and algorithms. — Reading, MA, USA: Addison-Wesley, 1983. 427 p.)

Поступила в редакцию 04.08.2025

Принята к публикации 15.10.2025

COMPLEXITY OF THE ALGORITHM FOR FINDING COINCIDENCES IN SEVERAL SEQUENCES

A. A. Grusho¹, M. I. Zabezhalo¹, V. V. Kulchenkov², and E. E. Timonina¹

¹Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

²VTB Bank, 43-1 Vorontsovskaya Str., Moscow 109147, Russian Federation

Abstract: Analysis of data for the presence of effects of an unknown but present cause arises in many tasks and one of the examples given in the work is related to the search for signs of fraud among many recipients of a consumer loan in a bank. To build the initial data, a method was chosen in which signs of fraud appear in

transactional activity after receiving a loan, namely, the signs are based on how the funds are withdrawn. The example given is a special case of situations when in a limited set of precedents of data having a large dimension, the effects of one cause are present and repeated. Under these conditions, the task of finding repetitions of consequences of effects is of great importance. An algorithm for such a search has been built, which has a complexity less than quadratic. The complexity of the constructed algorithm for finding all coincidences in m ordered precedents does not exceed mN where N is the length of all precedents. Given the complexity of ordering each precedent when there is the initial ordering of the entire set of characteristics, the complexity of solving the problem does not exceed $mN \log_2 N$.

Keywords: complexity of the classification task; machine learning; cause and effect relationships; searching for coincidences in the sequence of sets

DOI: 10.14357/08696527250411

EDN: GAAUSE

References

1. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, V. V. Kulchenkov, and E. E. Timonina. 2025. Klassifikatsiya malykh naborov dannykh bol'shoy razmernosti [Classification of small sets of data of large dimension]. *Informatika i ee Primeneniya — Inform. Appl.* 19(3):67–72. doi: 10.14357/19922264250308. EDN: SUSRRC.
2. Fahd, E., A. Fahad, and A. Amal. 2023. Enhancing loan fraud detection process in the banking sector using data mining techniques. *Indonesian J. Electrical Engineering Computer Science* 32(2):956–967. doi: 10.11591/ijeecs.v32.i2.pp956-967.
3. Akinjole, A., O. Shobayo, J. Popoola, O. Okoyeigbo, and B. Ogunleye. 2024. Ensemble-based machine learning algorithm for loan default risk prediction. *Mathematics* 12(21):3423. 31 p. doi: 10.3390/math12213423.
4. Vorobyev, I. A. 2024. Issledovaniya po razrabotke metodov protivodeystviya moshenichestvu v finansovykh organizatsiyakh s primeneniem mashinnogo obucheniya [Research on the development of methods for combating fraud in financial organizations using machine learning]. Moscow: HSE University. PhD Thesis. 103 p.
5. Glymour, C., K. Zhang, and P. Spirtes. 2019. Review of causal discovery methods based on graphical models. *Frontiers Genetics* 10:524. 15 p. doi: 10.3389/fgene.2019.00524.
6. Richens, J. G., C. M. Lee, and S. Johri. 2020. Improving the accuracy of medical diagnosis with causal machine learning. *Nat. Commun.* 11:3923. 9 p. doi: 10.1038/s41467-020-17419-7.
7. Grusho, A., N. Grusho, M. Zabezhailo, and E. Timonina. 2021. Localization of the root cause of the anomaly. *Autom. Control Comp. S.* 55(8):978–983. doi: 10.3103/S0146411621080137. EDN: YALOOQ.
8. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, V. V. Kulchenkov, E. E. Timonina, and S. Ya. Shorgin. 2023. Prichinno-sledstvennyye svyazi v zadachakh klassifikatsii [Causal relationships in classification problems]. *Informatika i ee Primeneniya — Inform. Appl.* 17(1):43–49. doi: 10.14357/19922264230106. EDN: DTQZPK.

9. Grusho, A. A., M. I. Zabezhailo, V. V. Kulchenkov, D. V. Smirnov, E. E. Timonina, and S. Ya. Shorgin. 2023. Prichinno-sledstvennyye svyazi v zadachakh analiza nenablyudaemykh svoystv protsessov [Cause-and-effect relationships in analysis of unobservable process properties]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(2):71–78. doi: 10.14357/08696527230207. EDN: HMOFPG.
10. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, E. E. Timonina, and S. Ya. Shorgin. 2023. Slozhnye prichinno-sledstvennyye svyazi [Complex cause-and-effect relationships]. *Informatika i ee Primeneniya — Inform. Appl.* 17(2):84–89. doi: 10.14357/19922264230212. EDN: TGXQIW.
11. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2023. Klassifikatsiya s pomoshch'yu prichinno-sledstvennykh svyazey [Classification by cause-and-effect relationships]. *Informatika i ee Primeneniya — Inform. Appl.* 17(3):71–75. doi: 10.14357/19922264230310. EDN: AKWBZD.
12. Hófler, M. 2005. Causal inference based on counterfactuals. *BMC Med. Res. Methodol.* 5:28. 12 p. doi: 10.1186/1471-2288-5-28.
13. Nievergelt, J., J. C. Farrar, and E. M. Reingold. 1974. *Computer approaches to mathematical problems*. Englewood Cliffs, NJ: Prentice-Hall. 257 p.
14. Aho, A. V., J. E. Hopcroft, and J. D. Ullman. 1983. *Data structures and algorithms*. Reading, MA: Addison-Wesley. 427 p.

Received August 4, 2025

Accepted October 15, 2025

Contributors

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; grusho@yandex.ru

Zabezhailo Michael I. (b. 1956) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; m.zabezhailo@yandex.ru

Kulchenkov Vladislav V. (b. 1989) — deputy head, Portfolio Analysis Department, VTB Bank, 43-1 Vorontsovskaya Str., Moscow 109147, Russian Federation; vlad.kulchenkov@gmail.com

Timonina Elena E. (b. 1952) — Doctor of Science in technology, professor, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; eltimon@yandex.ru

Адамович Игорь Михайлович (р. 1934) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Бесчастный Виталий Александрович (р. 1992) — кандидат физико-математических наук, доцент кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов им. Патриса Лумумбы

Верхотуров Андрей Леонидович (р. 1986) — научный сотрудник Вычислительного центра Дальневосточного отделения Российской академии наук

Волков Олег Игоревич (р. 1964) — ведущий программист Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Дружинина Ольга Валентиновна (р. 1967) — доктор физико-математических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Дубровин Константин Николаевич (р. 1993) — научный сотрудник Вычислительного центра Дальневосточного отделения Российской академии наук

Дьяченко Юрий Георгиевич (р. 1958) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Забежайло Михаил Иванович (р. 1956) — доктор физико-математических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Илларионова Любовь Викторовна (р. 1977) — кандидат физико-математических наук, старший научный сотрудник Вычислительного центра Дальневосточного отделения Российской академии наук

Кан Анна Владимировна (р. 1977) — кандидат технических наук, заместитель директора Всероссийского института научной и технической информации Российской академии наук; доцент Московского авиационного института (национального исследовательского университета)

Кононов Николай Олегович (р. 2000) — инженер Московского авиационного института (национального исследовательского университета)

Корепанов Эдуард Рудольфович (р. 1966) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кульченков Владислав Владимирович (р. 1989) — заместитель начальника управления портфельного анализа Банка ВТБ

Макаренкова Ирина Вячеславовна (р. 1964) — ведущий программист Федерального исследовательского центра «Информатика и управление» Российской академии наук

Максимова Вера Валентиновна (р. 1975) — ведущий программист Федерального исследовательского центра «Информатика и управление» Российской академии наук

Надоров Иван Сергеевич (р. 2003) — ассистент кафедры «Математическая кибернетика» Московского авиационного института (национального исследовательского университета)

Никитин Юрий Викторович (р. 1977) — начальник отдела Национального исследовательского центра «Институт имени Н. Е. Жуковского»

Никишин Дмитрий Александрович (р. 1976) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Орлов Георгий Александрович (р. 1994) — младший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Пантелеев Андрей Владимирович (р. 1955) — доктор физико-математических наук, профессор, заведующий кафедрой «Математическая кибернетика» Московского авиационного института (национального исследовательского университета)

Петров Алексей Алексеевич (р. 1991) — кандидат технических наук, доцент кафедры математического моделирования, компьютерных технологий и информационной безопасности Елецкого государственного университета имени И. А. Бунина

Плеханов Леонид Петрович (р. 1943) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Румовская София Борисовна (р. 1985) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Сомин Николай Владимирович (р. 1947) — кандидат физико-математических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Степанов Алексей Сергеевич (р. 1974) — доктор фармацевтических наук, ведущий научный сотрудник Дальневосточного научно-исследовательского института сельского хозяйства

Степанов Никита Вячеславович (р. 1996) — старший преподаватель кафедры инфокоммуникационных технологий и систем связи Санкт-Петербургского государственного университета аэрокосмического приборостроения

Ступников Сергей Александрович (р. 1978) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Турбин Николай Васильевич (р. 1989) — ведущий инженер Передовой инженерной школы Московского авиационного института (национального исследовательского университета)

Тюрликов Андрей Михайлович (р. 1957) — доктор технических наук, профессор, заведующий кафедрой инфокоммуникационных технологий и систем связи Санкт-Петербургского государственного университета аэрокосмического приборостроения

Фомина Елизавета Андреевна (р. 1998) — инженер Вычислительного центра Дальневосточного отделения Российской академии наук

Хилько Дмитрий Владимирович (р. 1987) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Хорошилов Александр Алексеевич (р. 1952) — доктор технических наук, ведущий научный сотрудник Всероссийского института научной и технической информации Российской академии наук; профессор Московского авиационного института (национального исследовательского университета)

Хорошилов Алексей Алексеевич (р. 1958) — кандидат технических наук, заведующий лабораторией Всероссийского института научной и технической информации Российской академии наук

Чечулин Иван Андреевич (р. 1988) — заместитель заведующего отделом Всероссийского института научной и технической информации Российской академии наук

Шарнин Михаил Михайлович (р. 1959) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шоргин Всеволод Сергеевич (р. 1978) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

АВТОРСКИЙ УКАЗАТЕЛЬ ЗА 2025 г.

	№	Стр.
Агаларов Я. М. Унимодальность взвешенной суммы частных критериев модели $G/M/1$ с обновлением и двухпороговым RED-подобным алгоритмом	3	71–89
Адамович И. М., Волков О. И. Подход к использованию анализа временных рядов для формирования гипотез в конкретно-историческом исследовании	4	143–153
Архипов П. О., Филиппских С. Л., Цуканов М. В. Особенности практической реализации метода аугментации малых объектов на основе сетей сверхвысокого разрешения	3	17–32
Бесчастный В. А., Тюрликов А. М., Степанов Н. В., Шоргин В. С. О применении древовидных алгоритмов разрешения конфликтов в процедуре случайного множественного доступа для устройств 5G NB-IoT	4	20–32
Брюхов Д. О. см. Ступников С. А.		
Бунтман Н. В. см. Игнатова С. Д.		
Васенин И. А. см. Тюрин С. Ф.		
Верхотуров А. Л. см. Степанов А. С.		
Волков О. И. см. Адамович И. М.		
Гончаров А. А. см. Игнатова С. Д.		
Греков М. М. Модель автоматизированной системы имитации атак на основе машинного обучения	3	105–116
Гридин В. Н., Салем Б. Р., Смирнов Д. С., Солодовников В. И. Исследование влияния механизма Лапласа на статистические характеристики исходных данных	1	71–94
Грушо А. А., Грушо Н. А., Забежайло М. И., Смирнов Д. В., Тимонина Е. Е., Шоргин С. Я. Анализ сбоев по косвенным признакам	1	59–70
Грушо А. А., Забежайло М. И., Кульченков В. В., Тимонина Е. Е. Сложность алгоритма поиска совпадений в нескольких последовательностях	4	154–163
Грушо Н. А. см. Грушо А. А.		
Добровольский Д. О., Зацман И. М. Интеграция электронного словаря с текстами параллельного корпуса: новый теоретический подход	1	111–124

	№	Стр.
Дружинина О. В., Корепанов Э. Р., Макаренко И. В., Максимова В. В., Петров А. А. Построение и анализ модели прогнозирования технического состояния буксовых узлов вагонов с применением интеллектуальных методов предиктивной аналитики	4	92–110
Дубровин К. Н. см. Степанов А. С.		
Дьяченко Д. Ю. см. Дьяченко Ю. Г.		
Дьяченко Ю. Г., Плеханов Л. П., Морозов Н. В., Степченков Д. Ю., Орлов Г. А., Дьяченко Д. Ю. Реализация функциональности синхронных триггеров в самосинхронном базисе	3	3–16
Дьяченко Ю. Г. см. Захаров В. Н.		
Дьяченко Ю. Г. см. Плеханов Л. П.		
Дьяченко Ю. Г. см. Тюрин С. Ф.		
Егоров В. Б. Процессоры данных как новый компонент инфраструктуры дата-центров	1	170–180
Забежайло М. И. см. Грушо А. А.		
Забежайло М. И. см. Грушо А. А.		
Захаров В. Н., Степченков Ю. А., Дьяченко Ю. Г., Морозов Н. В., Плеханов Л. П., Степченков Д. Ю. Свойства и оптимизация самосинхронных схем	1	149–169
Захаров В. Н. см. Соколов И. А.		
Зацаринный А. А., Козлов С. В. Эволюция концептуальных подходов к интеграции инфокоммуникационных и управляющих систем как основы перспективных организационных систем	3	130–144
Зацаринный А. А. см. Соколов И. А.		
Зацман И. М. см. Добровольский Д. О.		
Игнатова С. Д., Гончаров А. А., Бунтман Н. В. Метод автоматизированного выявления асимметрии пунктуации в параллельных текстах	2	103–115
Илларионова Л. В. см. Степанов А. С.		
Казанцев С. Ю., Сапожников М. В., Терехин Д. Н., Мионов Ю. Б., Шабанов Б. М. Оценка доступности атмосферной оптической связи в различных субъектах Российской Федерации	2	61–80
Кан А. В., Хорошилов Ал-др А., Чечулин И. А., Ступников С. А., Никитин Ю. В., Хорошилов Ал-ей А. Архитектура системы семантического анализа научно-технических документов	4	45–59

	№	Стр.
Кириков И. А., Листопад С. В., Румовская С. Б. Гибридная интеллектуальная многоагентная система как модель малого коллектива специалистов для решения практических проблем	1	125–148
Козлов С. В. см. Зацаринный А. А.		
Колин К. К. см. Соколов И. А.		
Конашенкова Т. Д. см. Сеницын И. Н.		
Конашенкова Т. Д. см. Сеницын И. Н.		
Кононов Н. О. см. Пантелеев А. В.		
Корепанов Э. Р. см. Дружинина О. В.		
Корепанов Э. Р. см. Сеницын И. Н.		
Корепанов Э. Р. см. Сеницын И. Н.		
Кривенко М. П. ARIMA-моделирование последовательности времен пребывания в системе массового обслуживания	3	90–104
Кульченко В. В. см. Грушо А. А.		
Ладиков А. В. Обезличивание пользовательских данных с привлечением доверенного центра деперсонализации	3	117–129
Листопад С. В., Лучко А. С. Анализ подходов к гибридизации метода рассуждений по прецедентам в системах поддержки принятия врачебных решений	2	81–102
Листопад С. В. см. Кириков И. А.		
Лучко А. С. см. Листопад С. В.		
Макаренкова И. В. см. Дружинина О. В.		
Максимова В. В. см. Дружинина О. В.		
Миронов Ю. Б. см. Казанцев С. Ю.		
Морозов Н. В. см. Дьяченко Ю. Г.		
Морозов Н. В. см. Захаров В. Н.		
Надоров И. С. см. Пантелеев А. В.		
Никитин Ю. В. см. Кан А. В.		
Никишин Д. А. Современные методы извлечения географического знания из больших массивов пространственных данных	4	129–142
Орлов Г. А. см. Дьяченко Ю. Г.		
Орлов Г. А. см. Плеханов Л. П.		
Пантелеев А. В., Турбин Н. В., Надоров И. С., Кононов Н. О. Метаэвристические алгоритмы для идентификации параметров трехмерной модели снижения усталостной жесткости композиционных материалов	4	73–91
Петров А. А. см. Дружинина О. В.		
Плеханов Л. П., Дьяченко Ю. Г., Хилько Д. В., Орлов Г. А. Оптимизация синтеза последовательностных самосинхронных схем по синхронному описанию	4	4–19
Плеханов Л. П. см. Дьяченко Ю. Г.		

	№	Стр.
Плеханов Л. П. см. Захаров В. Н.		
Румовская С. Б. Нейро-нечеткая модель и прецедентная экспертная система трансформационного гибридного интеллекта прогнозирования исходов острого панкреатита	4	33–44
Румовская С. Б. см. Кириков И. А.		
Салем Б. Р. см. Гридин В. Н.		
Сапожников М. В. см. Казанцев С. Ю.		
Синицын В. И. см. Синицын И. Н.		
Синицын В. И. см. Синицын И. Н.		
Синицын И. Н. Методы нормальной субоптимальной фильтрации в наблюдаемых неявных гауссовских стохастических системах	1	41–58
Синицын И. Н., Синицын В. И., Корепанов Э. Р., Конашенкова Т. Д. Алгоритм моделирования векторного стохастического процесса посредством его канонического разложения на основе многослойной вейвлет-нейронной сети	2	17–30
Синицын И. Н., Синицын В. И., Корепанов Э. Р., Конашенкова Т. Д. Нейросетевой байесовский синтез многомерной линейной стохастической системы	3	33–53
Скворцов Н. А. см. Ступников С. А.		
Смирнов Д. В. см. Грушо А. А.		
Смирнов Д. С. см. Гридин В. Н.		
Соколов И. А., Зацаринный А. А., Захаров В. Н., Колин К. К. О роли академика И. А. Мизина в развитии информационных технологий и их применении в государственных интересах: к 90-летию со дня рождения	1	3–40
Солодовников В. И. см. Гридин В. Н.		
Сомин Н. В. см. Шарнин М. М.		
Степанов А. С., Илларионова Л. В., Дубровин К. Н., Фомина Е. А., Верхотуров А. Л. Картографирование пахотных земель муниципальных образований Хабаровского края по данным спутника «Метеор-М» № 2	4	111–128
Степанов Н. В. см. Бесчастный В. А.		
Степченков Д. Ю. см. Дьяченко Ю. Г.		
Степченков Д. Ю. см. Захаров В. Н.		
Степченков Ю. А. Безыбыточное самосинхронное кодирование: реализация информационного канала	2	45–60
Степченков Ю. А. см. Захаров В. Н.		
Степченков Ю. А. см. Тюрин С. Ф.		

	№	Стр.
Ступников С. А., Скворцов Н. А., Брюхов Д. О. Перспективные методы реализации инкрементального обновления материализованных представлений в современных реляционных системах управления базами данных	1	95–110
Ступников С. А. см. Кан А. В.		
Сучков А. П. Некоторые подходы к управлению мониторингом ситуационной обстановки	2	116–126
Терехин Д. Н. см. Казанцев С. Ю.		
Тимонина Е. Е. см. Грушо А. А.		
Тимонина Е. Е. см. Грушо А. А.		
Турбин Н. В. см. Пантелеев А. В.		
Тюрин С. Ф., Васенин И. А., Степченков Ю. А., Дьяченко Ю. Г. Комбинированное кодирование в элементах программируемых логических интегральных схем	2	3–16
Тюрликов А. М. см. Бесчастный В. А.		
Урюпин И. В. Об универсальных моделях состояния для задач слежения и навигации беспилотных движущихся объектов	2	31–44
Урюпин И. В. Типовые модели системы наблюдения для задач отслеживания и навигации беспилотных движущихся объектов	3	54–70
Филиппских С. Л. см. Архипов П. О.		
Фомина Е. А. см. Степанов А. С.		
Хилько Д. В. см. Плеханов Л. П.		
Хорошилов Ал-др А. см. Кан А. В.		
Хорошилов Ал-ей А. см. Кан А. В.		
Цуканов М. В. см. Архипов П. О.		
Чечулин И. А. см. Кан А. В.		
Шабанов Б. М. см. Казанцев С. Ю.		
Шарнин М. М., Сомин Н. В. Движение, скорость и траектории представлений ключевых слов в векторном пространстве языковой модели	4	60–72
Шоргин В. С. см. Бесчастный В. А.		
Шоргин С. Я. см. Грушо А. А.		

2025 AUTHOR INDEX

	No.	Page
Adamovich I. M. and Volkov O. I. An approach to using time series analysis in hypotheses forming for concrete historical investigation	4	143–153
Agalarov Ya. M. Unimodality of the weighted sum of partial criteria of the $G/M/1$ model with updating and a two-threshold red-like algorithm	3	71–89
Arkhipov P. O., Philippskih S. L., and Tsukanov M. V. Development of a small-object augmentation method based on super-resolution networks	3	17–32
Beschastnyi V. A., Turlikov A. M., Stepanov N. V., and Shorgin V. S. Tree-splitting algorithms for massive random access of 5G NB-IoT communications	4	20–32
Briukhov D. O. see Stupnikov S. A.		
Buntman N. V. see Ignatova S. D.		
Charnine M. M. and Somin N. V. Movement, velocity, and trajectories of keyword representations in the vector space of the language model	4	60–72
Chechulin I. A. see Kan A. V.		
Diachenko D. Yu. see Diachenko Yu. G.		
Diachenko Yu. G., Plekhanov L. P., Morozov N. V., Stephenkov D. Yu., Orlov G. A., and Diachenko D. Yu. Implementation of synchronous flip-flop and latch functionality in a self-timed basis	3	3–16
Diachenko Yu. G. see Plekhanov L. P.		
Diachenko Yu. G. see Tyurin S. F.		
Diachenko Yu. G. see Zakharov V. N.		
Dobrovol'skij D. O. and Zatsman I. M. Integration of a digital dictionary with parallel corpus texts: A new theoretical approach	1	111–124
Druzhinina O. V., Korepanov E. R., Makarenkova I. V., Maksimova V. V., and Petrov A. A. Construction and analysis of model to predict the technical state of railway car axle boxes using intelligent predictive analytics methods	4	92–110
Dubrovin K. N. see Stepanov A. S.		

	No.	Page
Egorov V. B. Data processing units as a new component of the data center infrastructure	1	170–180
Fomina E. A. see Stepanov A. S.		
Goncharov A. A. see Ignatova S. D.		
Grekov M. M. Model of automated attack simulation system based on machine learning	3	105–116
Gridin V. N., Salem B. R., Smirnov D. S., and Solodovnikov V. I. The Laplace mechanism impact on the statistical characteristics of the original data	1	71–94
Grusho A. A., Grusho N. A., Zabezhalo M. I., Smirnov D. V., Timonina E. E., and Shorgin S. Ya. Analysis of failures by indirect features	1	59–70
Grusho A. A., Zabezhalo M. I., Kulchenkov V. V., and Timonina E. E. Complexity of the algorithm for finding coincidences in several sequences	4	154–163
Grusho N. A. see Grusho A. A.		
Ignatova S. D., Goncharov A. A., and Buntman N. V. Method of automated detection of punctuation asymmetry in parallel texts	2	103–115
Illarionova L. V. see Stepanov A. S.		
Kan A. V., Khoroshilov Alexander A., Chechulin I. A., Stupnikov S. A., Nikitin Yu. V., and Khoroshilov Alexey A. An architecture of a system for the semantic analysis of the scientific and technical documents	4	45–59
Kazantsev S. Yu., Sapozhnikov M. V., Terekhin D. N., Mironov Yu. B., and Shabanov B. M. Assessment of the availability of atmospheric optical communications in various regions of the Russian Federation	2	61–80
Khilko D. V. see Plekhanov L. P.		
Khoroshilov Alexander A. see Kan A. V.		
Khoroshilov Alexey A. see Kan A. V.		
Kirikov I. A., Listopad S. V., and Rumovskaya S. B. Hybrid intelligent multiagent system as a model of a small team of specialists for solving practical problems	1	125–148
Kolin K. K. see Sokolov I. A.		
Konashenkova T. D. see Sinitsyn I. N.		
Konashenkova T. D. see Sinitsyn I. N.		
Kononov N. O. see Panteleev A. V.		
Korepanov E. R. see Druzhinina O. V.		
Korepanov E. R. see Sinitsyn I. N.		
Korepanov E. R. see Sinitsyn I. N.		
Kozlov S. V. see Zatsarinny A. A.		

	No.	Page
Krivenko M. P. ARIMA-modeling of the sequence of sojourn times in queueing systems	3	90–104
Kulchenkov V. V. see Grusho A. A.		
Ladikov A. V. Depersinalization of user data with the involvement of a trusted depersonalization center	3	117–129
Listopad S. V. and Luchko A. S. Analysis of approaches to hybridization of the case-based reasoning method in medical decision support systems	2	81–102
Listopad S. V. see Kirikov I. A.		
Luchko A. S. see Listopad S. V.		
Makarenkova I. V. see Druzhinina O. V.		
Maksimova V. V. see Druzhinina O. V.		
Mironov Yu. B. see Kazantsev S. Yu.		
Morozov N. V. see Diachenko Yu. G.		
Morozov N. V. see Zakharov V. N.		
Nadorov I. S. see Pantelev A. V.		
Nikishin D. A. Modern methods of extracting geographical knowledge in large arrays of spatial data	4	129–142
Nikitin Yu. V. see Kan A. V.		
Orlov G. A. see Diachenko Yu. G.		
Orlov G. A. see Plekhanov L. P.		
Pantelev A. V., Turbin N. V., Nadorov I. S., and Kononov N. O. Metaheuristic optimization algorithms of parametric identification of a three-dimensional model for reducing the fatigue stiffness of composite materials	4	73–91
Petrov A. A. see Druzhinina O. V.		
Philippskih S. L. see P. O. Arkhipov		
Plekhanov L. P., Diachenko Yu. G., Khilko D. V., and Orlov G. A. Template method in synthesis of self-timed digital circuits	4	4–19
Plekhanov L. P. see Diachenko Yu. G.		
Plekhanov L. P. see Zakharov V. N.		
Rumovskaya S. B. Neuro-fuzzy model and precedent expert system of the transformational hybrid intelligence for predicting outcomes of the acute pancreatitis	4	33–44
Rumovskaya S. B. see Kirikov I. A.		
Salem B. R. see Gridin V. N.		
Sapozhnikov M. V. see Kazantsev S. Yu.		
Shabanov B. M. see Kazantsev S. Yu.		
Shorgin S. Ya. see Grusho A. A.		
Shorgin V. S. see Beschastnyi V. A.		

	No.	Page
Sinitsyn I. N. Normal suboptimal filtering methods in implicit observable Gaussian stochastic systems	1	41–58
Sinitsyn I. N., Sinitsyn V. I., Korepanov E. R., and Konashenkova T. D. Modeling algorithms for vector stochastic process by canonical expansions based on multi-layer wavelet neural network	2	17–30
Sinitsyn I. N., Sinitsyn V. I., Korepanov E. R., and Konashenkova T. D. Neural networks Bayes synthesis of multidimensional linear stochastic system	3	33–53
Sinitsyn V. I. see Sinitsyn I. N.		
Sinitsyn V. I. see Sinitsyn I. N.		
Skvortsov N. A. see Stupnikov S. A.		
Smirnov D. S. see Gridin V. N.		
Smirnov D. V. see Grusho A. A.		
Sokolov I. A., Zatsarinny A. A., Zakharov V. N., and Kolin K. K. The role of the academician I. A. Mizin in information technology development and application for government needs: On the 90th anniversary of his birth	1	3–40
Solodovnikov V. I. see Gridin V. N.		
Somin N. V. see Charnine M. M.		
Stepanov A. S., Illarionova L. V., Dubrovin K. N., Fomina E. A., and Verkhoturov A. L. Mapping of Khabarovsk Krai arable lands using Meteor-M No. 2 satellite data	4	111–128
Stepanov N. V. see Beschastnyi V. A.		
Stepchenkov D. Yu. see Diachenko Yu. G.		
Stepchenkov D. Yu. see Zakharov V. N.		
Stepchenkov Yu. A. Nonredundant self-timed coding: Implementation of the information channel	2	45–60
Stepchenkov Yu. A. see Tyurin S. F.		
Stepchenkov Yu. A. see Zakharov V. N.		
Stupnikov S. A., Skvortsov N. A., and Briukhov D. O. Advanced methods for implementation of incremental view maintenance in modern relational database management systems	1	95–110
Stupnikov S. A. see Kan A. V.		
Suchkov A. P. Some approaches to managing situational monitoring	2	116–126
Terekhin D. N. see Kazantsev S. Yu.		
Timonina E. E. see Grusho A. A.		
Timonina E. E. see Grusho A. A.		
Tsukanov M. V. see P. O. Arkhipov		
Turbin N. V. see Panteleev A. V.		
Turlikov A. M. see Beschastnyi V. A.		

	No.	Page
Tyurin S. F., Vasenin I. A., Stepchenkov Yu. A., and Diachenko Yu. G. Combined encoding in elements of field-programmable gate arrays	2	3–16
Uryupin I. V. On universal state models for tracking and navigation tasks of unmanned moving objects	2	31–44
Uryupin I. V. Typical models of observation system for tracking and navigation of unmanned moving objects	3	54–70
Vasenin I. A. see Tyurin S. F.		
Verkhoturov A. L. see Stepanov A. S.		
Volkov O. I. see Adamovich I. M.		
Zabezhailo M. I. see Grusho A. A.		
Zabezhailo M. I. see Grusho A. A.		
Zakharov V. N., Stepchenkov Yu. A., Diachenko Yu. G., Morozov N. V., Plekhanov L. P., and Stepchenkov D. Yu. Properties and optimization of self-timed circuits	1	149–169
Zakharov V. N. see Sokolov I. A.		
Zatsarinny A. A. and Kozlov S. V. Evolution of conceptual approaches to the integration of information, control, and telecommunication systems as the basis of promising organizational systems	3	130–144
Zatsarinny A. A. see Sokolov I. A.		
Zatsman I. M. see Dobrovol'skij D. O.		

Правила подготовки рукописей статей для публикации в журнале «Системы и средства информатики»

Журнал «Системы и средства информатики» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информационных технологий.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- информационно-телекоммуникационные системы и средства их построения;
- архитектура и программное обеспечение вычислительных машин, комплексов и сетей;
- методы и средства защиты информации.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Если при подготовке статьи авторы использовали инструменты на основе интеллекта, они обязаны включить эту информацию в текст статьи.

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или L^AT_EX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.
7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — A4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху и снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 10 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны оформления представлены в интернете:

http://www.ipiran.ru/journal/template_iiep_ssi_2024.zip

8. Статья должна содержать следующую информацию на **русском и английском языках**:
 - название статьи;
 - Ф.И.О. авторов, на английском можно только имя и фамилию;
 - место работы, с указанием города и страны и электронного адреса каждого автора;
 - сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/collected/2019_29_03_rus/authors.asp и
http://www.ipiran.ru/journal/collected/2019_29_03_eng/authors.asp;
 - аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
 - ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами.
 - источники финансирования работы (ссылка на гранты, проекты, поддерживающие организации и т. п.).
9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала.
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/bgn/> бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Примеры ссылок на различные виды публикаций в списке “References”:

Описание статьи из журнала:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursoberegayushchie tekhnologii nedropol’zovaniya i povyseniya neftegazootdachi”* [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W.* Vibration problems in engineering. — 4th ed. — New York, NY, USA: Wiley, 1974. 521 p.)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. New York, NY: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor [Mathematical modeling of the plasma in the compact torus]. Moscow. D.Sc. Diss. 272 p.

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovary informatzionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
 - указывать в поле subject (тема) название журнала и фамилию автора;
 - указывать в тексте письма название статьи, авторов и журнал, в который направляется статья;
 - использовать attach (присоединение);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Системы и средства информатики» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Системы и средства информатики»:

Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05

e-mail: ssi@frccsc.ru (Стригина Светлана Николаевна)

<http://www.ipiran.ru/journal/collected>

Requirements for manuscripts submitted to Journal “Systems and Means of Informatics”

Journal “Systems and Means of Informatics” publishes theoretical, review, and discussion articles on the research and development in the field of information technology.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

Topics covered include the following areas:

- information and communication systems and tools of their design;
- architecture and software of computational complexes and networks; and
- methods and tools of information protection.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . . ,” pass to the Founder and the Editorial Board of the Journal “Systems and Means of Informatics” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).”

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

The Editorial Board has the right to request from the authors an official expert conclusion that the submitted article has no classified data prohibited for publication.

If authors used artificial intelligence (AI)-based tools in preparing their manuscript, they must include this information in the text of the article.

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If authors receive their article for correction after reviewing, it does not mean that the article is approved to be published. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or L^AT_EX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.

7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 10 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site:

http://www.ipiran.ru/journal/template_jiep_ssi_2024.zip

8. Articles should enclose data both in **Russian and English**:

- title;
- author's name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format (see site):
http://www.ipiran.ru/journal/collected/2019_29_03_rus/authors.asp and
http://www.ipiran.ru/journal/collected/2019_29_03_eng/authors.asp;
- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae.
- Indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.
Important! Keywords must not be sentences.
- Acknowledgments.

9. References. Russian references have to be presented both in English translation and in Latin transliteration (refer <http://www.translit.net/ru/bgn/>).

Please take into account the following examples of Russian references appearance:

Article in journal:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Journal article in electronic format:

Swaminathan, V., E. Lepkoswka-White, and B.P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Conference proceedings:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povyshe-niya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Books and other monographs:

Lindorf, L. S., and L. G. Mamikoniants, eds. 1972. *Ekspluatatsiya turbogenera-torov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publ. 352 p.

Dissertation and Thesis:

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovary informat-ionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 23 p.

State standards and patents:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichstva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publ. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
 - the journal title and author's name in the "Subject" field;
 - the article title, authors' names, and the journal title, whereto the paper is being submitted, in the text of the e-mail;
 - an article and additional materials have to be attached using the "attach" function;
 - an electronic version of the article should contain the file with the text and a separate file with figures.
12. "System and Means of Informatics" journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia

Ph.: +7 (499)135 86 92, Fax: +7 (495)930 45 05

e-mail: ssi@frccsc.ru (to Svetlana Strigina)

http://www.ipiran.ru/english/journal_systems.asp

SYSTEMS AND MEANS OF INFORMATICS (SISTEMY I SREDSTVA INFORMATIKI)

SCIENTIFIC JOURNAL

Volume 35 No.4 Year 2025

Editor-in-Chief and Chair of Editorial Council

Academician I. A. Sokolov

IN THIS ISSUE:

TEMPLATE METHOD IN SYNTHESIS OF SELF-TIMED DIGITAL CIRCUITS <i>L. P. Plekhanov, Y. G. Diachenko, D. V. Khilko, and G. A. Orlov</i>	4
TREE-SPLITTING ALGORITHMS FOR MASSIVE RANDOM ACCESS OF 5G NB-IoT COMMUNICATIONS <i>V. A. Beschastnyi, A. M. Turlikov, N. V. Stepanov, and V. S. Shorgin</i>	20
NEURO-FUZZY MODEL AND PRECEDENT EXPERT SYSTEM OF THE TRANSFORMATIONAL HYBRID INTELLIGENCE FOR PREDICTING OUTCOMES OF THE ACUTE PANCREATITIS <i>S. B. Rumovskaya</i>	33
AN ARCHITECTURE OF A SYSTEM FOR THE SEMANTIC ANALYSIS OF THE SCIENTIFIC AND TECHNICAL DOCUMENTS <i>A. V. Kan, Alexander A. Khoroshilov, I. A. Chechulin, S. A. Stupnikov, Yu. V. Nikitin, and Alexey A. Khoroshilov</i>	45
MOVEMENT, VELOCITY, AND TRAJECTORIES OF KEYWORD REPRESENTATIONS IN THE VECTOR SPACE OF THE LANGUAGE MODEL <i>M. M. Charnine and N. V. Somin</i>	60
METAHEURISTIC OPTIMIZATION ALGORITHMS OF PARAMETRIC IDENTIFICATION OF A THREE-DIMENSIONAL MODEL FOR REDUCING THE FATIGUE STIFFNESS OF COMPOSITE MATERIALS <i>A. V. Panteleev, N. V. Turbin, I. S. Nadorov, and N. O. Kononov</i>	73
CONSTRUCTION AND ANALYSIS OF MODEL TO PREDICT THE TECHNICAL STATE OF RAILWAY CAR AXLE BOXES USING INTELLIGENT PREDICTIVE ANALYTICS METHODS <i>O. V. Druzhinina, E. R. Korepanov, I. V. Makarenkova, V. V. Maksimova, and A. A. Petrov</i>	92