№4-2024

ISSN 1999-9429

известия юфу
технические науки

- Алгоритмы обработки информации
- Анализ данных и моделирование
- Электроника, нанотехнологии и приборостроение

ИЗВЕСТИЯ ЮФУ. ТЕХНИЧЕСКИЕ НАУКИ
IZVESTIYA SFedU. ENGINEERING SCIENCES

Свидетельство о регистрации средства массовой информации
ПИ № ФС77-28889 от 12.07.2007
Федеральная служба по надзору в сфере массовых коммуникаций, связи
и охраны культурного наследия
Научно-технический и прикладной журнал
Издается с 1995 года, до середины 2007 года под названием «Известия ТРТУ»
Подписной индекс ПС704

№ 4 (240). 2024 г.

Журнал включен в «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».

Учредитель Южный федеральный университет.
Издатель Южный федеральный университет.
Ответственный за выпуск Курейчик В.В.
Технический редактор Ярошевич Н.В.
Оригинал-макет выполнен Ярошевич Н.В.

Дата выхода в свет 13.09. 2024 г. Формат $70\times108\ {}^{1}\!/_{16}$. Бумага офсетная.

Офсетная печать. Усл. печ. л. – 17,9. Уч.-изд. л. – 12,5.
Заказ № 9633. Тираж 250 экз.

Адрес издателя: 344090, г. Ростов-на-Дону, пр. Стачки, 200/1, тел. 8(863)243-41-66.

Адрес типографии: Отпечатано в отделе полиграфической, корпоративной и сувенирной продукции Издательско-полиграфического комплекса КИБИ МЕДИА ЦЕНТРА ЮФУ. 344090, г. Ростов-на-Дону, пр. Стачки, 200/1, тел. 8(863)243-41-66.

Адрес редакции: 347922, г. Таганрог, ул. Чехова, 22, ЮФУ, тел. +7 (928) 909-57-82, e-mail: iborodyanskiy@sfedu.ru, http://izv-tn.tti.sfedu.ru/.

16+ Цена свободная

ISSN 1999-9429 (Print)
ISSN 2311-3103 (Online)

СОДЕРЖАНИЕ

РАЗДЕЛ I. АЛГОРИТМЫ ОБРАБОТКИ ИНФОРМАЦИИ

РАЗДЕЛ II. АНАЛИЗ ДАННЫХ И МОДЕЛИРОВАНИЕ

РАЗДЕЛ III. ЭЛЕКТРОНИКА, НАНОТЕХНОЛОГИИ И ПРИБОРОСТРОЕНИЕ

CONTENT

SECTION I. INFORMATION PROCESSING ALGORITHMS

SECTION II. DATA ANALYSIS AND MODELING

SECTION III. ELECTRONICS, NANOTECHNOLOGY AND INSTRUMENTATION

Раздел I. Алгоритмы обработки информации

УДК 004.912 DOI 10.18522/2311-3103-2024-4-6-14

В.В. Курейчик, П.С. Герасименко

ОСНОВНЫЕ ПОДХОДЫ К ИЗВЛЕЧЕНИЮ ТЕКСТОВОЙ ИНФОРМАЦИИ (ОБЗОР)

Данная статья посвящена обзору известных и современных подходов, методов и алгоритмов полнотекстового поиска. Описана краткая история решения задачи поиска в неструктурированных текстовых данных, её развитие и актуальность. Сформулирована основная задача поиска в текстовых данных. Приведено определение индекса базы данных. В общем виде определена целевая функция поисковой информационной системы и описаны возможные компромиссные вариации её параметров при решения различных прикладных задач. Приведена обобщённая архитектура современной поисковой информационной системы с разделением задачи поиска на две фазы: первичное извлечение релевантных записей и их последующее ранжирование для формирования окончательных результатов поиска. Даны базовые описания основных алгоритмов и методов полнотекстового поиска, таких как: поиск по термам (логический поиск), поиск с помощью деревьев и их разновидностей (B-деревья, UB-деревья, tries), поиск на основе n-грамм (в том числе поиск на основе частотного представления), использование векторной модели пространства (VSM), поиск на основе инвертированного (обратного) индекса, поиск с использованием аппарата нечёткой логики и биоинспирированных методов. Приведены основные достоинства и недостатки этих методов, описана их применимость в различных условиях, а также рассмотрены возможные методы оптимизации поиска текстовых данных для улучшения точности, скорости поиска и эффективности использования ресурсов. Представлены возможные перспективные направления в области решения задачи первичного извлечения информации. Приведены некоторые способы определения сходства текстовых записей для решения задачи ранжирования на основе аппарата нечёткой логики. Затронуты вопросы повышения релевантности первичного извлечения с помощью методов искусственного интеллекта, нейронных сетей, аппарата нечёткой логики и биоинспирированных методов, в частности методы расширения поискового запроса и/или расширения обрабатываемых текстовых записей. Описано влияние граничных условий построения поисковой системы на повышение её эффективности. В заключение статьи подводятся итоги обзора и обсуждаются перспективы дальнейшего развития различных методов полнотекстового поиска.

Полнотекстовый поиск; B-деревья; векторная модель пространства; обратный индекс; n-грамм индексирование; двухфазовый поиск текста; индексы; извлечение информации; ранжирование; нейронные сети; нечёткая логика; биоинспирированные алгоритмы.

V.V. Kureichik, P.S. Gerasimenko

BASIC APPROACHES TO EXTRACTING TEXTUAL INFORMATION (OVERVIEW)

This article is devoted to the review of known and modern approaches, methods and algorithms of full-text search. A brief history of the solution of the problem of search in unstructured text data, its development and relevance are described. The main task of search in text data is formulated. The definition of the database index is given. The target function of the search information system is defined in general terms and possible compromise variations of its parameters when solving various applied problems are described. A generalized architecture of a modern search information system is given with the division of the search problem into two phases: the primary extraction of relevant records and their subsequent ranking to form the final search results. The article provides basic descriptions of the main algorithms and methods of full-text search, such as: search by terms (logical search), search using trees and their varieties (B-trees, UB-trees, tries), search based on n-grams (including search based on frequency representation), use of the vector space model (VSM), search based on an inverted (reverse) index, search using the

apparatus of fuzzy logic and bioinspired methods. The main advantages and disadvantages of these methods are given, their applicability in various conditions is described, and possible methods for optimizing the search for text data to improve the accuracy, speed of search and efficiency of resource use are considered. Possible promising directions in the field of solving the problem of primary information extraction are presented. Some methods for determining the similarity of text records for solving the ranking problem based on the apparatus of fuzzy logic are given. The article touches upon the issues of increasing the relevance of primary extraction using artificial intelligence methods, neural networks, fuzzy logic and bioinspired methods, in particular methods for expanding the search query and/or expanding the processed text records. The influence of the boundary conditions of the search system construction on increasing its efficiency is described. In conclusion, the article summarizes the review and discusses the prospects for further development of various full-text search methods.

Full-text search; B-trees; vector space model; inverse index; n-gram indexing; two-phase text search; indexes; information extraction; ranking; neural networks; fuzzy logic; binned algorithms.

Введение. Поиск похожей текстовой информации в виде отдельных документов или записей в базах данных является актуальной задачей и представляет всё больший интерес по мере увеличения объёмов хранимых и постоянно генерируемых данных. В данной области ведутся интенсивные исследования, направленные как на оптимизацию и усовершенствование уже хорошо зарекомендовавших себя решений, так и на поиск новых подходов и алгоритмов.

Основная задача систем извлечения информации – определить и получить ту информацию, которая наилучшим образом связана с запросом пользователя. Поскольку релевантными могут быть несколько записей, результаты часто ранжируются в соответствии с их оценкой релевантности запросу пользователя.

Решение этой задачи развивалось от традиционного поиска (когда к тексту относятся как к набору символов, букв или строк различной длины) к осуществлению поиска по смыслу и до современных систем семантического поиска [1] (направленных на сопоставление смыслов текста и поискового запроса, расширенного контекстом самого пользователя). Отдельно выделим гипотезу о статистической семантике (statistical semantics hypothesis) [2]: статистические зависимости употребления слов человеком могут быть использованы для нахождения заложенного в них смысла, так как она лежит в основе ряда широко используемых алгоритмов (например, VSM или TF-IDF).

В конечном счёте большинство подходов к поиску по текстовым данным сводится к преобразованию исходного текста к какой-либо форме, которая позволит вычислить меру близости двух записей или документов. Лучше всего это заметно на попытках представить текст в виде вектора, где существует целый ряд правил представления текста в виде вектора, при этом важно, чтобы близкие по смыслу и содержанию документы преобразовывались в близкие векторы [3]. В таких подходах постоянно предлагают новые методы приведения текста к какой-либо другой, удобной для вычислений, форме и сами способы вычисления меры схожести этих представлений.

Основная цель информационного поиска – предоставить конечному пользователю информацию наиболее релевантную его запросу, поэтому любая задача неточного поиска подразумевает ранжирование первичных результатов по релевантности. Как правило, критерий релевантности задаётся отдельно. Некоторые исследования [4] показывают перспективность глубокого разделения поиска кандидатов для ранжирования и самого ранжирования, так если мы можем выделить небольшую группу из n кандидатов (где n существенно меньше, чем общее число N всех объектов поиска) с использованием максимально простого и эффективного алгоритма, то задача ранжирования выполняется уже на этой небольшой группе, что открывает перспективы для применения более ресурсоемких, но более точных алгоритмов. Обобщённая схема поисковой информационной системы с разделением фазы первичного отбора и ранжирования представлена на рис. 1.

Рассмотрим более подробно некоторые подходы к организации поиска в текстовых данных. Отметим, что большинство из них может быть использовано как для решения задачи первичного извлечения, так и ранжирования данных.

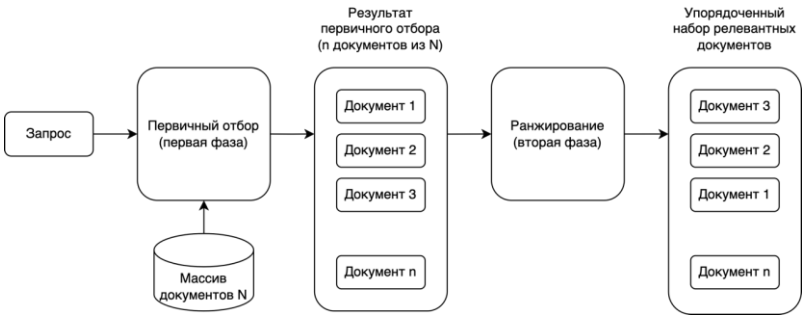


Рис. 1. Архитектура поисковой информационной системы

1. Анализ методов извлечение информации. Оценить качество системы извлечения информации можно с помощью целевой функции от трёх переменных: размер представления текстовой записи в памяти *s*, качество метода определения сходства двух (или более) записей *q* и время, затрачиваемое на работу, *t*.

$$F(s,q,t) = \begin{cases} S(s) \to min \\ Q(q) \to max \\ T(t) \to min. \end{cases}$$

Заметим, что время работы также можно разделить на составляющие: добавление, обновление, удаление, извлечение записей. При этом в идеале поисковая система работает с минимально возможным представлением *s*, которое находит наиболее похожие записи за минимальное время *t*. Тем не менее, для каждой конкретной практической задачи на переменные целевой функции *F* вводятся дополнительные ограничения. Например, для компактных устройств с ограничениями по памяти её размер может быть существенно ограничен. Однако для систем нахождения максимально релевантной (ценной) информации на первый план выходит максимизация качества определения сходства записей *q*. При этом для коммерческих решений важным является минимальное время нахождения необходимых записей *t*.

Совокупность функций преобразования и сравнения используется для построения индекса данных – компактного и удобного для обработки представление массива текстовых записей [5]. Индекс в большинстве случаев существенно меньше по размеру, чем исходные данные и существенно эффективнее с точки зрения обработки.

Создание индексов для полнотекстового поиска в базах данных является ключевым моментом для оптимизации производительности поиска. Рассмотрим достоинства и недостатки основных подходов к поиску в текстовых данных.

1.1. Обычный поиск по термам. Классический поиск по термам, также известный как логический поиск, представляет собой традиционный подход к поиску информации, который сопоставляет содержимое запроса содержимому документов [5, 6]. Это простой и легкий в реализации метод, но имеет некоторые ограничения, такие как неспособность обрабатывать синонимы, многозначность и контекст. Несмотря на эти ограничения, он широко используется в практике, особенно в небольших системах поиска.

1.2. Деревья и их разновидности. Использование различных вариантов деревьев (не всегда похожих на классические деревья, именно поэтому они называются в иностранной литературе *tries*) применяется в задачах поиска текста достаточно давно [5]. Безусловно современные деревья значительно сложнее и эффективнее, чем изначально предложенные B-деревья [7, 8]. Под деревом понимается такая структура данных, которая позволяет извлекать данные посредством весьма короткого пути, построенного на основе поискового запроса.

При использовании деревьев поиск данных происходит очень быстро и практически не зависит от размера набора данных, поскольку индекс и, следовательно, время поиска имеет логарифмическую сложность. Основание этого логарифма очень велико, по мень-

шей мере 1000 для современных архитектур хранения данных. Также дерево легко адаптируется к изменениям в хранимых данных, поскольку оно представляет собой самоорганизующуюся структуру, которая реорганизуется при каждой вставке или удалении данных. Таким образом, это позволяет осуществлять постоянную непрерывную обработку без перерывов на реорганизацию.

Один из современных вариантов реализации поиска с помощью деревьев представлен в [8]. Он демонстрирует высокие показатели по скорости поиска и обновлению самого дерева для данных, которые умещаются в оперативной памяти. Также современные деревья предлагают поиск не чувствительный к незначительным опечаткам/ошибкам в пользовательском запросе.

К основным недостаткам деревьев относят чрезмерное использование памяти и сложности с поддержанием оптимальной структуры дерева при интенсивных операциях вставки/обновления/удаления записей.

1.3. Инвертированное индексирование (обратный индекс). Инвертированный индекс – это структура данных, используемая в извлечении информации для эффективного и быстрого нахождения документов, содержащих определенные слова или термины. Идея инвертированных индексов для поиска была предложена в работе [9] и с тех пор этот подход активно используется в поисковых системах и постоянно модифицируется [10]. Идея инвертированного индекса состоит в следующем: для каждого слова/терма есть список документов, в которых это слово/терм используется. Он состоит из двух основных частей: словаря и списка обратных ссылок.

В словаре каждому уникальному слову или термину из всей коллекции документов присваивается уникальный идентификатор или токен. Этот словарь хранит отображение между словами и списками документов, в которых они встречаются.

Для каждого термина создается список обратных ссылок, который содержит идентификаторы (или другую информацию) всех документов, где встречается данный термин. Эти списки обеспечивают прямой доступ к документам, содержащим интересующие слова или термины.

Когда пользователь вводит поисковый запрос, система использует инвертированный индекс для быстрого определения документов, содержащих все или часть терминов из запроса. Это позволяет значительно ускорить процесс поиска и повысить его качество, особенно при работе с большими объемами текстовой информации.

Этот вид индекса широко используется в популярных поисковых системах, таких как Sphinx, Manticore, Lucene (ElasticSearch). Это широко используемый, эффективный и масштабируемый для больших объёмов данных подход. К основным недостаткам инвертированных индексов относят интенсивное использование памяти, обслуживание и обновление при этом являются ресурсоёмкими. Этот вид индекса хорошо работает для поиска точных совпадений, но для поиска по сложным фразам необходимо применять специальные техники.

1.4. Индексирование на основе n-грамм. Индексирование на основе n-грамм – это метод, используемый в информационном поиске и обработке естественного языка для эффективного поиска и извлечения текстовых документов на основе подстрок [11]. Термин «n-грамма» относится к последовательности из n элементов обрабатываемого текста. В контексте индексации эти «элементы» обычно представляют собой отдельные слова. Поиск на основе n-грамм состоит из четырёх этапов [12]:

1. Токенизация текста, разбивка его на отдельные слова или токены. В этом процессе как правило удаляются знаки пунктуации, слова приводятся к одному регистру, удаляются стоп слова и обрабатываются специальные случаи, такие как сокращения и слова с дефисами.

2. Генерация n-грамм. Текст разделяется на n-граммы. N-грамма – это последовательность из n токенов, если $n = 1$, то n-граммы – это отдельные слова, если $n = 2$, то n-граммы – пары последовательных слов, и так далее. Например, для предложения «Быстрый поиск текстовых данных», при $n = 2$, 2-граммы будут «Быстрый поиск», «поиск текстовых», «текстовых данных».

3. Индексация. Каждая n-грамма индексируется вместе с документами, в которых она встречается. Это может быть сделано с использованием структур данных, таких как хеш-таблицы или инвертированные индексы. Индекс присваивает каждой n-грамме идентификаторы всех документов, которые её содержат.

4. Обработка запроса. Когда пользователь вводит запрос, применяется та же самая токенизация и генерация n-грамм (1-3 этапы). После этого происходит поиск в индексе документов, содержащих n-граммы из запроса, и получается список всех документов, которые содержат предложенные набор *n*-грамм.

Индексация *n*-грамм имеет преимущество в точности перед поиском отдельных слов, так как позволяет работать со связками слов. Однако индексация n-грамм также имеет некоторые ограничения, такие как увеличенные требования к хранению и возможность появления шума от часто встречающихся n-грамм.

Эти подходы постоянно подвергаются модификациям и улучшениям [11, 12].

Ещё одним развитием этого метода индексирования является преобразование исходного текста к «частотному» виду [13]. Модификация метода состоит в разделении текста на n-граммы, для построения «спектрограммы» текстового документа с помощью преобразования Фурье.

1.5. Векторная модель пространства (Vector Space Model). Модель векторного пространства (Vector Space Model, VSM) – это математическая модель, используемая для информационного поиска в области обработки естественного языка и текстового анализа [3]. Она представляет документы и запросы в виде векторов в многомерном пространстве, где каждое измерение соответствует термину в словаре. Обобщённо процесс извлечения информации с помощью векторной модели пространства можно представить в виде следующих этапов [14]:

1. Создание словаря слов на основе документов. Для этого используют токенизацию текста, удаление стоп слов, приводят слова к базовым формам.

2. Каждый документ и поисковый запрос представляются в виде вектора в векторном пространстве. Размерности этих векторов соответствуют количеству терминов в словаре, а значение каждой размерности представляет собой некоторую меру важности или частоты соответствующего термина в документе или запросе. Перед построением векторов обычно применяются схемы взвешивания терминов к матрице документ-термин. Одна из популярных схем – TF-IDF (частота термина-обратная частота документа) [14], которая придает больший вес терминам, часто встречающимся внутри документа, но редко встречающимся во всем массиве документов. В [4, 15] приведены другие подходы к приведению текста к векторам.

3. После того как документы и запросы представлены в виде векторов, следующим шагом является расчет сходства между ними. Самая часто используемая метрика сходства – косинусное сходство, которое измеряет косинус угла между двумя векторами [3]. В [16, 17] приведены другие подходы к определению меры сходства.

4. Чтобы извлечь соответствующие документы для данного запроса, сходство между вектором запроса и каждым вектором документа вычисляется с использованием меры сходства. В результате мы получаем набор ранжированных документов, наиболее соответствующих запросу.

В целом, модель векторного пространства предоставляет гибкий подход для представления текстовых данных и извлечения соответствующей информации на основе мер сходства. Она широко используется в поисковых системах и для кластеризации документов. К недостаткам относят сложности с лексической неоднозначностью и вариативностью (например, омонимы всегда будут иметь ненулевое сходство, а документы со схожим значением, но с разным словарем терминов не будут связаны), также возникают сложности с выделением слов, специфичных для предметной области, так как они существенно реже встречаются в тексте и поэтому им приходится увеличивать вес при сравнении с помощью дополнительных методов.

1.6. Использование нейросетей, глубокое обучение. Широкое применение для решения задач семантического поиска в последнее время приобрели различные виды нейронных сетей [18–20]. Они показывают очень хорошие результаты в области поиска похожих по смыслу текстов даже с минимальными пересечениями на уровне отдельных слов. Зачастую нейросети используют как сложный классификатор, который позволяет отнести пользовательский запрос к какому-либо классу и сравнить его с классом интересующего нас документа/записи. Кроме того, с помощью нейронных сетей можно расширить поисковый запрос и/или добавляемую в хранилище запись [21–23], что позволит включить в результаты поиска в том числе записи, которые не содержат слова поискового запроса, но отвечают его смыслу.

Одним из подходов к семантическому поиску является представление текстовых документов и запросов в непрерывном векторном пространстве (аналогично векторной модели, описанной выше) с помощью свёрточных (англ. convolutional neural network, CNN) и рекурентных нейронных сетей (англ. recurrent neural network, RNN) [23].

Кроме того, механизмы, основанные на внимании, такие как архитектура трансформер [24], используются для улучшения способности систем выделять и обрабатывать важные части как самого запроса, так и документов. В ряде работ [25] было показано, что использование предварительно обученных языковых моделей, таких как BERT и GPT-4, повышает производительность систем извлечения информации, обеспечивая лучшее понимание семантики и контекста на естественном языке.

Обработка текста с помощью нейронных сетей является одним из самых перспективных направлений семантического сравнения. Однако, одним из недостатков применения нейронных сетей является необходимость их обучения, то есть решения дополнительных весьма непростых задач по подбору обучающих данных и самой стратегии обучения.

1.7. Нечёткий поиск. Нечётка логика может быть применена в системах информационного поиска для работы с неточными или двусмысленными запросами и улучшения релевантности результатов поиска [26]. Большой интерес представляет использование нечёткой логики для определения меры сходства документов независимо от формы их представления (вектор, n-граммы, инвертированные индексы). Введение нечёткой логики позволяет улучшить ранжирование результатов поиска, учитывая степень схожести между терминами запроса и содержимым документа. Вместо бинарных совпадений (точное совпадение или его отсутствие), использование нечёткой логики позволяет получать частичные совпадения, где документы, содержащие термины, сходные с запросом, получают более высокие оценки релевантности, что позволяет извлекать документы, тесно связанные с информационными потребностями пользователя, но не обязательно точно соответствующие терминам запроса [27]. Также в [27] предлагается заменить метрики требованием к нечеткому сходству, удовлетворяющему свойству транзитивности с настраиваемым нечетким конъюнктором.

Также нечёткую логику используют для расширения запросов, учитывая синонимы, родственные термины или даже приблизительные совпадения с исходными терминами запроса [28]. Это особенно эффективно при работе с полисемией или синонимией, так как позволяет извлекать семантически связанные документы, не содержащие точных терминов, указанных в поисковом запросе.

Отметим, что данный подход позволяет совмещать нескольких источников доказательств сходства документов, таких как текстовое содержание, метаданные, контекстная информация и любые другие. Агрегируя доказательства из разных источников с использованием нечётких операторов, система может принимать более обоснованные решения о релевантности документов по отношению к заданному запросу, что потенциально приводит к более точным и всесторонним результатам поиска.

1.8. Биоинспирированные алгоритмы. Существует ряд работ, посвящённых использованию биоинспирированных методик для решения задач извлечения информации. Так, например, генетические алгоритмы используют для поиска наилучшего набора документов, связанных с ключевыми словами из запроса пользователя. В [29] приведён обзор применения различных биоинспирированных методов для решения задач классификации и кластеризации текстовых данных.

Ещё одним направлением использования биоинспирированных методов является расширение (оптимизация) запросов [30] – добавление, удаление или изменение исходных запросов. Они применяются для устранения неточностей в системах поиска информации, которые возникают из первоначального запроса, предоставленного пользователями. Тем не менее, в некоторых работах отмечается отсутствие значимых улучшений показателей поиска при использовании генетических алгоритмов для расширения поисковых индексов.

2. Направления оптимизации поиска и улучшения его качества. В ряде случаев разбиение задачи поиска на две фазы: поиск документов для ранжирования и непосредственно ранжирование, позволяет повысить эффективность поисковых систем. Основной задачей первой фазы является нахождение максимально полного подмножества записей релевантных запросу. При этом основной задачей второй фазы является максимально точное ранжирование результатов первой фазы. Безусловно, у этого подхода есть недостаток, так как чрезмерное упрощение предварительного поиска может привести к исключению из выдачи записей, которые могут быть близкими по смыслу, но существенно различаться по лексическому составу.

Актуальным является поиск новых форм и способов приведения текстовых документов к максимально удобному для поиска и извлечения виду. Также по мере развития аппаратной базы совершенствуются структуры данных, используемых для решения задач извлечения информации.

Также для улучшения качества поиска применяют различные техники расширения как самих анализируемых текстов, так и запросов пользователя. Под расширением понимается добавление новых слов в исходные данные по определённым правилам. Эти слова не встречаются в исходном тексте или запросе, но являются схожими, синонимичными или релевантными по другим критериям. Такие подходы позволяют находить релевантные записи по смыслу даже если они не содержат слова из запроса пользователя.

Кроме того, на этапе ранжирования сравнительно небольшой группы документов применяются алгоритмы машинного обучения, нейронные сети, учитывающие предпочтения конкретного пользователя поисковой системы и конкретную практическую задачу её владельца.

При некоторых граничных условиях, накладываемых на предметную область, добиваются существенных преимуществ с точки зрения быстродействия и размеров индексов. Яркий пример – разграничение двух видов поиска: поиск в документах (как правило достаточно больших текстовых массивах) и поиск в записях баз данных (в большинстве случаев текстовые данные в них представлены относительно короткими фрагментами). Если строить индекс исходя из работы только с небольшими фрагментами текста, то требования к нему снижаются, что позволяет получить существенный выигрыш в быстродействии.

Заключение. В данной работе представлен обзор и анализ современных моделей, методов и алгоритмов полнотекстового поиска. Рассмотрена обобщённая архитектура поисковой информационной системы, предложена целевая функция. В результате проведенного обзора и анализа выявлены основные достоинства и недостатки этих методов. Представлены возможные перспективные направления в области первичного извлечение информации. Предложены направления для повышения релевантности первичного извлечения на основе методов искусственного интеллекта и нейронных сетей, а также методы расширения поискового запроса и/или расширения обрабатываемых текстовых записей.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Guha R.V., McCool R., Miller E.* Semantic search, *Proceedings of the 12th International Conference on World Wide Web, Budapest,* 2003, pp. 700-709.
2. *Weaver W.* Machine Translation of Languages. MIT Press, Cambridge, MA, Reprinted from a memorandum written by Weaver in 1949, 1955.
3. *Salton G., Wong A., Yang C.S.* A vector space model for automatic indexing, *Communications of the ACM,* 1975, Vol. 18, No. 11, pp. 613-620.
4. *Rygl J., Pomikálek J., Řehůřek R., Růžička M.* Semantic Vector Encoding and Similarity Search Using Fulltext Search Engines, *Proceedings of the 2nd Workshop on Representation Learning for NLP*, 2017.

5. *Baeza-Yates R., Ribeiro-Neto B.* Modern Information Retrieval. ACM Press New York, 1999.
6. *Manning C.D., Raghavan P., Sch"utze, H.* Introduction to Information Retrieval. Cambridge University Press, 2008.
7. *Yan W., Zhang X.* A Concise Concurrent B + -Tree for Persistent Memory, *ACM Transactions on Architecture and Code Optimization,* 2023, Vol. 21 (2).
8. *Leis V., Kemper A., Neumann T.* The adaptive radix tree: ARTful indexing for main-memory databases, *2013 IEEE 29th International Conference on Data Engineering (ICDE), Brisbane, QLD, Australia,* 2013, pp. 38-49.
9. *Salton G.* Information Retrieval: Data Structures and Algorithms. Reading, Massachussetts: Addison-Wesley, 1989.
10. *Giulio Ermanno P., Rossano V.* Techniques for Inverted Index Compression, *ACM Computing Surveys,* 2020, Vol. 53, pp. 1-36.
11. *Miller E., Shen D., Liu J., Nicholas C.* Performance and Scalability of a Large-Scale N-gram Based Information Retrieval System, *Journal of Digital Information,* 2000, Vol 1, No 5.
12. *Srinivasa K., Devi B.* GPU Based N-Gram String Matching Algorithm with Score Table Approach for String Searching in Many Documents, *Journal of The Institution of Engineers (India) Series B,* 2017, Vol. 98, pp. 467-476.
13. *Kudryavtsev K.* Search in text documents based on N-grams and Fourier window transformation, *Computer Science,* 2014, Vol. 65, pp. 871-880.
14. *Jones K.* A Statistical Interpretation of Term Specificity in Retrieval, *Journal of Documentation,* 2004, Vol. 60, pp. 493-502.
15. *Chebil W., Soualmia L.* Improving semantic information retrieval by combining possibilistic networks, vector space model and pseudo-relevance feedback, *Journal of Information Science,* 2023.
16. *Eminagaoglu M.* A new similarity measure for vector space models in text classification and information retrieval, *Journal of Information Science,* 2020, Vol. 48.
17. *Singh A., Yadav A., Rana A.* K-means with Three different Distance Metrics, *International Journal of Computer Applications,* 2013, Vol. 67, pp. 13-17.
18. *Nogueira R., Yang W., Lin J., Cho K.* Document expansion by query prediction, *arXiv preprint arXiv:1904.08375,* 2019.
19. *Mao Y., He P., Liu X., Shen Y., Gao J., Han J., Chen W.* Generation-augmented retrieval for open-domain question answering, *arXiv preprint arXiv:2009.08553,* 2020.
20. *Yan M., Li C., Bi B., Wang W., Huang S.* A unified pretraining framework for passage ranking and expansion, *Proceedings of the AAAI Conference on Artificial Intelligence,* 2021, Vol. 35, pp. 4555-4563.
21. *Efron M., Organisciak P., Fenlon K.* Improving retrieval of short texts through document expansion, *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval,* 2012, pp. 911–920.
22. *Agirre E., Arregi X., Otegi A.* Document expansion based on WordNet for robust IR, *Coling 2010: Posters,* 2010, pp. 9-17.
23. *Shen Y., He X., Gao J., Deng L.* Latent semantic models with deep neural networks for information retrieval, *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval. ACM,* 2014, pp. 269-278.
24. *Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I.* Attention Is All You Need, *Advances in Neural Information Processing Systems,* 2017, pp. 6000-6010.
25. *Devlin J., Chang M.W., Lee K., Toutanova K.* BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers),* 2019, pp. 4171–4186.
26. *Chimah J., Ude F.* Current trends in information retrieval systems: review of fuzzy set theory and fuzzy Boolean retrieval models, *Journal of Library Services and Technologies,* 2020, Vol. 2, pp. 48-56.
27. *Bartos T., Eckhardt A., Skopal T.* Fuzzy approach to non-metric similarity indexing, *Proceedings - 4th International Conference on SImilarity Search and APplications, SISAP 2011,* 2011, pp. 115-116.
28. *Mounira C., Jedidi A., Gargouri F.* Semantic/Fuzzy Information Retrieval System, *International Journal of Information Technology and Web Engineering,* 2017.
29. *Abualigah L., Hanandeh E.* Applying Genetic Algorithms to Information Retrieval Using Vector Space Model, *International Journal of Computer Science, Engineering and Applications,* 2015, Vol. 5, pp. 19-28.
30. *Russell A.* A Genetic Algorithm for Query Optimization, *Department of Computer and Information Sciences University of Strathclyde, Glasgow August 26,* 2019.

Статью рекомендовала к опубликованию д.т.н., профессор Л.С. Лисицына.

Курейчик Владимир Викторович – Южный федеральный университет; e-mail: vkur@sfedu.ru; г. Таганрог, Россия; тел.: 88634371651; кафедра систем автоматизированного проектирования им. В.М. Курейчика; зав. кафедрой; д.т.н.; профессор.

Герасименко Петр Сергеевич – e-mail: pege@sfedu.ru; тел.: 88634371651; кафедра систем автоматизированного проектирования им. В.М. Курейчика; аспирант.

Kureichik Vladimir Victorovich – Southern Federal University; e-mail: vkur@sfedu.ru; Taganrog, Russia; phone: +78634371651; the department of Computer Aided Design; head of department; dr. of eng. sc.; professor.

Gerasimenko Petr Sergeevich – e-mail: pege@sfedu.ru; phone: +78634371651; the department of Computer Aided Design; post graduate student.

УДК 004.896 DOI 10.18522/2311-3103-2024-4-14-30

Б.К. Лебедев, О.Б. Лебедев, М.А. Ганжур

БИОИНСПИРИРОВАННЫЙ АЛГОРИТМ ПЛОТНОЙ УПАКОВКИ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ РАСКРОЯ ПОЛУОГРАНИЧЕННОЙ ПОЛОСЫ

Предлагается архитектура и методология раскроя-упаковки полуограниченной полосы на основе методов биоинспирированного поиска. В основе подхода к декомпозиции обшей задачи упаковки и методологии формированию карт раскроя лежат эвристики уровневого подхода к упаковке полосы. Архитектура сформирована на основе декомпозиции общей задачи и включает 5 основных секций: управление процессом поиска; формирования блоков; формирование контейнеров; компакция контейнеров; заполнения полосы контейнерами. Упаковка ориентирована на двухуровневый раскрой полосы. На первом уровне путем гильотинного разреза выполняется раскрой на контейнеры. На втором уровне два варианта раскроя: путем гильотинного или путем не гильотинного разреза выполняется раскрой контейнеров на детали (элементы прямоугольной формы). Упаковка выполняется путем последовательного заполнения уровней полосы контейнерами. В основу методологии раскроя-упаковки полуограниченной полосы положен иерархический подход снизу вверх. Задача, решаемая на первом уровне иерархии, заключается в формировании множества блоков B одинаковой ширины на базе исходного набора A прямоугольников, включаемых в блоки. Для решения поставленной задачи авторами разработан биоинспирированный алгоритм одномерной упаковки элементов в одинаковые блоки. На втором уровне иерархии решается задача распределения блоков по контейнерам. Все контейнеры и блоки имеют одинаковую ширину D, равную ширине полосы. В каждом контейнере помещаются два блока. Задача распределения блоков по контейнерам сведена к задаче нахождения максимального паросочетания минимальной стоимости. В отличие от канонической метаэвристики муравьиного алгоритма в работе агентом на графе поиска решений строится максимальная клика, которая является интерпретацией решения. На третьем уровне иерархии решается задача компакции контейнеров. Процесс распределения блоков по контейнерам сопровождается процедурой сжатия каждой пары блоков, назначаемых в один контейнер. Целью компакции является минимизация общей площади контейнера путем плотного размещении блоков. Компакцию последовательно проводят во всех контейнерах. На четвертом уровне иерархии решается задача заполнения полосы контейнерами. В качестве модели для представления решения на графе поиска решений служит клика. Разработана база данных коллективной эволюционной памяти. Разработана методика формирования феромоновых точек и структур данных коллективной эволюционной памяти. Для проведения объективных экспериментов были использованы известные тестовые задачи, представленные в литературе и сети Интернет. По сравнению с существующими алгоритмами достигнуто улучшение результатов на 3-5%. Временная сложность алгоритма, полученная экспериментальным путем, практически совпадает с теоретическими исследованиями и для рассмотренных тестовых задач составляет (BCA ≈ $O(n^2)$).

Раскрой-упаковка; полуограниченная полоса; декомпозиция; методология; задача об упаковке в контейнеры; компакция; поисковая оптимизация; роевой интеллект; муравьиная колония; адаптивное поведение.

B.K. Lebedev, O.B. Lebedev, M.A. Ganzhur

BIO-INSPIRED DENSE PACKING ALGORITHM TO INCREASE THE EFFICIENCY OF SEMI-LIMITED STRIP CUTTING

A methodology has been developed for finding solutions to the semi-infinite strip packing problem based on models of adaptive behavior of biological systems. To reduce the overall labor intensity of the search procedure, an approach based on decomposition of the problem being solved is proposed. The packaging is designed for cutting by guillotine cutting of the tape into containers and non-guillotine cutting of containers into elements. Packaging is carried out by sequentially filling the strip with containers. The problem of packing rectangles into strips is solved in three stages. At the first stage, the agent solves the problem of distributing a set A of rectangular-shaped elements in a set of blocks B. The problem of forming a set of blocks B, including sets of rectangular-shaped elements A, is solved by an algorithm for one-dimensional packing of elements into identical blocks. At the second stage, the problem of distributing blocks among containers is solved. All containers have the same width D, equal to the width of the strip. Each container holds two blocks. The process of distributing blocks into containers is accompanied by a compaction procedure for each pair of blocks assigned to one container. The purpose of compaction is to minimize the total area of the container by densely placing the blocks. Compaction is carried out sequentially in all containers. The problem of distributing blocks into containers is reduced to the problem of finding the maximum matching of the minimum cost. In contrast to the canonical paradigm of the ant algorithm, when working as an agent, a clique is built on the solution search graph, on the edges of which a pheromone is deposited. A technique has been developed for the formation of pheromone points and data structures of collective evolutionary memory. To conduct objective experiments, well-known test problems presented in the literature and on the Internet were used. Compared to existing algorithms, a 3-5% improvement in results was achieved. The time complexity of the algorithm, obtained experimentally, practically coincides with theoretical studies and for the considered test problems is ≈ $O(n^2)$.

Packing in an endless strip; decomposition; distribution of blocks into containers; compaction; swarm intelligence; ant colony; adaptive behavior; search engine optimization.

Введение. В работе рассматривается задача упаковки полуограниченной полосы (*1.5 Dimensional Bin Packing, 1.5 DBP*). Задачи раскроя-упаковки (*Bin Packing*) занимают важное место в современной комбинаторной оптимизации и привлекают внимание многих ученых, как в России, так и за рубежом. Интерес к задачам раскроя-упаковки объясняется, в частности, их большой практической значимостью [1–3]. Как правило, приложения задач раскроя-упаковки относятся к материалоемким производствам, где одним из основных факторов снижения себестоимости выпускаемой продукции является рациональное использование ресурсов [4, 5].

Задачи раскроя-упаковки (ЗРУ) относятся к классу *NP*–трудных задач комбинаторной оптимизации. Для решения задачи раскроя-упаковки российскими и зарубежными учеными разработано большое количество методов и алгоритмов, различающихся как областью применения, так и эксплуатационными свойствами. Оптимальное решение задач раскроя, как правило, дает высокий экономический эффект. Методы решения ЗРУ принято разделять на две большие группы: точные и приближенные [6–8]. Поиск эффективных методов решения *NP*-полных задач является предметом исследования многих специалистов различных областей науки и практики [9, 10].

На сегодняшний день для решения задач раскроя-упаковки используются различные оптимизационные алгоритмы [11, 12]. Во многих работах уделяется большое внимание разработке приближенных и эвристических методов. В связи с тем, что точные методы, основанные на полном переборе решений, нереализуемы при большой размерности данных, используют более эффективные методы сокращенного перебора.

Одним из перспективных и интенсивно развивающихся направлений в этой области является использование бионических методов и алгоритмов [13, 14]. Результатом непрекращающегося поиска наиболее эффективных методов упаковки стало использование методов биоинспирированного поиска [15–17].

В работе для уменьшения общей трудоемкости поисковой процедуры предлагается подход, основанный на декомпозиции решаемой задачи. Постулат декомпозиции применяется для снижения уровня сложности задачи упаковки путем расчленения общей задачи на подзадачи [18].

Для решения задачи прямоугольной упаковки в полуограниченную полосы полосу предложен новый гибридный подход на основе декомпозиции общей задачи упаковки и построения поисковой процедуры на основе модифицированного метода муравьиной колонии [19]. Для оптимизации раскроя путем минимизации общей площади полосы в общий маршрут упаковки включаются процедуры компакции (сжатия). Предложена новая методология и метод распределения блоков по контейнерам, являющийся ядром решения задачи упаковки. Разработаны поисковые методы для решения задач гильотинного и не гильотинного прямоугольного раскроя.

Постановка задачи упаковки. Основная задача методологии проектирования состоит в декомпозиции задачи с целью получения последовательности действий, наилучшим образом организующей процесс проектирования. При выборе маршрута решения задачи упаковки, решающее значение имеет связь этапов проектирования в единый процесс. В соответствии с принципом модульности и иерархичности (многоуровневости), работы по упаковке полосы делятся по этапам таким образом, что решения задач, получаемых на одних этапах, служат исходными данными для задач следующих этапов.

Необходимо: произвести упаковку прямоугольников в полуограниченную полосу шириной D не выходят за границы полосы, без перекрытия друг друга с минимизацией занимаемой площади.

Дано: полуограниченная полоса P; множество прямоугольников $A=\{a_i|i=1,2,\ldots,n\}$, размеры которых задаются: размеры прямоугольников вектором $W=\{w_i|i=1,2,\ldots,n\}$ ширин и вектором $H=\{h_i|i=1,2,\ldots,n\}$ высот прямоугольников.

Требуется найти такую упаковку полосы, для которой высота H^P ее занятой части имеет минимальное значение.

Упаковка полосы осуществляется путем последовательного заполнения контейнерами (рис. 1).

В основе метода декомпозиции обшей задачи упаковки и методологии формирования карт раскроя лежат эвристики уровневого подхода к упаковке полосы [3–6]. Уровневые алгоритмы обладают лучшими возможностями улучшения и коррекции решения. Упаковка выполняется путем последовательного заполнения уровней полосы контейнерами.

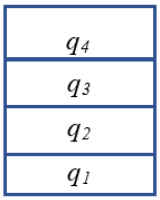


Рис. 1. Упаковка полосы осуществляется путем последовательного заполнения

Архитектура, сформированная на основе декомпозиции общей задачи, включает 5 основных секций: управление процессом поиска; формирования блоков; формирование контейнеров; компакция контейнеров; заполнения полосы контейнерами. В состав стандартого маршрута поиска решения дополнительно включена процедура распределения блоков между контейнерами и процедура компакции контейнера. Целью добавленных процедур является минимизация общей площади контейнера путем плотного размещении блоков. Это позволило улучшить характеристики решений, что подтверждается результатами экспериментов. У всех блоков и контейнеров одинаковая ширина, равная ширине полосы – D.

В соответствии с иерархическим подходом к упаковке первой решается задача формирования множества $B=\{b_j|j=1,2,\ldots,n_b\}$ одинаковых по ширине блоков. Элементы в блоках располагаются на горизонтальном основании внутри контейнера вплотную друг к другу.

Рис. 2. Множество блоков

Задача распределения множества прямоугольников по блокам сводится к задаче разбиения множества $A=\{A_j|j=1,2,...,m\}$ на n_b непустых и непересекающихся подмножеств A_j. $A=\cup A_j$, $A_j\neq\varnothing$, $(\forall i,j)[A_i\cap A_j=\varnothing]$. Суммарная ширина W_j прямоугольников множества A_j., входящих в состав блока b_j, определяется как $W_j=\sum w_i$, для $\{i|a_i\in A_j\}$. Значение W_j ограничено размером ширины контейнера, $W_j\leq D$.

В качестве высоты h^b_j блока b_j служит высота самого высокого элемента в блоке.

Критерий оптимизации – суммарная высота сформированного множества блоков: $F=\sum h^b_j$ для $\{j=1,..,m\}$. Цель оптимизации – минимизация F.

Для решения задачи формирования множества блоков $B=\{b_j|j=1,2,...,n_b\}$, включающих наборы элементов $A=\{a_i|i=1,2,...,n_a\}$, используется разработанный авторами биоинспирированный алгоритм одномерной упаковки элементов в одинаковые блоки [20].

Второй по важности является задача распределения блоков по контейнерам. В каждом контейнере помещаются два блока. Один над другим. Процесс распределения блоков по контейнерам сопровождается процедурой компакции (сжатия) Суть компакции заключается в максимально возможном сближении по вертикали блоков, назначаемых в один контейнер. В результате исследования подходов к компакции была установлена зависимость плотности упаковки полосы от порядка расположения элементов в блоке. В связи с этим в работе используется эвристики предварительного упорядочивания расположения элементов в блоках по возрастанию высоты элементов в нижнем блоке контейнера с начала блока, в верхнем блоке с конца.

Целью компакции является минимизация общей площади контейнера путем плотного размещении блоков. Компакцию последовательно проводят во всех контейнерах. Высота H^q_k.контейнера q_k определяется после компакции. Критерий оптимизации $F=\sum H^q_{k.}$ для $\{k|q_k\in Q\}$. Цель оптимизации – минимизация F.

Размещение блоков в контейнере. В работе [20] рассматриваются два возможных варианта размещения блоков в контейнере, в зависимости от типа разрезов и компакции контейнера.

Первый вариант сближения блоков в контейнере реализуется при использовании одного типа разреза – гильотинного при раскрое полосы на контейнеры и раскрое контейнеров на элементы. В этом случае блок представляется в виде прямоугольника, у которого верхняя и нижняя стороны являются прямыми линиями. Компакция реализуется путем сближения пары блоков b_i и b_j без взаимного перекрытия до соприкосновения границ блоков (рис. 3).

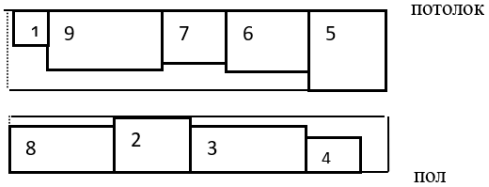


Рис. 3. Первый вариант компакции

В этом случае высота контейнера q_k определится как:

$$H^q_{k.}=(h^b_i + h^b_j), \quad (1)$$

где h^b_i и h^b_j – высота соответственно верхнего и нижнего блоков, входящих в состав контейнера q_k.

Второй вариант сближения блоков в контейнере реализуется при использовании двух типов разреза:

- гильотинного при раскрое полосы на контейнеры и не гильотинного разреза при раскрое контейнеров на элементы.

В этом случае блок с одной стороны ограничен прямой линией, а другая стороны блок ограничен ломанной линией. Блоки сближаются до первого касания границ блоков. (рис. 4).

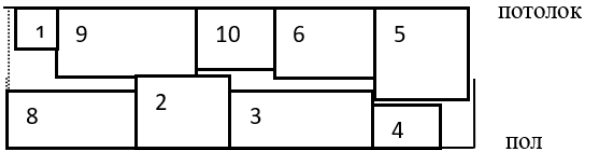


Рис. 4. Сближения блоков с неупорядоченным размещением элементов

Сближение блоков контейнера выполняется до первого касания границ элементов по вертикали. Отметим, что положение элементов друг относительно друга по вертикали в обоих блоках зафиксировано. В результате сближения блоков отдельные прямоугольники верхнего уровня будут соприкасаться с прямоугольниками нижнего уровня. На рис. 4 упорядочивание элементов в блоке не проводилось. Девятый элемент соприкасается с вторым. $H^q_k=(h_9 + h_2)$.

Третий вариант. предусматривает предварительное упорядочивание элементов в блоках: в верхнем блоке элементы расположены по убыванию, а в нижнем блоке по возрастанию высоты прямоугольников.

На рис. 5 соприкасаются второй и десятый элементы. Высота контейнера $H^q_k=(h_{10} + h_2)$. В большинстве случаев такой подход приводит к минимизации высоты контейнера.

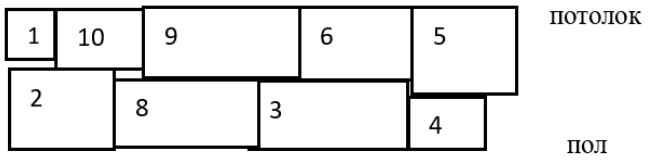


Рис. 5. Вариант сближения упорядоченных блоков

Процедура компакции контейнера. В каждом контейнере помещаются два блока. На рис. 6 представлен контейнер с размещенными в нем блоками. В первый (верхний) блок входят элементы a_1 - a_5, во второй (нижний) блок входят элементы $a_6 – a_{10}$.

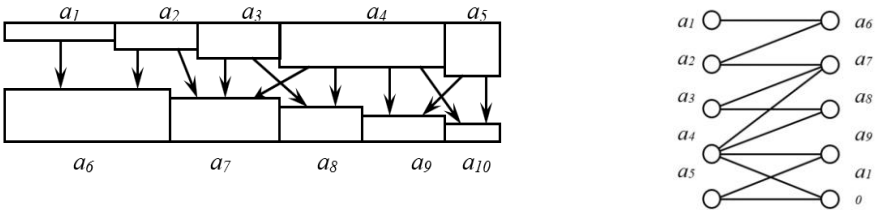


Рис. 6. Формирование графа ограничений

Для формирования графа ограничений разработан алгоритм распространения тени. «Тень» элемента распространяется вдоль направления компакции. От каждого элемента a_i верхнего блока отбрасывается тень на элементы нижнего блока контейнера. Для каждого элемента a_i верхнего блока определяется множество T_i элементов нижнего блока, на которые падает тень элемента a_i верхнего блока. $T_1=\{a_6\}$, $T_2=\{a_6, a_7\}$, $T_3=\{a_7, a_8\}$, $T_4=\{a_7, a_8, a_9, a_{10}\}$, $T_5=\{a_9, a_{10}\}$. На базе a_i и T_i формируется множество пар $P_i=\{p_{ij}\}$ элементов контейнера q_k, находящихся в конфликте, где $p_{ij}=<a_i,(a_j\in T_i)>$. Элемент a_i принадлежит верхнему блоку контейнера, а элемент a_j, лежащий в тени элемента a_i, принадлежит нижнему блоку контейнера. На рис. 6:

$$P_1=\{<a_1, a_6>\}, P_2=\{<a_2, a_6>, <a_2, a_7>\}, P_3=\{<a_3, a_7>, <a_3, a_8>\},$$
$$P_4=\{<a_4, a_7>, <a_4, a_8>, <a_4, a_9>, <a_4, a_{10}>\}, P_5=\{<a_5, a_9>, <a_5, a_{10}>\}.$$

На базе множества $P = \cup P_i$ пар $\underline{p_{ij}}$ формируется взвешенный граф ограничений $G=(A,E)$ каждая вершина $a_i\in A$ соответствует элементу, а ребра отображают ограничения.

Для каждой пары $p_{ij}\in P$ элементов a_i и a_j, находящихся в конфликте друг с другом в контейнере, рассчитывается сумма $h_{ij}=h_i+h_j$ высот элементов, входящих в состав пары p_{ij}. Ребру, связывающему вершины a_i и a_j, присваивается вес, равный h_{ij}. Отыскивается максимальное значение h^*_k, $(\forall i\,\forall j)[h^*_k\geq h_{ij}]$. Величина h^*_k является минимально возможной высотой h^q_k контейнера q_k. В рассматриваемом примере суммарная высота пары конфликтующих блоков имеет значение: $h_{16}=4$; $h_{26}=4{,}5$; $h_{27}=4$; $h_{37}=4{,}5$; $h_{47}=5$; $h_{48}=4{,}5$; $h_{49}=4{,}5$; $h_{4,10}=3{,}5$; $h_{5,10}=4$.

Максимальное значение имеет $h_{47}=5$. Следовательно значение $h_{47}=5$ будет минимальной высотой H^q_k контейнера q_k.

Подход к решению задачи распределения блоков по контейнерам с компакцией. Алгоритм распределения блоков по контейнерам разработан на основе модифицированных метаэвристик адаптивного поведения муравьиной колонии [5, 12].

Так как в каждом контейнере q_k помещаются два блока, то задача распределения множества $B=\{b_j|j=1,2,\ldots,n_b\}$ блоков по контейнерам сводится к задаче разбиения множества блоков B на множество $P=\{p_k|k=1,2,\ldots,n_k\}$ пар, таких, что каждая пара $p_k=(b_i, b_j)$ блоков размещается в одном контейнере q_k. $(\forall i,j)[p_i\cap p_j=\varnothing]$, $n_k=n_b/2$.

Если число блоков нечетное, то в множество $B=\{b_j|j=1,2,\ldots,n_b\}$ добавляется один пустой блок. Поскольку число блоков становится четным, то формируется $n_b/2$ пар блоков. Распределение блоков по контейнерам осуществляется одновременно с процедурой компакции контейнера.

В качестве пространства поиска решений служит полный граф $G=(X,U)$. Вершинам множества X и соответствует множество блоков $B=\{b_j|j=1,2,\ldots,n_b\}$. $S_1:X\rightarrow B$; $x_j=S_1(b_j)$.

Каждому ребру $u_k\in U$, $u_k=(x_i, x_j)=(S_1(b_i),S_1(b_j))$ графа G, связывающего вершины x_i и x_j, соответствует пара блоков $p_k=(b_i,b_j)$, $b_i=S_1(x_i)$ и $b_j=S_1(x_j)$. Между ребрами u_k, и множеством P пар $p_k=(b_i,b_j)$ блоков $B=\{b_j|j=1,2,\ldots,n_b\}$ установлено соответствие $S_2:U\rightarrow P$; $p_k=S_2(u_k)$.

Вес $\varphi(u_k)$ ребра $u_k=(x_i, x_j)$ графа G, связывающего вершины x_i и x_j, равен высоте H^q_k контейнера q_k в состав которого входит пара блоков $b_i=S_1(x_i)$ и $b_j=S_1(x_j)$. $\varphi(u_k)=H^q_k$.

В случае первого варианта сближения пары блоков в контейнере высота контейнера определяется по формуле (1).

Если реализуется второй вариант сближения пары блоков в контейнере, то предварительно в верхнем блоке прямоугольники располагаются по убыванию, а в нижнем блоке по возрастанию высоты прямоугольников. Высота контейнера рассчитывается после процедуры компакции контейнера.

Задача разбиения множества блоков B на множество пар, таких, что каждая пара $p_k=(b_i, b_j)$ блоков размещается в одном контейнере q_k, сводится к задаче построения паросочетания U_1 в графе $G=(X,U)$. $U_1\subset U$. Основное свойство паросочетания U_1 в графе $G=(X,U)$ заключается в том, что любые два ребра $u_i\in U_1$ и $u_j\in U_1$ не имеют общих вершин, т.е. не смежны. Например: в графе G (рис. 7) паросочетание $U_1=\{u_1,u_4,u_7,u_9\}$.

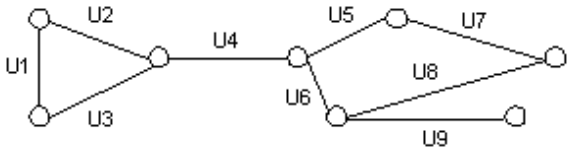


Рис. 7. Пример графа G

Паросочетание максимальной мощности U_1 полного графа поиска решений $G=(X,U)$ у которого $n_b=|X|$ – четное число, включает $n_b/2$ не смежных ребер.

Обозначим как F суммарный вес ребер, входящих в паросочетание U_1.

В ряде прикладных задач граф G трансформируется в реберный граф $G_r=(U,V)$. Вершины графа G_r соответствуют рёбрам графа G. Если в графе G рёбра u, и u_j смежны, т.е. инциденты одной вершине, то вершины u_i и u_j в графе G_r связаны ребром.

На рис. 8. представлен реберный граф G_r графа поиска решений $G=(X,U)$ представленного на рис.7.

Подмножество вершин $P=\{u_1, u_4, u_8\}$ реберного графа G_r (рис. 8) является внутренне устойчивым, т.к. любые две вершины подмножества P не смежны. Паросочетанию $P=\{u_1, u_4, u_7, u_9\}$ графа G соответствует внутренне-устойчивое подмножество $P=\{u_1, u_4, u_7, u_9\}$ реберного графа G_r.

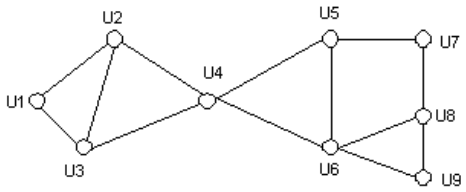


Рис. 8. Реберный граф G_r графа G

Авторами разработан метод и алгоритм последовательного формирования в полном графе $G_r=(U,V)$ максимального внутренне-устойчивого множества вершин $u_k \in U_1$, $U_1 \subset U$, соответствующего паросочетанию V_1 в графе G таких, что любые две любые вершины $u_i \in U_1$ и $u_j \in U_1$ не являются смежными, $u_i \cap u_j = \varnothing$. Отличительной особенностью является то, что вершина u_k в графе G_r соответствует паре вершин $u_k=(x_i, x_j)$ графа G, которые в свою очередь соответствуют паре блоков $p_k=(b_i, b_j)$ размещаемых в одном контейнере. Пусть $|B|=n_b$, $|U_1|=n_b/2$. Критерий оптимизации – суммарная стоимость вершин в подмножестве U_1. Цель оптимизации минимизация критерия $F=\sum_k (\varphi(u_k))\ (k|u_k \in U_1)$.

Муравьиный алгоритм распределения блоков по контейнерам. Алгоритм распределения блоков по контейнерам базируется на методе муравьиной колонии Поиск решений осуществляется в виде двух итерационных процессов – внешнего и внутреннего, вложенного во внешний.

Каждая итерация с внутреннего цикла включает две процедуры, выполняемые агентом z_c, которые выполняются последовательно одна за другой. С помощью первой процедуры на каждой итерации с внутреннего цикла каждый агент z_c формирует конструктивным алгоритмом свое собственное решение – распределение блоков по контейнерам и рассчитывает оценку F_c решения. Лучшее решение внутреннего цикла запоминается.

Вторая процедура заключается в формировании множества феромонных точек, соответствующих решению, и отложении феромона агентом в промежуточные матрицы коллективной эволюционной памяти (КЭП) в количестве, пропорциональном целевой функции.

После выхода из внутреннего цикла выполняются две постпроцедуры.

Первая постпроцедура добавляет в основную матрицу КЭП феромон, накопленный в промежуточной матрице КЭП.

Второй постпроцедурой осуществляется испарение феромона в феромонных точках.

На каждой итерации внешнего цикла последовательно выполняются процедуры внутреннего цикла и две постпроцедуры.

Работа муравьиного алгоритма базируется на использовании коллективной эволюционной памяти (КЭП). КЭП муравьиной колонии представляет собой набор статистических показателей показывающее полезность фрагмента при построении решений на предыдущих итерациях алгоритма. Ядром муравьиного алгоритма является конструктивный алгоритм, с помощью которого каждый член популяции на каждой итерации находит решение задачи.

Для отражения коллективной эволюционной памяти в течение жизни популяции агентов и для формирования решения задачи используется граф учета качества решений $R(U,V^*)$, являющийся дополнением реберного графа $G_r=(U,V)$.

Граф $R(U,V^*)$, имеющий то же множество вершин, что и реберный граф $G_r=(U,V)$, но в котором две несовпадающие вершины u_i и u_j смежны тогда и только тогда, когда они не смежны в графе $G_r=(U,V)$. Формально дополнение графа $G_r=(U,V)$, определяется как граф $R(U,V^*)$, с исходным набором множества вершин U и с набором ребер V^*, полученным из полного графа на множестве вершин U, удалением множества ребер V имевшихся в заданном графе G_r. Дополнение пустого графа (содержащего только вершины, но нет рёбер) является полным графом, и наоборот. Независимое множество графа $G_r=(U,V)$ является кликой в дополнении $R(U,V^*)$ графа $G_r=(U,V)$, и наоборот.

Максимальному по мощности паросочетанию в графе G соответствует предельное внутренне-устойчивое подмножество реберного графа G_r, которое является кликой в дополнении $R(U,V^*)$ графа G_r.

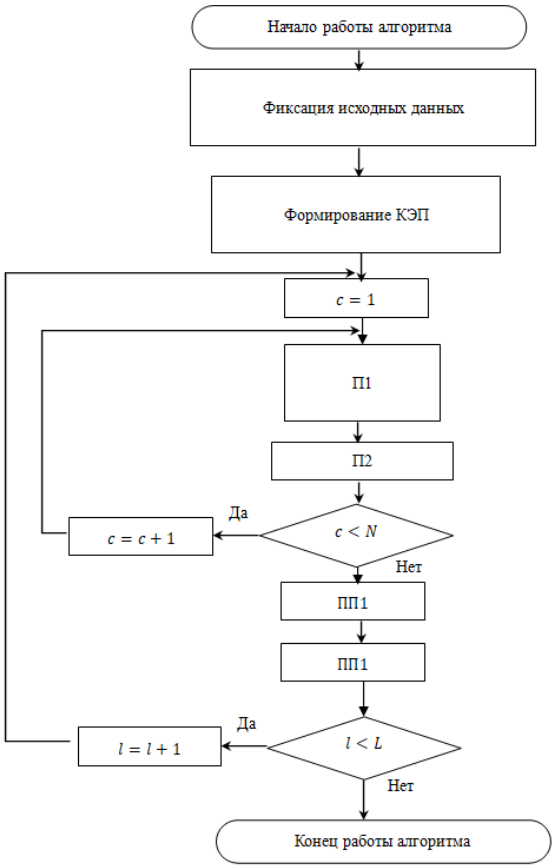


Рис. 9. Структурная схема муравьиного алгоритма

Механизмы формирования в графе независимого подмножества вершин на основе метода муравьиной колонии. Процесс поиска решений в МА итерационный. Каждая итерация l включает три процедуры, которые выполняются последовательно одна за другой. На первом этапе каждой итерации каждый агент z_c формирует конструктивным алгоритмом свое собственное решение – распределение блоков по контейнерам.

На втором этапе агенты откладывают феромон в матрицы коллективной эволюционной памяти (КЭП) в количестве пропорциональном целевой функции. На третьем этапе итерации осуществляется испарение феромона.

В основе конструктивного алгоритма распределения блоков по контейнерам лежит процедура формирования в графе $G_r=(U,V)$ максимального внутри устойчивого множества U_1 с минимальным значением суммарного веса элементов (контейнеров), входящих в его состав. Вершине $u_k \in U_1$ соответствует контейнер q_k. Вес сформированного агентом внутренне устойчивого множества U_1 определяется после компакции каждого контейнера, «входящего» в состав ВУМ U_1.

Классическая формулировка муравьиного алгоритма включает процедуру отложения феромона в феромонных точках, в качестве которых обычно служат ребра графов. В начальной части МА для откладывания феромона формируется множество $\Phi=\{\varphi_k|k=1,2,\ldots,n\}$ феромонных точек, соответствующих ребрам графа $R(U,V^*)$. Между Φ и V^* установлено соответствие $S_2:\Phi\rightarrow V^*$; $\varphi_k=S_2(v_k)$. $v_k \in V^*$ – ребро, связывающее вершины u_i и u_j графа $R(U,V^*)$. $v_k=(u_i, u_j)$.

В качестве коллективной эволюционной памяти (КЭП) используется матрица $\Delta=||\delta_{ij}||_{n\times n}$, в которой для каждого ребра $v \in V^*$ графа решений $R(U,V^*)$, хранится оценка степени пригодности δ_{ij} ребра $v \in V^*$, связывающего вершины u_i и u_j. Степень пригодности ребра определяется количеством отложенного феромона на ребре. В работе используется циклический (ant-cycle) метод муравьиных систем. В этом случае феромон откладывается в матрицу $\Delta=||\delta_{ij}||_{n\times n}$ после полного формирования всеми агентами решений на текущей итерации. Для накопления феромона, откладываемого агентами за одну итерацию, используется накопительная матрица $\Delta^*=||\delta^*_{ij}||_{n\times n}$. Другими словами, в процессе текущей итерации агенты по очереди откладывают феромон в $\Delta^*=||\delta^*_{ij}||_{n\times n}$, а после завершения итерации феромон, накопленный агентами в $\Delta^*=||\delta^*_{ij}||_{n\times n}$, откладывается в матрице $\Delta=||\delta_{ij}||_{n\times n}$.

Формирование ВУМ выполняется агентом z_c по шагам на базе опорного упорядоченного вектора $V=<1,2,3,4,\ldots,n>$ номеров блоков с 1 по n с динамически изменяемым составом. На каждом шаге t агентом z_c формируется вершина $u_k \in U$, $u_k=(x_i, x_j)$ графа $G_r=(U,V)$, которая включается в ВУМ $U_{1c}(t)$ графа $G_r=(U,V)$, а номера блоков x_i и x_j исключается из V. $U_{1c}(t)\subset U$. $U_{1c}(t+1)=U_{1c}(t)\cup u_k$. Отметим, что, если число блоков равно n, то число пар блоков равно $n/2$. Следовательно число вершин, входящих в ВУМ полного графа $G_r=(U,V)$ равно $n/2$. Каждый блок входит в состав только одной вершины $u_k \in U$.

Выбор вершины $u_k \in U$ включает две стадии. На первой стадии каждого шага t агентом z_c осуществляется формирование множества $P(t)\subset U$ вершин u_k графа G_r (пар $u_k=(b_i,b_j)$ блоков), рассматриваемых в качестве претендентов для включения в $U_{1c}(t)$, отличающихся тем, что в состав каждой пары блоков $u_k \in P(t)$ входит блок $b_i=S_1(x_i)$.

Формирование выполняется следующим образом. В векторе $V(t)$ выбирается первый по порядку номер элемента b_i в составе $V(t)$, который будет в составе всех пар множества $P(t)$ блоков первым элементом. Вторыми элементами у множества претендентов $P(t)$ являются оставшиеся элементы вектора $V(t)\backslash b_i$.

Каждая пара $u_k \in P(t)$ может быть включена в $U_{1c}(t)$, но включена будет только одна из них из них.

На второй стадии шага t, выбирается пара блоков $u_k \in P(t)$ которая вносится в состав ВУМ $U_{1c}(t)$, при этом блоки (b_i,b_j), входящие в u_k удаляются из состава вектора $V(t)$.

Выбор пары блоков $u_k=(b_i$ и $b_j)$, входящей в список претендентов $P(t)$, осуществляется вероятностным образом.

Для каждой $u_k \in P(t)$ на графе $R(U,V^*)$ определяется подмножество U_k вершин множества $U_{1c}(t)$, смежных u_k. $U_k \subset U_{1c}(t)$.

Рассчитываются стоимость S_k связи вершины $u_k \in P(t)$ с вершинами множества $U_1(t)$ по одной из формул:

$$S_k=(\lambda 1_k)^{\alpha} / (\lambda 2_k+1)^{\beta} \quad \text{при мультипликативной свертке,} \qquad (1)$$

$$S_k= \alpha(\lambda 1_k)+ \beta\, \lambda 2_k \quad \text{при аддитивной свертке,} \qquad (2)$$

где $\lambda 1_k$ – суммарный объем феромона на ребрах графа $R(U,V^*)$, связывающих u_k с вершинами подмножества$U_{1c}(t)$; $\lambda 2_k$ – вес вершины u_k.

Настройка значений параметров α и β производится экспериментально.

Далее для каждой вершины $u_k \in P(t)$ рассчитывается вероятность π_k включения в формируемое подмножество $U_1(t)$:

$$\pi_k = S_k / \sum_k (S_k). \qquad (3)$$

Выбранная с вероятностью π_k вершина $u_k \in P(t)$ (пара $u_k=(b_i,b_j)$ блоков) включается в множество $U_{1c}(t)$, а блоки b_i и b_j, входящие в u_k=(b_i и b_j), удаляются из вектора $V(t)$.

При $\alpha = 0$ вероятность выбора вершины u_k зависит только от веса вершины, без учета связей подмножества $U_1(t)$.

При β=0 вероятность выбора вершины u_k зависит только от объема феромона на связях с $U_1(t)$, что приводит к субоптимальным решениям.

Соотношение между значениями α и β находится экспериментально.

Пример: число блоков $n=10$. Необходимо сформировать 5 пар. Формируется вспомогательный вектор $V(1)$=<1,2,3,4,5,6,7,8,9,10 >. На первом шаге (t=1) выбирается первый элемент вектора $V(1)= 1$.

На первой стадии формируется множество пар претендентов для включения в $U_{1c}(1)$.

P_1 = (1,2), (1,3), (1,4), (1,5), (1,6), (1,7), (1,8), (1,9), (1,10).

Пусть на второй стадии с помощью приведенного выше правила выбрана пара u_1=(1,10).

u_1 включается в $U_{1c}(1)$, а блоки 1,10 исключаются из вектора $V(1)$.

Вектор $V(2)$=<2,3,4,5,6,7,8,9 >.

На втором шаге (t=2) выбирается первый элемент вектора $V(2) – 2$.

На первой стадии формируется множество пар претендентов для включения в $U_{1c}(2)$:

P_2 = (2,3), (2,4), (2,5), (2,6), (2,7), (2,8), (2,9).

Пусть на второй стадии с помощью приведенного выше правила выбрана пара u_2=(2,3).

u_2 включается в $U_{1c}(2)$, а блоки 2,3 исключаются из вектора $V(2)$.

Вектор $V(3)$=<4,5,6,7,8,9 >.

На третьем шаге (t=3) выбирается первый элемент вектора $V(3) – 4$.

На первой стадии формируется множество пар претендентов для включения в $U_{1c}(3)$:

P_3 = (4,5), (4,6), (4,7), (4,8), (4,9).

Пусть на второй стадии с помощью приведенного выше правила выбрана пара u_3=(4,6).

u_3 включается в $U_{1c}(3)$, а блоки 4,6 исключаются из вектора $V(3)$.

Вектор $V(4)$=<5,7,8,9>.

На четвертом шаге (t=4) выбирается первый элемент вектора $V(4) – 5$.

На первой стадии формируется множество пар претендентов для включения в $U_{1c}(4)$:

P_4= (5,7), (5,8), (5,9).

Пусть на второй стадии с помощью приведенного выше правила выбрана пара u_4=(5,8).

u_3 включается в $U_{1c}(4)$, а блоки 5,8 исключаются из вектора $V(4)$.

Вектор $V(5)$=<7,9> включает только два элемента.

В состав пары u_5 входят блоки 7,9.

ВУМ U_{1k}, построенное z_k, включает множество вершин (пар):

u_1=(1,10), u_2=(2,3), u_3=(4,6), u_4=(5,8), u_5=(7,9).

Для построенного агентом z_c ВУМ U_{1c} рассчитывается оценка F_c.

$$F_c=\sum_k \varphi(u_k),\ (k|u_k\in U_{1c}).$$

На построенном агентом z_c ВУМ U_{1c} в графе $R(U,V^*)$ формируется полный подграф Q_{1c}.

В ячейки матрицы $\Delta^*=||\delta^*_{ij}||_{n\times n}$, соответствующие всем ребрам подграфа Q_{1c} графа $R(U,V^*)$, добавляется порция феромона в объеме:

$$\varepsilon_k(l)=\Theta/F_c(l). \quad (4)$$

После того, как все агенты сформировали решения и добавили каждый из них свою порцию феромона в накопительную матрицу $\Delta^*=||\delta^*_{ij}||_{n\times n}$, производится сложение матриц $\Delta=||\delta_{ij}||_{n\times n}$ и $\Delta^*=||\delta^*ij||_{n\times n}$.

$$\Delta(t+1) = \Delta(t)+ \Delta^*(t).$$

На третьем этапе происходит общее испарение феромона на ребрах полного графа R (соответственно в ячейках матрицы КЭП) в соответствии с формулой.

$$\tau_k = \tau_k(1 - \rho). \quad (5)$$

После выполнения всех действий среди решений, полученных агентами на текущей итерации, выбирается лучшее, которое запоминается. Далее осуществляется переход на следующую итерацию. Временная сложность этого алгоритма зависит от времени жизни колонии l (число итераций), количества вершин графа n и числа агентов m, и определяется как $O(l\cdot n^2\cdot m)$.

Алгоритм распределения блоков по контейнерам. На начальном этапе алгоритма распределения множества блоков по контейнерам формируется стартовый набор данных:

- формируется секция исходных данных поставленной задачи, включающая параметры полосы и прямоугольных элементов, ограничения;
- формируется модели пространства поиска решений $G=(X,U)$, $G_r=(U,V)$, $R(U,V^*)$;
- для каждого ребра $u_k=(x_i,x_j)$ графа G определяется соответствующий ему контейнер q_k, включающий два блока $b_i=S_1(x_i)$ и $b_j=S_1(x_j)$;
- формируются структуры базы данных КЭП в виде матриц $\Delta(t)$ и Δ^*;
- выполняется компакция каждого контейнера q_k, и определяется его вес $\varphi(u_k)$;
- формирование начального набора UP решений в виде максимальных паросочетаний в графе G и рассчитывается оценка каждого решения;
- формирование исходной базы данных МА, путем отложения в ячейках матрицы $\Delta(t)$ феромона, пропорционального оценкам паросочетаний набора UP;
- задаются начальные значения управляющих параметров: C – число агентов, L – число итераций.

Агенты имеют собственную «память». На каждом шаге t в памяти агента z_c имеется:

- множество $U_{1c}(t)\subset U$ вершин, уже вошедших в состав формируемого ВУМ;
- опорный упорядоченный вектор $V(t)$;
- матрицы $\Delta(t)=||\delta_{ij}||_{n\times n}$ и $\Delta^*(c)=||\delta^*_{ij}||_{n\times n}$;
- δ_{ij} – количество феромона, на каждом ребре $v_{iv}=(x_i, p_v)$ графа $R(U,V^*)$.

1. Формирование исходных данных.

2. Формирование моделей поиска решений $G=(X,U)$, $G_r=(U,V)$, $R(U,V^*)$.

3. Для каждого ребра $u_k=(x_i, x_j)$ графа G определяется соответствующий ему контейнер q_k, выполняется компакция контейнера q_k, и определяется его вес (высота) $\varphi(u_k)$.

4. Формирование структур коллективной эволюционной памяти – основной и вспомогательной матриц

$\Delta(l)=||\delta_{ij}||_{n\times n}$, и $\Delta^*(c)=||\delta^*_{ij}||_{n\times n}$ степеней пригодности решений.

5. Обнуление матриц $\Delta(l)=||\delta_{ij}||_{n\times n}$ и $\Delta^*(c)=||\delta^*_i||_{n\times n}$.

Формирование начальной базы данных КЭП МА.

6. Формирование на базе графа $G=(X,U)$ начального набора UP максимальных паросочетаний U_{1i},

$UP=\{U_{1i}|i=1,2,\ldots,n\}$. $U_{1i}\in UP$ является ВУМ в $G_r=(U,V)$ и кликой в $R(U,V^*)$.

7. Расчет для каждого паросочетания U_{1i} суммарного веса F_i входящих в его состав вершин.

$F_i=\sum_k (\varphi(u_k))\ (k|u_k\in U_{1i})$.

8. На каждом множестве вершин U_{1i} графа $R(U,V^*)$ формируется полный подграф Q_{1i}.

9. Для каждого Q_{1i} формируется множество феромонных точек Φ_{1i}, соответствующих ребрам подграфа

Q_{1i}.

10. На каждом множестве феромонных точек Φ_{1i} в матрице $\Delta=||\delta_{ij}||_{n\times n}$ откладывается начальное

количество феромона $\tau_k(l)=\Theta/F_i$

11. Задается число итераций – L. Задается число агентов – N_a. $l=1$.

12. Начало новой итерации. $l= l +1$. $c=0$. (l – номер итерации, c– номер агента).

13. Выбор очередного агента. $c=c+1$.

Начало формирования паросочетания агентом z_c.

Приводится в начальное состояние память агента z_c.

Обнуляется вспомогательная матрица $\Delta^*(c)=||\delta^*_{ij}||_{n\times n}$.

14. Формируются начальные значения параметров памяти агента z_c. $t=0$. $U_{1c}(1)=\varnothing$.

15. Формируется опорный упорядоченный вектор $V(t)=<1,2,3,4,\ldots,n>$ номеров блоков с 1 по n с динамически изменяемым составом.

16. $t=t+1$. *(t – номер шага)*.

В векторе $V(t)$ выбирается первый по порядку в составе $V(t)$ номер блока b_i.

17. Формирование множества $P(t)\subset U$, $|P(t)|=n_b$ вершин u_k (соответствующих парам блоков), рассматриваемых в качестве претендентов для включения в $U_{1c}(t)$, отличающихся тем, что в состав каждой пары блоков $u_k\in P(t)$ в качестве первого элемента входит блок b_i.

Вторыми элементами у множества претендентов $P(t)$ являются оставшиеся элементы вектора $V(t)\backslash b_i$.

18. Для каждой u_k из списка претендентов $P(t)$ на графе $R(U,V^*)$:

- Определяется множество $U_k\subset U$ смежных ей вершин $u_j\in U_{1c}(t)$.
- Рассчитываются показатели:

  $\lambda 1_k$ – общий вес феромона на ребрах графа $R(U,V^*)$, связывающих $u_k\in P(t)$ со смежными ей вершинами $u_i\in U_{1c}(t)$;

  $\lambda 2_k$ – вес вершины u_k.
- Определяется стоимость S_k связей вершины $u_k\in P(t)$ с вершинами множества $U_1(t)$

$$S_k= \alpha(\lambda 1_k)+\beta\ \lambda 2_k$$

- Рассчитывается вероятность π_k включения вершины $u_k\in P(t)$ в формируемое подмножество $U_1(t)$

$$\pi_k =S_k/\sum_k (S_k).$$

19. Агент с вероятностью $\pi_k =W_k/\sum_k (W_k)$ выбирает одну из вершин $u_k\in P(t)$, которая включается в множество $U_{1c}(t)$, а блоки, входящие в состав выбранной пары u_k, исключается из состава вектора $V(t)$.

20. Если $t<n/2$, то переход к п. 14, иначе переход к п. 19.

21. Рассчитывается оценка F_c ВУМ U_{1c}

$$F_c=\sum_k (\varphi(u_k)\ (k|u_k\in U_{1c}).$$

22. Если $F_c < F_{опт}$ то $F_{опт}=F_c$, $U_{1опт}=U_{1c}$.

23. На множестве вершин U_{1c} графа $R(U,V^*)$ формируется клика Q_{1c}.

24. На множестве феромонных точек, соответствующих ребрам клики Q_{1c}, в матрице $\Delta^*(c)=||\delta^*_{ij}||_{n\times n}$ откладывается феромон в количестве $\tau_c(l)=\Theta/F_c(l)$.

25. Если $c<N_a$, то переход к п. 13, иначе переход к п. 26.

26. Феромон, накопленный агентами популяции в накопительной матрице $\Delta^*(c)=||\delta^*_{ij}||_{n\times m}$ на итерации l добавляется в матрицу $\Delta=||\delta_{ij}||_{n\times m}$. $\Delta(l)=\Delta(l) + \Delta^*(c)$.

27. Этап испарения феромона на всех ребрах графа $R(U,V^*)$. Значения всех ячеек матрицы $\Delta(l)=||\delta_{ij}||_{n\times m}$ уменьшаются в соответствии с формулой: $\delta_{ij} = \delta_{ij}(1 - \rho)$.

28. Если $l<L$, то переход к п. 12, иначе переход к п. 29.

29. Конец работа алгоритма.

Результаты эксперимента. Разработанный авторами алгоритм упаковки в полуограниченную полосу на основе декомпозиции и гибридизации биоиспирированных методов поисковой оптимизации реализован в виде программы DBK. Эксперименты проводились на основе методики, изложенной в работе [20], в которой исследовался алгоритм GBA.

Для оценки эффективности разработанных методик и программного обеспечения использовались данные из электронной библиотеки OR-library. Эта библиотека содержит примеры упаковки, для которых заранее известно оптимальное значение высоты занятой полосы. Разработанный алгоритм сравнивался с алгоритмами GAGA, PH, DDA и GBA.

Алгоритм GAGA, описанный в исследовании [21], разработан на основе гибридизации конструктивного и генетического методов. Результаты исследования показали высокую вычислительную эффективность. Алгоритм исследования PH [22] разработан на основе механизмов, инспирированных природой. Алгоритм DDA основан на эвристике динамической декомпозиции [23]. Алгоритмом GBA [20] задача упаковки реализуется при использовании гильотинного разреза ленты на контейнеры и гильотинного разреза контейнеров, реализован на основе методов роевого интеллекта.

На первом этапе агентом решается задача упаковки набора элементов прямоугольной формы в множестве блоков прямоугольной формы с фиксированными размерами высоты и ширины. Все блоки имеют одинаковую ширину D, равную ширине полосы. Тестирования вышеуказанных алгоритмов проводилось на тестовых задачах SCPL1 ~ SCPL9 из OR-Benchmark. Результаты расчетов представлены в табл. 1, где N – число прямоугольников, H^* оптимальное значение.

Таблица 1

Сравнение с тестами из OR-Benchmark

Inst	*N*	*Opt*	*GAHA*	*PH*	*DDA*	*GBA*	*DBK*
		*H**	*H*	*H*	*H*	*H*	*H*
SCPL1	425	110	126.08	183.5	123	121	116
SCPL2	127	120	140.16	163.5	128	125	122
SCPL3	225	84	98.84	102	98	90	88
SCPL4	365	102	108.263	109.5	106	102	102
SCPL5	165	102	118.89	118	110	102	102
SCPL6	657	126	142.4	134	132	132	128
SCPL7	357	198	229.93	242	216	204	200
SCPL8	475	156	168.419	167	165	164	162
SCPL9	175	117	129.96	127	125	121	119

Как видно из табл. 1, результаты алгоритма DBK лучше, чем у сравниваемых алгоритмов, и некоторые результаты достигли предельной высоты.

Набор данных случайно выбран из OR-Benchmark. (http://www.laria.u-picardie.fr/hifi/OR-Benchmark). Элементы имеют ширину 135, высота не ограничена. Используя тестовые данные OR-Library, разработанный алгоритм DBK сравнивался с алгоритмами EA, DDA и GBA. Результаты приведены в табл. 2.

Таблица 2

Сравнение с тестами из OR- Library

Test	H^*	*Type*	*n*	*EA*	*DDA*	*GBA*	*DBK*
				H	*H*	*H*	*H*
cgcut 1	23	7	16	23	23	23	23
cgcut 2	63	10	23	67	63	63	63
cgcut 3	636	20	62	670	642	640	638
ngcut 1	20	5	10	23	20	20	20
ngcut 2	28	7	17	30	28	28	28
ngcut 3	28	10	21	28	28	28	28
ngcut 4	18	5	7	20	18	18	18
ngcut 5	36	7	14	36	36	36	36
ngcut 6	29	10	15	31	29	29	29
ngcut 7	10	5	8	20	10	10	10
ngcut 8	33	7	13	33	33	33	32
ngcut 9	49	10	18	50	49	49	49
ngcut 10	59	5	13	80	59	59	59
ngcut 11	51	7	15	52	51	51	51
ngcut12	77	10	22	87	77	77	77
beng01	30	20	20	30	30	30	30
beng02	57	40	40	58	57	57	57
beng03	84	60	60	85	84	84	84
beng04	107	80	80	108	107	107	107
beng05	134	100	100	134	134	134	134
beng06	36	40	40	37	36	36	36
beng07	67	80	80	67	67	67	67

Сравнение с результатами исследований алгоритмов *EA, DDA, GBA* показывают, что разработанный алгоритм имеет явное преимущество с точки зрения времени вычислений и точности вычислений.

По сравнению с существующими алгоритмами достигнуто улучшение результатов раскроя-упаковки полуограниченной полосы на 3-5%. Полученная временная сложность алгоритма, практически совпадает с теоретическими исследованиями и для рассмотренных тестовых задач составляет (ВСА$\approx O(n^2)$).

Заключение. Предложена архитектура и методология раскроя-упаковки полуограниченной полосы на основе методов биоинспирированного поиска. В основе метода декомпозиции обшей задачи упаковки и методологии формирования карт раскроя лежат эвристики уровневого подхода к упаковке полосы. Уровневые алгоритмы обладают лучшими возможностями улучшения и коррекции решения. Упаковка выполняется путем последовательного заполнения уровней полосы контейнерами. Архитектура, сформированная на основе декомпозиции общей задачи, включает 5 основных секций: управление процессом поиска; формирования блоков; формирование контейнеров; компакция контейнеров; заполнения полосы контейнерами. В состав стандартного маршрута поиска решения дополнительно включена процедура распределения блоков между контейнерами и процедура компакции контейнера. Целью добавленных процедур является минимизация общей площади контейнера путем плотного размещении блоков. Это позволило улучшить характеристики решений, что подтверждается результатами экспериментов.

Компакцию последовательно проводят во всех контейнерах. Анализ результатов распределения
блоков по контейнерам показал, что максимальная плотность упаковки полосы, зависит от распределения блоков по контейнерам. В работе рассматриваются два возможных варианта размещения блоков в контейнере, в зависимости от сочетания типов разрезов: гильотинный и не гильотинный, и компакции контейнера.

Разработана методика и конструктивный метод распределения блоков по контейнерам, используемый агентами, отличающийся линейной трудоемкостью. Формирование контейнеров выполняется последовательно по шагам. На первой стадии шага *t* формиру-

ется множество пар блоков-претендентов для назначения в очередной контейнер. На второй стадии вероятностным образом выбирается пара блоков, входящей в список претендентов.

Разработана методика и метод компакции контейнера. Компакция выполняется на основе графа ограничений, формируемого методом распространения тени.

Разработана методика формирования феромоновых точек и структур данных коллективной эволюционной памяти. Предложены новые модели и структуры данных для организации коллективной эволюционной памяти. Задача распределения блоков по контейнерам сведена к задачам построения паросочетания и выделения внутренне-устойчивого множества, для решения которых разработаны эффективные биоинспирированные алгоритмы.

Разработанный алгоритм формирования карты раскроя на основе биоинспирированного поиска позволяет значительно сократить отходы материала.

По сравнению с существующими алгоритмами достигнуто улучшение результат раскроя-упаковки полуограниченной полосы на 3-5%. Временная сложность алгоритма (ВСА), полученная экспериментальным путем, практически совпадает с теоретическими исследованиями и для рассмотренных тестовых задач составляет (ВСА ≈ $O(n^2)$).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Stephen C.H., Defu Z.* A fast layer-based heuristic for non-guillotine strip packing // Expert System with Application. – 2011. – P. 13032-13042.
2. *Тимофеева О.П., Соколова Э.С., Милов К.В.* Генетический алгоритм в оптимизации упаковки контейнеров // Труды НГТУ. Информатика и системы управления. – 2013. – № 4 (101). – С. 167-172.
3. *Валеева А.Ф.* Применение конструктивных эвристик в задачах раскроя-упаковки // Приложение к журналу «Информационные технологии». – 2006. – № 11. – С. 1-24.
4. *Лебедев Б.К., Лебедев О.Б., Лебедева Е.М.* Модернизированный муравьиный алгоритм синтеза идентифицированного дерева гильотинного разреза при планировании СБИС // Известия ЮФУ. Технические науки. – 2017. – № 7. – С. 15-28.
5. *Лебедев О.Б., Зорин В.Ю.* Упаковка на основе метода муравьиной колонии // Известия ЮФУ. Технические науки. – 2010. – № 12. – С. 25-30.
6. *Defu Zhang, Yuxin C., Furong Y., Yain-Whar S., Stephen C.* A hybrid algorithm based on variable neighbourhood for the strip packing problem // Journal of Combinatorial Optimization. – 2016. – P. 513-530.
7. *Ахтямов А.А., Картак В.М.* Упаковка и оценка плотности упаковки ортогональных многоугольников в полубесконечную полосу // Информационные технологии интеллектуальной поддержки принятия решений. – 2014. – С. 117-121.
8. *Jaya T., Narendra S.* Hybrid approach for 2d strip packing problem using genetic algorithm // Proceedings of the 12th international conference on Artificial Neural Networks: advances in computational intelligence. – 2013.
9. *Месягутов М.А., Мухачева Э.А., Белов Г.Н., Шайтхауэр Г.* Упаковка одномерных контейнеров с продолженным выбором идентичных предметов: точный метод поиска оптимального решения // Автоматика и телемеханика. – 2011. – № 1. – С. 154-173.
10. *Lei W., Aihua Y.* A quasi-human algorithm for the two dimensional rectangular strip packing problem: in memory // Journal of Combinatorial Optimization. – 2016. – P. 416-444.
11. *Потарусов Р.В., Курейчик В.М.* Проблема одномерной упаковки элементов // Известия ТРТУ. – 2006. – № 8. – С. 88-93.
12. *Карпенко А.П.* Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой: учеб. пособие. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2021. – 448 с.
13. *Лебедев О.Б.* Модели адаптивного поведения муравьиной колонии в задачах проектирования. – Таганрог: Изд-во ЮФУ. 2013. – 199 с.
14. *Лебедев Б.К., Лебедев О.Б.* Муравьиные алгоритмы разбиения, использующие представления задачи, отличные от канонического // Вестник Ростовского государственного университета путей сообщения. – 2016. – № 3 (63). – С. 42-47.
15. *Blum C., Roli A.* Metaheuristics in combinatorial optimization: overview and conceptual comparison // ACM computing surveys. – 2003. – P. 268-308.

16. *Лебедев Б.К., Лебедев О.Б.* Распределение ресурсов на основе гибридных моделей роевого интеллекта // Научно-технический вестник информационных технологий, механики и оптики. – 2017. – № 6. – С. 1063-1073.
17. *Житников В.П.* Задача прямоугольной упаковки в полубесконечную полосу: поиск решения в окрестности локальной нижней границы // Информационные технологии. – 2007. – № 6. – С. 55-61.
18. *Raidl G.R.* A Unified View on Hybrid Metaheuristics // Lecture Notes In CS. – 2006. – P. 1-12.
19. *Лебедев Б.К., Лебедев О.Б.* Роевой алгоритм планирования работы многопроцессорных вычислительных систем // Инженерный вестник Дона. – 2017. – № 3.
20. *Лебедев Б.К., Лебедев О.Б., Ганжур М.А.* Упаковка в полубесконечную полосу на основе декомпозиции и гибридизации биоинспирированных методов // Известия ЮФУ. Технические науки. – 2023. – № 4. – С. 109-125.
21. *Cong J., Romesis M., Xie M.* Optimality, Scalability and Stability Study of Partitioning and Placement Algorithms // Proc. of the International Symposium on Physical Design. – 2003. – P. 88-94.
22. *Hifi M., M'Hallah R.* A hybrid algorithm for the two-dimensional layout problem: The cases of regular and irregular shapes // International Transactions in Operational Research. – 2003. – P. 195-216.
23. *Zhang DeFu, Han ShuiHua, Ye WeiGuo.* A Bricklaying Heuristic Algorithm for the Orthogonal Rectangular Packing Problem // Chinese Journal of Computers. – 2008. – P. 509- 515.

REFERENCES

1. *Stephen C.H., Defu Z.* A fast layer-based heuristic for non-guillotine strip packing, *Expert System with Application,* 2011, pp. 13032-13042.
2. *Timofeeva O.P., Sokolova E.S., Milov K.V.* Geneticheskiy algoritm v optimizatsii upakovki konteynerov [Genetic algorithm in optimization of container packaging], *Trudy NGTU. Informatika i sistemy upravleniya* [Proceedings of NSTU. Informatics and control systems], 2013, No. 4 (101), pp. 167-172.
3. *Valeeva A.F.* Primenenie konstruktivnykh evristik v zadachakh raskroya-upakovki [Application of constructive heuristics in cutting-packing problems], *Prilozhenie k zhurnalu «Informatsionnye tekhnologii»* [Supplement to the journal «Information Technologies»], 2006, No. 11, pp. 1-24.
4. *Lebedev B.K., Lebedev O.B., Lebedeva E.M.* Modernizirovannyy murav'inyy algoritm sinteza identifitsirovannogo dereva gil'otinnogo razreza pri planirovanii SBIS [Modernized ant algorithm for synthesizing an identified guillotine cut tree when planning VLSI], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2017, No. 7, pp. 15-28.
5. *Lebedev O.B., Zorin V.Yu*. Upakovka na osnove metoda murav'inoy kolonii [Packaging based on the ant colony method], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2010, No. 12, pp. 25-30.
6. *Defu Zhang, Yuxin C., Furong Y., Yain-Whar S., Stephen C.* A hybrid algorithm based on variable neighbourhood for the strip packing problem, *Journal of Combinatorial Optimization*, 2016, pp. 513-530.
7. *Akhtyamov A.A., Kartak V.M.* Upakovka i otsenka plotnosti upakovki ortogonal'nykh mnogougol'nikov v polubeskonechnuyu polosu [A hybrid algorithm based on variable neighbourhood for the strip packing problem], *Informatsionnye tekhnologii intellektual'noy podderzhki prinyatiya resheniy* [Information technologies for intelligent decision support], 2014, pp. 117-121.
8. *Jaya T., Narendra S.* Hybrid approach for 2d strip packing problem using genetic algorithm, *Proceedings of the 12th international conference on Artificial Neural Networks: advances in computational intelligence,* 2013.
9. *Mesyagutov M.A., Mukhacheva E.A., Belov G.N., Shaytkhauer G.* Upakovka odnomernykh konteynerov s prodolzhennym vyborom identichnykh predmetov: tochnyy metod poiska optimal'nogo resheniya [Packing one-dimensional containers with continued selection of identical items: an exact method for finding the optimal solution], *Avtomatika i telemekhanika* [Automation and Remote Control], 2011, No. 1, pp. 154-173.
10. *Lei W., Aihua Y.* A quasi-human algorithm for the two dimensional rectangular strip packing problem: in memory, *Journal of Combinatorial Optimization,* 2016, pp. 416-444.
11. *Potarusov R.V., Kureychik V.M.* Problema odnomernoy upakovki elementov [The problem of one-dimensional packing of elements], *Izvestiya TRTU* [Izvestiya TSURE], 2006, No. 8, pp. 88-93.
12. *Karpenko A.P.* Sovremennye algoritmy poiskovoy optimizatsii. Algoritmy, vdokhnovlennye prirodoy: ucheb. posobie [Modern search engine optimization algorithms. Algorithms inspired by nature: a tutorial]. Moscow: Izd-vo MGTU im. N.E. Baumana, 2021, 448 p.
13. *Lebedev O.B.* Modeli adaptivnogo povedeniya murav'inoy kolonii v zadachakh proektirovaniya [Models of adaptive behavior of an ant colony in design problems]. Taganrog: Izd-vo YuFU. 2013, 199 p.

14. *Lebedev B.K., Lebedev O.B.* Murav'inye algoritmy razbieniya, ispol'zuyushchie predstavleniya zadachi, otlichnye ot kanonicheskogo [Ant partitioning algorithms using problem representations different from the canonical one], *Vestnik Rostovskogo gosudarstvennogo universiteta putey soobshcheniya* [Bulletin of the Rostov State University of Transport and Communications], 2016, No. 3 (63), pp. 42-47.
15. *Blum C., Roli A.* Metaheuristics in combinatorial optimization: overview and conceptual comparison, *ACM computing surveys,* 2003, pp. 268-308.
16. *Lebedev B.K., Lebedev O.B.* Raspredelenie resursov na osnove gibridnykh modeley roevogo intellekta [Resource distribution based on hybrid models of swarm intelligence], *Nauchno-tekhnicheskiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Scientific and Technical Bulletin of Information Technologies, Mechanics and Optics], 2017, No. 6, pp. 1063-1073.
17. *Zhitnikov V.P.* Zadacha pryamougol'noy upakovki v polubeskonechnuyu polosu: poisk resheniya v okrestnosti lokal'noy nizhney granitsy [The problem of rectangular packing in a semi-infinite strip: searching for a solution in the neighborhood of the local lower boundary], *Informatsionnye tekhnologii* [Information Technologies], 2007, No. 6, pp. 55-61.
18. *Raidl G.R.* A Unified View on Hybrid Metaheuristics, *Lecture Notes In CS,* 2006, pp. 1-12.
19. *Lebedev B.K., Lebedev O.B.* Roevoy algoritm planirovaniya raboty mnogoprotsessornykh vychislitel'nykh sistem [Swarm algorithm for scheduling the operation of multiprocessor computing systems], *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don], 2017, No. 3.
20. *Lebedev B.K., Lebedev O.B., Ganzhur M.A.* Upakovka v polubeskonechnuyu polosu na osnove dekompozitsii i gibridizatsii bioinspirirovannykh metodov [Packaging in a semi-infinite strip based on decomposition and hybridization of bioinspired methods], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2023, No. 4, pp. 109-125.
21. *Cong J., Romesis M., Xie M.* Optimality, Scalability and Stability Study of Partitioning and Placement Algorithms, *Proc. of the International Symposium on Physical Design,* 2003, pp. 88-94.
22. *Hifi M., M'Hallah R.* A hybrid algorithm for the two-dimensional layout problem: The cases of regular and irregular shapes, *International Transactions in Operational Research*, 2003, pp. 195-216.
23. *Zhang DeFu, Han ShuiHua, Ye WeiGuo.* A Bricklaying Heuristic Algorithm for the Orthogonal Rectangular Packing Problem, *Chinese Journal of Computers,* 2008, pp. 509-515.

Статью рекомендовал к опубликованию д.т.н., профессор А.В. Боженюк.

Лебедев Борис Константинович – Южный федеральный университет; e-mail: lebedev.b.k@gmail.com; г. Таганрог, Россия; тел.: 89282897933; кафедра систем автоматизированного проектирования; д.т.н.; профессор.

Лебедев Олег Борисович – МИРЭА – Российский технологический университет; e-mail: lebedev.ob@mail.ru; г. Москва, Россия; тел.: 89085135512; кафедра информатики; д.т.н.; профессор.

Ганжур Марина Александровна – Донской государственный технический университет; e-mail: mganzhur@yandex.ru; г. Ростов-на-Дону; тел.: 89081819111; кафедра вычислительных систем и информационной безопасности; старший преподаватель.

Lebedev Boris Konstantinovich – Southern Federal University; e-mail: lebedev.b.k@gmail.com; Taganrog, Russia; phone: +79282897933; the department of Computer Aided Design; dr. of eng. sc.; professor.

Lebedev Oleg Borisovich – MIREA – Russian Technological University; e-mail: lebedev.ob@mail.ru; Moscow, Russia; phone: +79085135512; the department of Computer Science; dr. of eng. sc.; professor.

Ganzhur Marina Aleksandrovna – Don State Technical University; e-mail: mganzhur@yandex.ru; Rostov-on-Don; phone: +79081819111; the department of Computing Systems and Information; security senior lecturer.

УДК 681.327 DOI 10.18522/2311-3103-2024-4-31-40

П.О. Никашина

ОПРЕДЕЛЕНИЕ МАКСИМАЛЬНОГО ПОТОКА В НЕЧЕТКОМ ПЕРИОДИЧЕСКОМ ГРАФЕ

Предложен метод нахождения максимального значения динамического потока с использованием периодических графов, представленном в виде обобщенной сети. Интерес к сетям такого вида объясняется их широким практическим применением в местах, где есть периодичность, например управление периодическими пассажирскими перевозками на различных видах транспорта, грузовые перевозки, в том числе товаров с коротким сроком годности, управление дорожно-транспортным потоком, а именно регулирование светофоров, с учетом периодичности и загруженности. В то же время значения пропускной способности дуг рассматриваемых сетей могут варьироваться в зависимости от времени отправления потока и возможных циклов, поэтому мы переходим к динамическим сетям. Параметры сети представлены в нечеткой форме из-за влияния факторов окружающей среды и деятельности человека. А выбор именно периодических графов обусловлен наличием циклов и периодичностью временных интервалов. Рассмотренные типы сетей могут быть реализованы на реальных дорогах в процессе транспортировки. Для решения выявленной проблемы, в рамках представленной работы приведен краткий обзор литературных источников, позволяющий оценить современный уровень развития систем подобного назначения. В результате выполнения данного обзора установлено, что наиболее эффективными методами решения поставленной проблемы, является применение методов нечетких периодических графов. В связи с этим принято решение о проведении исследования указанных методов. Новизна данной работы определяется исходя из применения периодических темпоральных нечетких графов в рамках решения задачи нахождения максимального потока динамической сети.

Нечеткая сеть; нечеткие динамические графы; периодические графы.

P.O. Nikashina

DETERMINATION OF MAXIMUM FLOW IN A FUZZY PERIODIC GRAPH

The article illustrates a method for finding the maximum value of a dynamic flow using periodic graphs, presented in the form of a generalized network. The interest in networks of this type is explained by their wide practical application in places where there is periodicity, for example, management of periodic passenger transportation on various types of transport, freight transportation, including goods with a short shelf life, management of road traffic flow, namely regulation traffic lights, taking into account frequency and workload. At the same time, the values of the bandwidth of the arcs of the networks under consideration may vary depending on the time of departure of the stream and possible cycles, so we turn to dynamic networks. Network parameters are presented in a fuzzy form due to the influence of environmental factors and human activity. And the choice of periodic graphs is due to the presence of cycles and the frequency of time intervals. The considered types of networks can be implemented on real roads during transportation. To solve the identified problem, within the framework of the presented work, a brief overview of literary sources is provided, which allows us to assess the current level of development of systems for such purposes. As a result of this review, it was found that the most effective methods of solving the problem posed are the use of fuzzy periodic graph methods. In this regard, it was decided to conduct a study of these methods. The novelty of this work is determined based on the use of periodic temporal fuzzy graphs in solving the problem of finding the maximum flow of a dynamic network.

Fuzzy network; fuzzy dynamic graphs; periodic graphs.

Введение. Обычные задачи управления потоками имеют различные вариации, практические реализации и предполагают передачу потока от источника к приемнику сети для нахождения максимального значения потока. Во всех возможных постановках задачи, исходными данными является емкость сети и цель состоит в том, чтобы найти экстремальное значение расхода без превышения пропускной способности сети по дуге. Другим сходством обычных задач о потоке является условие сохранения потока, в частности, предположение о том, что поток сохраняется на каждой дуге. Однако, если брать во внимание различные типы практических применений, например, периодичные перевозки с определенным циклом транспортировки, нарушение которых может привести к

повреждению или порче продукта, с одной стороны, перевозка скоропортящихся товаров приводит к снижению первоначальной перевозимой стоимости, а с другой стороны, некоторые виды товаров могут увеличить свою стоимость в процессе транспортировки.

Следовательно, в таких постановках задачи мы имеем дело не только с обобщенными сетями. Но и с периодическими графовыми системами.

В нашем алгоритме мы предлагаем учитывать неопределенность, свойственную всем видам сетей, и такие параметры, как пропускная способность дуг, затраты на транспортировку, а также учитывать возможную периодичность и цикличность в таких сетях, которые обусловлены, как природными явлениями, так техногенными и рукотворными.

Еще одно важное замечание заключается в том, что при решении задач с потоком важно учитывать динамический характер мощности дуги, поскольку они могут изменять свои значения во времени и зависеть от времени выхода потока. Эти модификации приводят к необходимости иметь дело с нечеткими динамическими сетями и рассматривать задачи в сетях такого типа.

1. Существующие методы. Нечеткие периодические графы представляют собой структуру данных, которая моделирует системы с циклическими процессами, где неопределенность и вариативность присутствуют как в объемах потока, так и в пропускных способностях дуг. Эта модель находит широкое применение в различных областях, включая управление трафиком, планирование поставок, проектирование телекоммуникационных сетей и распределение ресурсов. Нахождение максимального динамического потока в нечетких периодических графах представляет собой важную задачу, которая позволяет оптимизировать использование ресурсов в таких системах.

Существуют различные подходы к решению задачи поиска максимального динамического потока в нечетких периодических графах.

1. Методы на основе нечеткой линейной оптимизации.

Одна из первых работ в этой области была проведена А. Кауфманном и М. Гупта, где авторы предложили метод, основанный на нечеткой линейной оптимизации с использованием нечетких чисел [1].

М. Инуигучи, Дж. Рамик разработали метод, основанный на нечеткой многоцелевой оптимизации, который учитывает, как ограничения пропускной способности, так и потребности пользователей [2].

2. Методы на основе теории нечетких множеств.

Дж. Бакли, Т. Феринг в своей работе использовали теорию нечетких множеств для определения нечеткой пропускной способности дуг, а затем применили алгоритм поиска максимального потока для расчета нечеткого максимального потока [3].

Л. Заде предложил новый подход, который использует нечеткие отношения для определения нечетких графов и затем применяет алгоритмы поиска максимального потока к этим нечетким графам [4].

3. Гибридные методы:

Д. Голдберг разработал гибридный метод, который сочетает в себе нечеткую линейную оптимизацию и генетические алгоритмы для поиска оптимального решения [5].

Дж. Холланд предложил метод, который использует нечеткие правила и эволюционные алгоритмы для определения максимального динамического потока [6].

Разные методы имеют свои преимущества и недостатки. Методы на основе нечеткой линейной оптимизации просты в реализации, но могут быть недостаточно точными, особенно при больших объемах неопределенности. Методы, основанные на теории нечетких множеств, могут обеспечить более точные решения, но могут быть сложнее в реализации. Гибридные методы обычно обеспечивают компромисс между точностью и сложностью.

Несмотря на то, что существует множество исследований по данной теме, существует ряд открытых вопросов:

- Разработка более эффективных алгоритмов для решения задачи поиска максимального динамического потока в нечетких периодических графах.

- Разработка новых методов, которые учитывают динамическую природу неопределенности.
- Разработка методов, которые могут быть применены к реальным системам с большими объемами данных.

2. Метод нахождения потока. Обобщенная задача о потоке является обобщением обычной задачи о максимальном расходе, которая нарушает предположение о сохранении расхода. Аналогично обычным потоковым сетям рассматривается емкостная сеть, где каждая дуга имеет пропускную способность $\tilde{a}$, которая ограничивает количество потока, передаваемого в его дугу.

Более того, обобщенная сеть имеет дополнительный параметр – коэффициент усиления потока или эффективность потока. Этот параметр p является положительным множителем и указывает, что если $\tilde{m}$ единиц расхода войдут в дугу, затем $\tilde{a} \times p$ единиц измерения расхода выйдут [7].

В рамках определения периодических графов, стоит отметить их отличительные характеристики в сравнении с классическими графами, а именно: инициализация пропускной способности, совместно с заданием времени повторения (тактами), а также возможностью определения количества задержек на той или иной вершине. Пример данного графа представлен на рис. 1.

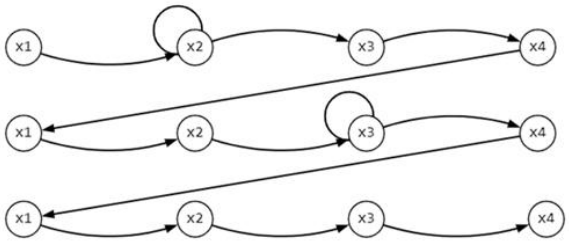


Рис. 1. Пример периодического нечеткого графа

Пусть задача о нахождении максимального динамического потока в нечеткой обобщенной сети имеет вид:

$$Maximize\ \tilde{b}(p). \tag{1}$$

$$\sum_{\vartheta=0}^{p}\left(\sum_{xj\in U(x_i)}\tilde{\delta}_{ij}(\vartheta)-\sum_{xj\in U^{-1}(x_i)}\tilde{\delta}_{ij}(\vartheta-\mu_{ij}(\vartheta))\right)=\tilde{b}(p), x_i=s. \tag{2}$$

$$\sum_{\vartheta=0}^{p}\left(\sum_{xj\in U(x_i)}\tilde{\delta}_{ij}(\vartheta)-\sum_{xj\in U^{-1}(x_i)}\tilde{\delta}_{ij}\left(\vartheta-\mu_{ij}(\vartheta)\right)\right)=\tilde{0}, x_i\neq s,t;\vartheta\in T. \tag{3}$$

$$\sum_{\vartheta=0}^{p}\left(\sum_{xj\in U(x_i)}\tilde{\delta}_{ij}(\vartheta)-\sum_{xj\in U^{-1}(x_i)}\tilde{\delta}_{ij}\left(\vartheta-\mu_{ij}(\vartheta)\right)\right)=-\tilde{b}(p), x_i=t. \tag{4}$$

$$\sum_{xj\in U^{-1}(x_i)}\tilde{\delta}^{ij}_{xi\in U^{-1}(x_i)}\left(\vartheta+\mu_{ij}\right)=p_{ij}\times\ \tilde{\delta}^{ij}_{xi\in U(x_i)}. \tag{5}$$

$$\tilde{0}\leq\tilde{\delta}^{ij}_{xi\in U(x_i)}\leq\tilde{a}_{ij}(\vartheta), \forall\left(x_i,x_j\right)\in\tilde{A}, \vartheta\in T. \tag{6}$$

Уравнение (1) $\tilde{b}(p)$ отражает максимизацию расхода. Уравнение (2) является условием сохранения расхода для источника $\sum_{\vartheta=0}^{p}\left(\sum_{xj\in U(x_i)}\tilde{\delta}_{ij}(\vartheta)\right), x_i=s$. Уравнение (4) является условием сохранения расхода для приемника $\sum_{\vartheta=0}^{p}\left(\sum_{xj\in U^{-1}(x_i)}\tilde{\delta}_{ij}(\vartheta-\mu_{ij}(\vartheta))\right), x_i=t$. Уравнение (3) представляет собой условие сохранения потока для промежуточных узлов. Равенство (5) указывает на зависимость между входящим потоком, коэффициентом полезного действия и выходящим потоком. Неравенство (6) указывает $\tilde{\delta}_{ij}(\vartheta)$, что потоки за все периоды времени должны быть меньше, чем пропускная способность соответствующих дуг $\tilde{a}_{ij}(\vartheta)$ за те же периоды времени.

Для представления предлагаемого алгоритма введем правила перехода к развернутому во времени варианту исходного динамического графа, перехода к нечеткой остаточной сети “развёрнутому во времени” графу и перехода в нечеткий периодический граф, основанными на [8–10].

Правило 1 – переход к развернутому во времени варианту исходного динамического графа.

Цель состоит в том, чтобы перейти от начальной версии динамической сети, представленной в виде графика, к развернутой во времени версии, путем создания отдельной копии каждого узла $x_i \in X$ в каждый период времени $\vartheta \in T$. $\tilde{G}_p = (X_p, \tilde{A}_p)$. Набор узлов X_p может быть определен как $X_p = \{(x_i, \vartheta): (x_j, \vartheta) \in X \times T\}$. В $\tilde{A}_p$ входят дуги из пары вершина-время $(x_i, \vartheta) \in X_p$ для каждого узла-временя пара $\left(x_j, \theta = \vartheta + \mu_{ij}(\vartheta)\right)$, где $x_j \in U(x_i)$ и $\vartheta + \mu_{ij}(\vartheta) \leq p$. Значение дуги $\tilde{a}_{ij}(x_i, x_j, \vartheta, \theta)$, соединяющей (x_i, ϑ) с (x_j, ϑ) равно $\tilde{a}_{ij}(\vartheta)$. Факторы эффективности потока $\tilde{p}(x_i, x_j, \vartheta, \theta)$ соединяющей (x_i, ϑ) с (x_j, ϑ) равны $\tilde{p}_{ij}(\vartheta)$. Транзитное время $p(x_i, x_j, \vartheta, \theta)$, соединяющее (x_i, ϑ) с (x_j, ϑ), равно $\mu_{ij}(\vartheta)$.

Правило 2 – переход к нечеткой сети "развернутого во времени" графа.

Построение нечеткой остаточной сети $\tilde{G}_p^\tau = (X_p^\tau, \tilde{A}_p^\tau)$ реализуется в соответствии с "развернутым во времени" графом $\tilde{G}_p = (X_p, \tilde{A}_p)$ и значениями потока $\tilde{\delta}\left(x_i, x_j, \vartheta, \theta = \vartheta + \mu_{ij}(\vartheta)\right)$, исходящими из узлов (x_i^τ, ϑ) и входящими в узлы (x_j^τ, θ) каждой дуги в остаточной сети $\tilde{G}_p^\tau$, головной узел которой равен (x_i^τ, ϑ) , а хвостовой узел равен (x_j^τ, θ) с расходом $\tilde{\delta}(x_i, x_j, \vartheta, \theta)$ исходящую из вершины x_i в момент $\vartheta \in T$, остаточная емкость нечетких дуг потока $\tilde{a}^\tau(x_i, x_j, \vartheta, \theta) = \tilde{a}(x_i, x_j, \vartheta, \theta) - \tilde{\delta}(x_i^+, x_j^-, \vartheta^+)$ с транзитным временем $\tilde{\mu}^\tau(x_i, x_j, \vartheta, \theta) = \tilde{\mu}(x_i, x_j, \vartheta, \theta)$ и обратная дуга $\tilde{p}^\tau(x_i, x_j, \vartheta, \theta) = \tilde{p}(x_i, x_j, \vartheta, \theta)$, с головой-узлом (x_i^τ, ϑ) и хвост-узел (x_j^τ, θ) с нечеткими дугами емкости $\tilde{a}^\tau(x_i, x_j, \vartheta, \theta) = \tilde{\delta}(x_j^-, x_i^+\vartheta + \mu_{ij}^-)$, время в пути $\mu^\tau(x_i, x_j, \vartheta, \theta) = -\mu(x_i, x_j, \vartheta, \theta)$ и КПД $p^\tau(x_i, x_j, \vartheta, \theta) = 1/p(x_i, x_j, \vartheta, \theta)$, где $\tilde{\delta}(x_i^+, x_j^-, \vartheta^+)$ – поток, истекающего из вершины x_i в период времени ϑ к узлу x_j, $\tilde{\delta}(x_j^-, x_i^+\vartheta + \mu_{ij}^-)$ - поток, поступающий в узел x_j в период времени $\vartheta + \mu_{ij}$ из узла x_i.

Правило 3 – составление периодических графов в сети.

Стоит отметить возможность повтора итераций бесконечное множество раз, в случае отсутствия заданного ограничения. Следовательно, во избежание возникновения подобной ситуации, задаются ограничения во времени (тактах), к примеру: $t = 6$. Данное ограничение предполагает окончание прохождения через $t = 6$, соответственно, при достижении данного значения, необходимо сформировать максимально возможное количество выходных значений равных U. Данный периодический темпоральный граф с прохождением тактов с задержкой в вершине, представляется следующим образом: пусть в графе G вершины x_i и x_j соединяются дугой (x_i, x_j), при условии, что в исходном графе $G = (X, A)$, пропускная способность x_i в вершину x_j за время равное $t = 1$, равна q. Например, пропускная способность вершины x_i по дуге (x_i, x_j) в момент времени 1 обладает значением, равным 3, способна представляться в графе G единицей потока, проходящей по дуге (x_1, x_2).

При этом мы видим, что в вершине x_2 у нас идет задержка и мы ждем следующего периода, чтобы из этой вершины пойти дальше. При этом пропускные способности дуг в развернутом во времени графе с учетом задержек, задаются в соответствии с уравнениями (7)-(9):

$$\tilde{q}(x_{0_i}, x_{1_{i+1}}) = \left[q_1(x_{0_i}, x_{1_{i+1}}); q_r(x_{0_i}, x_{1_{i+1}})\right] = [q_1(x_0, x_1); q_r(x_0, x_1)],$$
$$\forall_i = 0, \dots, p-1, \tag{7}$$
$$\tilde{q}(x_{1_i}, x_{2_{i+1}}) = \left[q_1(x_{1_i}, x_{2_{i+1}}); q_r(x_{1_i}, x_{2_{i+1}})\right] = [q_1(x_1, x_2); q_r(x_1, x_2)],$$
$$\forall_i = 0, \dots, p-1;$$

$$\tilde{q}(x_{1_i}, x_{2_{i+1}}) = \begin{cases} U_i + U_{i+1}, \text{если } \tilde{q}(x_{1_i}, x_{2_{i+1}}) > \max \tilde{q} \\ U_i, \text{если } \tilde{q}(x_{1_i}, x_{2_{i+1}}) \leq \max \tilde{q} \end{cases}, \quad (8)$$

$$t+1$$

$$\tilde{q}(x_{2_i}, x_{3_{i+1}}) = [q_1(x_{2_i}, x_{3_{i+1}}); q_r(x_{2_i}, x_{3_{i+1}})] = [q_1(x_2, x_3); q_r(x_2, x_3)],$$

$$\forall_i = 0, \dots, p-1. \quad (9)$$

Исходя из формулы (8) стоит внести пояснения: в случае если пропускная способность вершины 2 больше максимальной пропускной способности, осуществляется задержка, с последующим аккумулированием выходных значений и с переходом на новую итерацию цикла. В случае, если пропускная способность вершины окажется меньше или равной максимальному значению, итерация цикла продолжается.

Далеемы можем ввести предложенный алгоритм нахождения максимального динамического потока в обобщенной сети.

3. Алгоритм нахождения максимального динамического потока

Этап 1.

Переход к развернутой во времени версии $\tilde{G}_p$ данного ориентированного графа $\tilde{G}$ путем создания отдельной копии каждого узла $x_i \in X$ в каждый момент времени $\vartheta \in T$. В соответствии с правилом 1 цель состоит в определении максимального потока в динамической обобщенной нечеткой сети, проходящего из набора источников, расширенный на p периодов и входящий в набор приемников, расширенных на p периодов, не позднее p. Мы должны добавить искусственный источник s' и приемник t' и соединиться с каждым заданным источником и приемником соответственно. Искусственные дуги, выходящие из искусственных узлов и входящие в них, имеют бесконечную пропускную способность.

Этап 2.

Создаём нечеткую остаточную сеть $\tilde{G}_p^\tau$ с учетом значений потока, проходящих по дугам графа $\tilde{G}_p$. Нечеткая остаточная сеть $\tilde{G}_p^\tau = (X_p^\tau, \tilde{A}_p^\tau)$основана на "развернутом во времени" графе $\tilde{G}_p = (X_p, \tilde{A}_p)$ и значения расхода, полученные по правилу 2.

Этап 3.

Найдем путь $\tilde{P}_p^\tau$ с наибольшим произведением коэффициента эффективности

$$r = \prod_{(x_i, x_j, \vartheta, \theta) \in A_f} \tilde{p}^\tau(x_i, x_j, \vartheta, \theta) \times \prod_{(x_i, x_j, \vartheta, \theta) \in A_l} 1/\tilde{p}^\tau(x_i, x_j, \vartheta, \theta)$$

Перейти к этапу 4, если путь $\tilde{P}_p^\tau$ найден. Если такого пути нет, максимальное значение расхода $\tilde{\delta}(x_i, x_j, \vartheta, \theta) + \tilde{\varepsilon}_p^\tau \times \tilde{P}_p^\tau = \tilde{b}(p)$ найдено на расширенном во времени нечетком статическом графе, затем перейдите к шагу 6.

Этап 4.

Передаем максимальное значение расхода в зависимости от дуги с минимальной нечеткой остаточной емкостью выдающего потока и с учетом следующих коэффициентов эффективности:

$$\tilde{\varepsilon}_p^\tau = \min\left[\tilde{u}(\tilde{P}_p^\tau)\right], \tilde{u}(\tilde{P}_p^\tau) = \min\left[\tilde{u}^\tau(x_i, x_j, \vartheta, \theta) \times p_{ij}\right], (x_i^\tau, \vartheta), (x_j^\tau, \theta) \in \tilde{P}_p^\tau$$

принимая во внимание коэффициенты эффективности в начале и в конце дуги. Следовательно, полученное значение расхода может быть рассчитано как произведение переданного значения расхода $\tilde{\delta}(x_i^+, x_j^-, \vartheta^+)$ и r:

$$\tilde{\delta}(x_j^-, x_i^+, \vartheta + \mu_{ij}^-) = \tilde{\delta}(x_i^+, x_j^-, \vartheta^+) \times \prod_{(x_i, x_j, \vartheta, \theta) \in A_f} \tilde{p}^\tau(x_i, x_j, \vartheta, \theta) \times \prod_{(x_i, x_j, \vartheta, \theta) \in A_l} 1/\tilde{p}^\tau(x_i, x_j, \vartheta, \theta) \cdot$$

Этап 5.

Обновляем значения нечеткого потока на графике $\tilde{G}_p$: замените нечеткий поток для дуг $\tilde{\delta}\left(x_j^-, x_i^+, \vartheta + \mu_{ij}^-\right)$ на $\tilde{\delta}\left(x_j^-, x_i^+, \vartheta + \mu_{ij}^-\right) - \tilde{\varepsilon}_p^\tau$, соединяющих узел времени $\left(x_i^\tau, \vartheta\right)$ с $\left(x_j^\tau, \theta\right)$, таких как $\left(\left(x_i^\tau, \vartheta\right), \left(x_j^\tau, \theta\right)\right) \notin \tilde{A}_p, \left(\left(x_i^\tau, \vartheta\right), \left(x_j^\tau, \theta\right)\right) \in \tilde{A}_p^\tau$ в $\tilde{G}_p^\tau$ вдоль соответствующих дуг, идущих от $\left(x_i^\tau, \vartheta\right)$ в $\left(x_j^\tau, \theta\right)$ от $\tilde{G}_p$ и замените нечетких поток $\tilde{\delta}\left(x_i^+, x_j^-, \vartheta^+\right)$ на $\tilde{\delta}\left(x_j^-, x_i^+, \vartheta + \mu_{ij}^-\right) + \tilde{\varepsilon}_p^\tau$ для дуг, соединяющих узел-время пара $\left(x_i^\tau, \vartheta\right)$ с $\left(x_j^\tau, \theta\right)$, таких как $\left(\left(x_i^\tau, \vartheta\right), \left(x_j^\tau, \theta\right)\right) \in \tilde{A}_p^\tau$ в $\tilde{G}_p^\tau$ вдоль дуги происходит от $\left(x_i^\tau, \vartheta\right)$в $\left(x_j^\tau, \theta\right)$ от $\tilde{G}_p$. Заменить значение нечеткого потока в $\tilde{G}_p$: $\tilde{\delta}\left(x_i, x_j, \vartheta, \theta\right) \to \tilde{\delta}\left(x_i, x_j, \vartheta, \theta\right) + \tilde{\varepsilon}_p^\tau \times \tilde{P}_p^\tau = \tilde{b}(p)$ и перейдите к этапу 2, начиная с обновленных значений расхода.

Этап 6.

Переход к заданному динамическому нечеткому обобщенному графу $\tilde{G}$, если максимальное поток $\tilde{\delta}\left(x_i, x_j, \vartheta, \theta\right) + \tilde{\varepsilon}_p^\tau \times \tilde{P}_p^\tau = \tilde{b}(p)$ находится в $\tilde{G}_p$. Следовательно, максимальный поток значения $\tilde{b}(p)$ получен на исходном графике, что эквивалентно потоку из источников (начальный узел, расширенный на p периодах) к приемникам (конечный узел, расширенный на p периодах) в $\tilde{G}_p$ после отмены искусственных узлов. Каждый путь, соединяющий пары узел-время $\left(s, c\right)$ и $\left(t, \psi = c + \mu_{st}\left(c\right)\right), \psi \in T$, по которому проходит поток $\delta\left(s, t, \theta, c\right)$, равен потоку $\delta_{st}\left(\theta\right)$.

Этап 7.

После того, как мы перешли к данному динамическому нечеткому графу $\tilde{G}_p$, мы переходим к его периодической форме. Для этого, зная правила 2 и 3, мы корректируем наши значения периодичности в нашем графике. На выходе этого шага мы должны получить динамический нечеткий периодический граф.

4. Числовой пример. В этом разделе приводится расчет числового примера, отражающего предлагаемый метод. Необходимые данные для задачи взяты из геоинформационной системы «Elshayal Smart GIS».

Здесь можно создавать новые слои, которые фокусируются на точках, полилиниях, полигонах и множествах точек, а также на комбинациях этих элементов. Можно автоматически сбалансировать изображение, перейти к слою пользовательского масштаба в Google Земля, подтвердить границы слоя и отредактировать таблицу данных слоя.

Кроме того, можно использовать функции отмены и повтора, вставлять и привязывать вершины, поворачивать или масштабировать изображение, находить элементы по данным, разделять полигоны или полилинии, объединять несколько частей в одну, переключать режим 3D-просмотра, а также вычислять площадь и объем фигуры и многое другое.

Пользователь может выбрать часть карты для обработки и получить информацию. Постановка задачи состоит в том, чтобы найти максимальное значение потока в исходной динамической обобщенной сети в виде. Мы получим два значения максимального потока: для отправленного потока и для полученного потока, а также рассмотрим обобщенную сеть.

Пропускная способность дуг определяется экспертами и включается в ГИС в виде таблиц или в виде понятных исходных данных. Рассмотрим решение задачи нахождения максимального потока в динамической сети. Представим выбранную область в виде нечеткого периодического обобщенного графа, как показано на рис. 2.

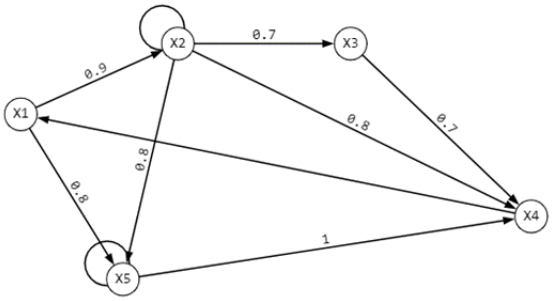


Рис. 2. Исходный периодический нечеткий обобщеный граф

Перейдем к статическому графу с развернутого во времени в соответствии с правилом 2 (рис. 3). Добавим частоту системы к этой сети и получим рис. 4. В табл. 1 и 2 приведены параметры, присвоенные дугам сети.

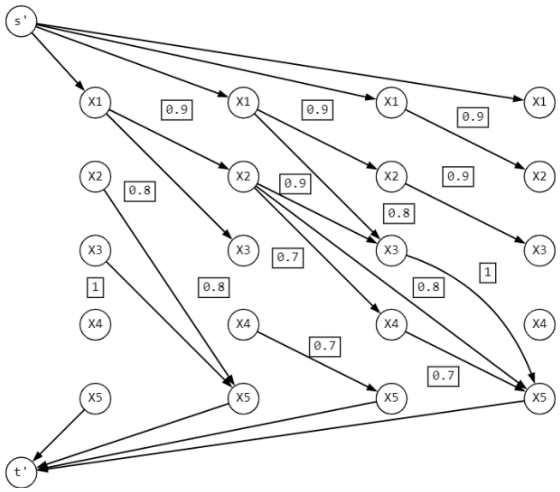


Рис. 3. Нечеткая развернутая во времени обобщенная сеть

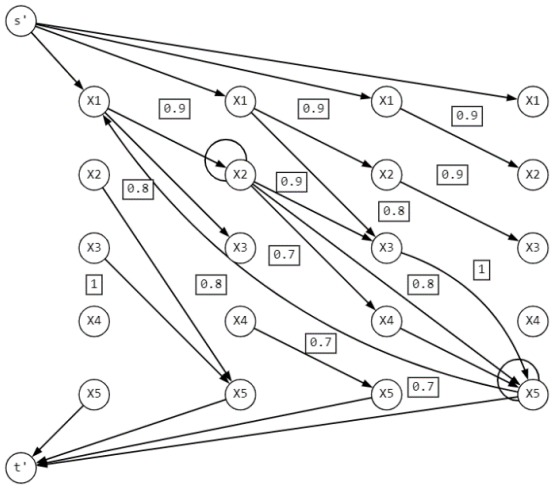


Рис. 4. Нечеткая периодическая развернутая во времени обобщенная сеть

Таблица 1

Нечеткие мощности дуги с коэффициентами эффективности, зависящими от времени задержки потока

Дуги графа	Нечеткие мощности дуги, в периодическом времени			
	0	1	2	3
(x_1, x_2)	40	20	25	40
(x_1, x_3)	16	12	55	37
(x_2, x_3)	18	25	55	35
(x_2, x_4)	15	30	35	18
(x_2, x_5)	42	10	38	16
(x_3, x_5)	36	45	18	55
(x_4, x_5)	20	22	19	26

Мы найдем путь с максимальным произведением эффективности потока и перенесем по нему минимальный расход в соответствии с коэффициентами производительности дуги и эффективности. Давайте построим остаточную сеть в соответствии с уже пройденным потоком.

Таблица 2

Временные параметры в зависимости от времени выхода потока

Дуги графа	Временные параметры во временных периодах			
	0	1	2	3
(x_1, x_2)	1	1	1	2
(x_1, x_3)	1	1	3	2
(x_2, x_3)	5	1	1	1
(x_2, x_4)	4	1	3	1
(x_2, x_5)	1	4	2	2
(x_3, x_5)	1	3	1	1
(x_4, x_5)	5	1	1	3

Итеративно повторяя следующую процедуру, мы получаем окончательное распределение потока в сети, поскольку последующее построение сети не показывает дополнительных путей.

Наконец, максимальная скорость потока, передаваемая от источника, составляет 52 единицы, а скорость потока, принимаемая приемником, составляет 34,2 единицы.

Предложенный алгоритм является оптимальным, поскольку на каждом шаге метода находится насыщенный путь, а на заключительном шаге метода нет расширяющегося пути.

Заключение. Рассмотрен метод нахождения максимального потока в нечеткой динамической обобщенной сети, основанный на предложенных правилах построения расширенной во времени сети с нечеткими периодическими графами. Данный метод учитывает параметры эффективности, присвоенные сетевым дугам, нечеткость и периодичность, свойственную реальным типам сетей из-за различных факторов, связанных с деятельностью человека, условиями окружающей среды, авариями и т.д. Рассмотрен динамический характер пропускных способностей дуг и их зависимостей от времени, периодичности и циклов отправления потока. Представлен численный пример. Рассматриваемый метод имеет значительную практическую ценность, поскольку позволяет решать реальные оптимизационные задачи.

Подтверждения. Исследование было профинансировано Российским научным фондом, проект №23-21-00206, https://rscf.ru/en/project/23-21-00206 / реализовано Южным федеральным университетом.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Kaufman A., Gupta M.M.* Introduction to Fuzzy Arithmetic: Theory and Applications. New York: Van Nostrand Reinhold, 1985.
2. *Inuiguchi M., Romik J.* Probabilistic linear programming: A brief overview of fuzzy mathematical programming and comparison with stochastic programming in portfolio selection, *Fuzzy Sets and Systems,* 2000, Vol. 111, No. 1, pp. 3-28.
3. *Buckley J., Fearing T.* Fuzzy Mathematics in Economics and Finance. Heidelberg: Physica-Verlag, 2000.
4. *Zadeh L.A.* Fuzzy sets, *Information and control,* 1965, Vol. 8, No. 3, pp. 338-353.
5. *Goldberg D.E.* Genetic algorithms in search, optimization and machine learning. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 1989.
6. *Holland J.* Adaptation in Natural and Artificial Systems: An Introductory Analysis with applications to biology, Management and Artificial Intelligence. University of Michigan Press, Ann Arbor, Michigan, 1975.
7. *Kureichik V, Gerasimenko E* Approach to the minimum cost flow determining in fuzzy terms considering vitality degree, *Advances in Intelligent Systems and Computing,* 2017, 573, pp. 200-210. DOI: 10.1007/978-3-319-57261-1_20.
8. *Bozhenyuk A.V., Gerasimenko E.M., Kacprzyk J., and Rozenberg I.N.* Flows in Networks Under Fuzzy Conditions. Springer International Publishing, Switzerland, 2017.
9. *Bozhenyuk A., Kosenko O., Knyazeva M., and Dolgiy A.* The Comparative Approach to Solving Temporal-Constrained Scheduling Problem Under Uncertainty, *Lecture Notes in Computer Science,* 2021, Vol. 13068, pp. 173-183.
10. *Bozhenyuk A., Belyakov S., Gerasimenko E., and Savelyeva M.* Fuzzy optimal allocation of service centers for sustainable transportation networks service, *Intelligent Systems Reference Library,* 2017, Vol. 113, pp. 415-437.
11. *Gorbachev S., Bozhenyuk A., Nikashina P.* Optimization of Traffic Flow Based on Periodic Fuzzy Graphs. Springer Tracts in Human-Centered Computing (STHC), 2023. – 374-383. Available at: https://doi.org/10.1007/978-981-99-3478-2_32.
12. *Knyazeva M., Bozhenyuk A., Kaymak U.* Managing temporal uncertainty in multi-mode Z-number fuzzy graph structures, *Proceedings of the 2019 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology,* 2019, Vol. 1, pp. 580-587. Available at: https://doi.org/10.2991/eusflat-19.2019.80.
13. *Çakır E., Ulukan Z., Acarman T.* Shortest Fuzzy Hamiltonian Cycle on Transportation Network Using Minimum Vertex Degree and Time-dependent Dijkstra's Algorithm, *IFAC-PapersOnLine,* 2021, 54 (2), pp. 348-353. Available at: https://doi.org/10.1016/j.ifacol.2021.06.048.
14. *Wayne K.D., Fleische L.* Faster approximation algorithms for generalized flow, *Proceedings of 10th Annual ACM-SIAM Symposium on Discrete Algorithms*. Baltimore, MD, USA, 1998.
15. *Eguchi A., Fujishige S., Takabatake T.* A polynomial-time algorithm for the generalized independent-flow problem, *J. Oper. Res*., 2004, 47 (1), pp. 1-17.
16. *Krumke S.O., Zeck C.* Generalized max flow in series–parallel graphs, *Discrete Optimization*, 2013, 10 (2), pp. 155-162. Available at: https://doi.org/10.1016/j.disopt.2013.01.001.
17. *Kartikasari R.* Optimization of Traffic Light Control Using Fuzzy Logic Sugeno Method, International *Journal of Global Operations Research,* 2020, pp. 51-61.
18. *Belyakov S., Bozhenyuk A., Samoylov L., Nikashina P.* Geoinformation Model for Smart Grid Sustainability Management, *Intelligent and Fuzzy Systems. INFUS 2023. Lecture Notes in Networks and Systems*, 2023, Vol. 759, pp. 651-658.
19. *Umare P.R.; Jayswal S.G.; Tambakhe S.R.; Upadhye P.D.; Gulhane N.D.* Smart Solution for Traffic Control, *In Proceedings of the 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, 2019, pp. 721-724.
20. *Kutlimuratov K., Khakimov S., Mukhitdinov A., Samatov R.* Modelling trffic flow emissions at signalized intersection with PTV vissim, *Mezhdunarodnaya nauchnaya konferentsiya «Stroitel'naya mekhanika, gidravlika i gidrotekhnika»* [International scientific conference "Structural mechanics, hydraulics and hydraulic engineering"], 2021, pp. 12.

Статью рекомендовал к опубликованию д.т.н., профессор Н.Е. Сергеев.

Никашина Полина Олеговна – Южный федеральный университет; e-mail: nikashina@sfedu.ru; г. Таганрог, Россия; тел.: +79381453823; кафедра информационно-аналитических систем безопасности имени профессора Берштейна Леонида Самойловича; аспирант.

Nikashina Polina Olegovna – Southern Federal University; e-mail: nikashina@sfedu.ru; Taganrog, Russia; phone: +79381453823; the department of Information and Analytical Security Systems named after Professor Bershtein Leonid Samoylovich; graduate student.

УДК 007:51 DOI 10.18522/2311-3103-2024-4-40-49

Ф.А. Хуссейн

РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДА ЦЕНТРАЛИЗОВАННОГО РАСПРЕДЕЛЕНИЯ ЗАДАЧ В МУЛЬТИАГЕНТНЫХ СИСТЕМАХ*

Проводится анализ задачи мульти-коммивояжера, которая является расширенной версией классической задачи коммивояжера. В отличие от последней, задача мульти-коммивояжера предполагает участие нескольких коммивояжеров, каждый из которых должен посетить определенное количество городов ровно один раз и вернуться в исходную точку, при этом минимизируя затраты на путешествие. Задача мульти-коммивояжера представляет значительный интерес в области оптимизации маршрутов и распределения задач между несколькими агентами. Основная цель исследования – разработка эффективного метода решения этой задачи, который сократит время выполнения и оптимизирует использование ресурсов. В рамках исследования был разработан метод, который базируется на сокращении размерности пространства решений. Данный метод позволяет более эффективно распределять нагрузку и управлять ресурсами, что в конечном итоге способствует сокращению общего времени выполнения задач. Одной из ключевых особенностей предлагаемого метода является его универсальность и адаптивность к различным сценариям, включая ситуации с различным количеством задач и коммивояжеров. Также проводилось исследование предложенного метода с точки зрения влияние его гиперпараметров (коэффициент испарения феромона, количество итерации, количество муравьёв) на качества решения и время расчета. Для оценки эффективности предлагаемого метода было проведено сравнительное исследование с использованием классического метода решения задачи мульти-коммивояжера. Оценка результатов проводилась по трем основным критериям: время вычислений для решения задачи мульти-коммивояжера, суммарная длина пройденных маршрутов и максимальная длина маршрута среди всех коммивояжеров. Анализ экспериментальных данных показал, что разработанный метод значительно превосходит классический по всем ключевым показателям. В частности, среднее время вычислений для решения задачи мульти-коммивояжера уменьшилось на 52% по сравнению с лучшими известными классическими результатами. Кроме того, средняя суммарная длина пройденных маршрутов уменьшилась на 12%, а максимальная длина маршрута (показатель дисбаланса нагрузки) снизилась на 14%. Эти результаты подтверждают высокую эффективность предложенного метода и его перспективность для практического применения в различных сферах, требующих оптимизации маршрутов и распределения задач между несколькими исполнителями.

Задача мульти коммивояжера; распределение задач; целераспределение; мультиагентные системы; централизованное управление; групповое управление.

F.A. Houssein

DEVELOPMENT AND STUDY OF A CENTRALIZED TASK ALLOCATION METHOD IN MULTI-AGENT SYSTEMS

This study provides a comprehensive analysis of the multi-traveling salesman problem, which is an extended version of the classical traveling salesman problem. In contrast to the latter, the multi-traveling salesman problem involves the participation of several traveling salesmen, each of whom must visit a certain number of cities exactly once and return to the starting point, while minimizing travel costs. The mul-

* Исследование выполнено при финансовой поддержке фонда содействия инновациям в рамках выполнения программы «УМНИК» по договору № 19041ГУ/2023 от 21.02.2024».

ti-traveling salesman problem is of significant interest in the field of route optimization and task distribution among multiple agents. The main goal of the research is to develop an effective method for solving this problem, which will reduce execution time and optimize the use of resources. As part of the study, an innovative method was developed, which is based on reducing the dimension of the solution space. This method allows you to more effectively distribute the load and manage resources, which ultimately helps reduce the overall time to complete tasks. One of the key features of the proposed method is its versatility and adaptability to various scenarios, including situations with varying numbers of tasks and traveling salespeople. A detailed study of the proposed method was also carried out from the point of view of the influence of its hyperparameters (pheromone evaporation coefficient, number of iterations, number of ants) on the quality of the solution and calculation time. To evaluate the effectiveness of the new method, a comparative study was conducted using the classical method for solving the multi-traveling salesman problem. The results were assessed according to three main criteria: the computation time for solving the multi-traveling salesman problem, the total length of the routes traveled, and the maximum route length among all traveling salesmen. Analysis of experimental data showed that the developed method significantly exceeds the classical one in all key indicators. These results confirm the high efficiency of the proposed method and its promise for practical application in various fields that require optimizing routes and distributing tasks among several performers. Thus, the study demonstrates that the developed method has significant potential for improving routing and resource allocation processes. Its application can significantly increase efficiency in various areas where coordination of the work of several agents is necessary, such as logistics, transport systems and other areas related to route optimization.

Multi traveling salesman problem; task distribution; target distribution; multi-agent systems; centralized management; group management.

Введение. Суть целераспределения является сопоставление задач агентам таким образом, чтобы минимизировать конфликты и максимизировать общее качество выполнения. Каждая задача должна быть назначена не более чем одному агенту, и каждый агент может выполнить не более определенного числа задач. Распределение считается завершенным после назначения определенного минимального числа задач. Глобальная целевая функция зависит от локальных функций, которые отражают задачи, стоящие перед каждым агентом.

Когда каждый агент должен выполнить несколько задач, возникает проблема оптимального распределения задач между агентами, а также определения оптимальной последовательности выполнения задач для каждого агента. Эта проблема эквивалентна задаче множественного коммивояжера, где целью является минимизация суммарного расстояния, пройденного всеми коммивояжерами, и/или максимального расстояния, пройденного одним из коммивояжеров.

Существуют различные подходы к организации распределения задач, которые могут быть классифицированы в соответствии с организационной парадигмой командной работы. Эта парадигма определяет структуру и взаимодействие между многочисленными агентами в системе, а также их конкретные роли. В литературе выделяются две основные организационные парадигмы: централизованная и децентрализованная.

Децентрализованная парадигма в многоагентных системах представляет собой процесс распределения административных задач и полномочий между агентами без централизованного контроля. В такой конфигурации каждый агент действует независимо, принимая решения на основе своей локальной информации, без учета других агентов. Однако иногда требуется обмен информацией между агентами для эффективного выполнения общей миссии.

Существует множество децентрализованных подходов к решению задачи распределения. Например, в работе [1] предложена децентрализованная реализация венгерского метода, а в [2] представлены алгоритмы аукционов на основе консенсуса для решения этой задачи. Также существуют децентрализованные методы, основанные на эволюционных вычислениях, как описано в [3], а также иерархический рыночный подход, предложенный в [4].

Однако у децентрализованных систем есть свои ограничения. Например, они часто страдают от отсутствия глобальной перспективы, что может привести к неоптимальным решениям. Агенты принимают решения, исходя из локальной информации, что может

снизить эффективность распределения задач. Кроме того, частая связь между агентами для обмена информацией и координации может привести к увеличению сетевого трафика и задержкам, особенно в крупномасштабных системах.

Для преодоления ограничений, связанных с децентрализованными подходами к распределению задач, часто применяют централизованные подходы. В таких системах каждый агент поддерживает связь с центральным агентом, который координирует распределение задач между агентами. Это означает, что агенты отправляют всю свою информацию центральному агенту, который затем анализирует её и отправляет инструкции каждому агенту для выполнения задач.

Основное преимущество централизованного подхода заключается в том, что он позволяет избежать дублирования усилий и ресурсов, а также экономит затраты и время. В то же время, несмотря на то, что централизованные системы широко представлены в литературе, они могут столкнуться с ограничениями масштабируемости. Управление распределением и координацией ресурсов для большого числа агентов может быть сложным и ресурсоемким, что может привести к проблемам с производительностью и масштабируемостью.

Централизованный подход является одним из наиболее широко рассматриваемых в литературе подходов к решению проблем распределения задач. Однако важно учитывать его ограничения, особенно при увеличении количества агентов или задач, чтобы избежать потенциальных проблем с производительностью и масштабируемостью. Для этого предлагается новый метод решения задачи целераспределения на основе уменьшения размера пространство решений.

Задача коммивояжера (Travelling Salesman Problem, TSP) является классической NP-трудной задачей комбинаторной оптимизации [5]. Она формулируется следующим образом: имея список городов и расстояния между каждой парой городов, необходимо найти кратчайший маршрут, который позволяет посетить каждый город ровно один раз и вернуться в исходный город. В отличие от широко исследуемой задачи коммивояжера, задача мульти-коммивояжеров (Multiple Travelling Salesmen Problem, MTSP) не привлекает столько внимания и не так интенсивно изучается. В MTSP несколько коммивояжеров должны посетить заданное количество городов ровно один раз и вернуться в исходную точку с минимальными затратами на поездку. MTSP тесно связана с другими задачами оптимизации, такими как проблема маршрутизации транспортных средств (Vehicle Routing Problem, VRP) [6] и проблема назначения задач [7]. Фактически, MTSP является упрощенной версией VRP, не учитывающей вместимость транспортных средств и требования клиентов. Тем не менее, MTSP можно использовать для моделирования многих реальных приложений, таких как производство и распространение газет [8, 9], планирование горячей прокатки [10, 11], планирование работы причальных кранов [12], и военные поисково-спасательные операции [13]. Поскольку MTSP является NP-полной задачей [14], поиск эффективных методов и разработка оптимальных алгоритмов для ее решения имеют большое практическое значение.

В последние годы интеллектуальные вычислительные технологии привлекают все больше внимания исследователей. Многие из этих технологий вдохновлены природными явлениями и пытаются использовать преимущества различных аспектов поведения живых организмов на разных уровнях их организации.

Харрат и коллеги [15] предложили гибридный алгоритм AC2OptGA для решения задачи MTSP, комбинируя генетический алгоритм (GA) и муравьиный алгоритм. В этом алгоритме муравьиная колония генерировала решения, а GA улучшал их. Гомес и соавторы [16] применили GA для задачи выбора маршрута транспортных средств (VRP) с целью оптимизации ежедневных маршрутов работников компании в Ковильяне, Португалия, минимизируя стоимость и расстояние.

Акбай и Калайчи [17] разработали решение на основе алгоритма поиска переменных окрестностей для задачи MTSP с учетом сбалансированной стоимости. Они использовали 22 набора данных разных масштабов для своих вычислений. Муньос-Эррера и

Сучан [18] исследовали фитнес-ландшафт (FLA) для задач MTSP и CVRP, предложив новую меру FLA для анализа структуры фитнес-ландшафта и взаимосвязей с производительностью алгоритмов.

Сюй и Чжан [19] реализовали гибридный алгоритм сбалансированного MTSP, основанный на генетических алгоритмах и методе локального поиска (two-opt). Хуфи и коллеги [20] провели обзор литературы по задачам оптимизации траектории беспилотных летательных аппаратов (БПЛА), связанным с задачами коммивояжера и VRP, предоставив обзор методов и показателей производительности.

Перейра и соавторы [21] разработали гибридный алгоритм GABC-LS для решения минимального MTSP, сочетая генетический алгоритм, пчелиный алгоритм и два метода локального поиска (move1-inside и two-opt).

Методика решения задач, основанная на поведении муравьев при поиске пищи (оптимизация колонии муравьев – ACO), доказала свою эффективность в различных задачах оптимизации [22, 23]. Оригинальный ACO был модифицирован для применения в разных областях, таких как задачи назначения, раскраска графов, планирование, проектирование схем, сети связи, биоинформатика и маршрутизация транспортных средств. Были разработаны расширения ACO, такие как элитарная AS [24], Ant-Q [25], Ant Colony System [26], ранговая AS [27], популяционная ACO [28] и другие.

Объём пространства решений задачи MTSP с n городами и m коммивояжёрами можно вычислить с помощью числа Лаха. Это коэффициент, который выражает возрастающие факториалы через падающие факториалы и наоборот. Числа Лаха были открыты Иво Лахом в 1954 году [29]. В комбинаторике беззнаковые числа Лаха имеют важное значение: они подсчитывают количество способов разделения множества из n элементов на m непустых линейно упорядоченных подмножеств. Беззнаковые числа Лаха рассчитываются по следующей формуле:

$$L(n,m) = \left(\frac{n!}{m!}\right)^2 * \frac{m}{n(n-m)!}. \quad (1)$$

Предложенный метод. Предложенный метод реализуется в несколько этапов:

Шаг 1: Первоначально решается задача коммивояжёра, где предполагается, что количество агентов равно одному. Это позволяет определить оптимальный маршрут, который должен пройти один агент, чтобы посетить все задачи и вернуться в начальную точку.

Шаг 2: После решения задачи коммивояжёра получается оптимальная последовательность выполнения задач. Эта последовательность затем разделяется между несколькими агентами, стараясь распределить задачи между ними как можно более равномерно.

Шаг 3: Далее происходит комбинаторный процесс выбора оптимальных первых и последних задач для каждого агента. Это позволяет определить начальную и конечную точки маршрута для каждого агента таким образом, чтобы минимизировать общее время выполнения всех задач.

Таким образом, этот метод включает в себя два основных алгоритма:

- Алгоритм решения задачи коммивояжёра, который позволяет определить оптимальный маршрут для посещения всех задач и возврата в начальную точку одному агенту.
- Алгоритм разделения решения задачи коммивояжёра на несколько агентов, который распределяет полученный оптимальный маршрут между несколькими агентами, стараясь добиться равномерного распределения задач и оптимального использования ресурсов.

В качестве алгоритма решения задачи коммивояжёра был выбран муравьиный алгоритм – это метаэвристический алгоритм оптимизации, вдохновленный поведением муравьев при поиске пищи. Он используется для решения задачи коммивояжера (TSP) и других задач комбинаторной оптимизации.

Основные шаги муравьиного алгоритма:

1. Инициализация феромонов: Начальные уровни феромонов на ребрах графа устанавливаются на некоторое начальное значение.
2. Выбор маршрутов: Каждый муравей выбирает следующий город для посещения, используя вероятностный метод, основанный на феромонах и эвристических информациях (например, расстояния между городами).
3. Обновление феромонов: После того как все муравьи завершили построение своих маршрутов, феромоны обновляются на основе эффективности маршрутов. Обычно феромоны усиливаются на кратчайших маршрутах, по которым прошли муравьи, а также испаряются с течением времени.
4. Повторение процесса: Шаги 2 и 3 повторяются некоторое количество раз или до достижения критерия останова.
5. Выбор лучшего решения: По завершении всех итераций выбирается лучший найденный маршрут.

Муравьиный алгоритм эффективно исследует пространство решений и обладает свойством самоорганизации благодаря обратной связи через феромоны. Он позволяет избежать застраивания в локальных оптимумах и находить приближенно оптимальные решения для задачи TSP и других комбинаторных задач.

После нахождения оптимального пути для задачи TSP необходимо разделить его между агентами для решения задачи MTSP. Этот процесс начинается с формирования матрицы расстояний между всеми городами, включая промежуточные города по маршруту.

Затем рассчитывается примерная длина маршрута (d*), которую каждый агент должен пройти. Путем перебора городов первого агента находится такой город, расстояние до которого на кусочно-ломаной траектории максимально близко к d*. Эти города становятся центрами для каждого агента.

Далее каждый следующий агент выбирает город, следующий за последним городом предыдущего агента, и этот процесс повторяется до тех пор, пока не будут назначены все агенты. Затем рассчитывается и сохраняется сумма длин маршрутов для всех агентов.

Начальный город первого агента изменяется с целью проверки всех возможных решений, после чего сохраняется решение с наименьшей суммой пройденных расстояний.

Таким образом, этот подход минимизирует суммарное расстояние, пройденное всеми агентами, за счет оптимизации решения задачи TSP и равномерного разделения маршрута TSP на сегменты одинаковой длины для каждого агента. Размер пространства возможных решений составляет (n-1)! + n–1, где n - количество городов в задаче MTSP.

Исследования гиперпараметров алгоритма. Предложенный метод решения задачи MTSP основан на подходе, который использует решение задачи TSP с применением муравьиного алгоритма. Эффективность найденного решения и время его вычисления зависят от настройки набора гиперпараметров муравьиного алгоритма, таких как количество итераций, количество муравьев и коэффициент испарения феромона. Для того чтобы проанализировать влияние изменения значений этих гиперпараметров на оптимальность решения, было проведено исследование. Для этого исследования были выбраны две эталонные задачи TSP: eil51 и kroA100.

Для исследования влияния коэффициента испарения феромона проведен следующий подход: фиксировались значения количества итераций и количества муравьев, и изменялись значения коэффициента испарения феромона. Каждый набор гиперпараметров повторялся 10 раз для получения статистики. Количество итераций составило 100, количество муравьев составило 80. Эксперименты проводились с различными значениями коэффициента испарения феромона в интервале от 0.05 до 0.95 с шагом 0.05. На рис. 1 изображено изменение качества решения задачи TSP при различных значениях коэффициента испарения феромона. Рис. 2 показывает время вычисления решения задачи TSP при различных значениях коэффициента испарения феромона.

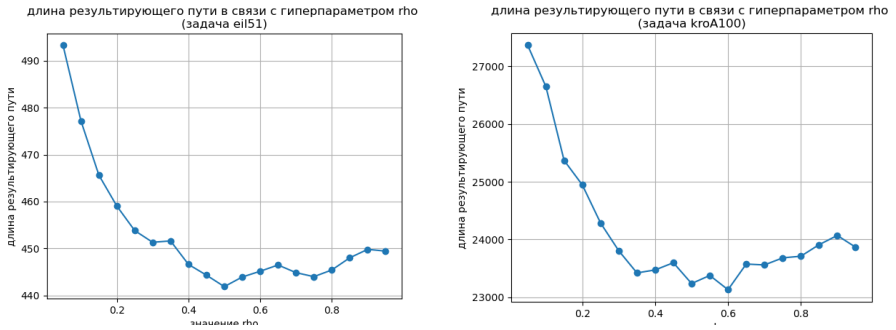


Рис. 1. Изменение качество решения при разных значениях коэффициента rho

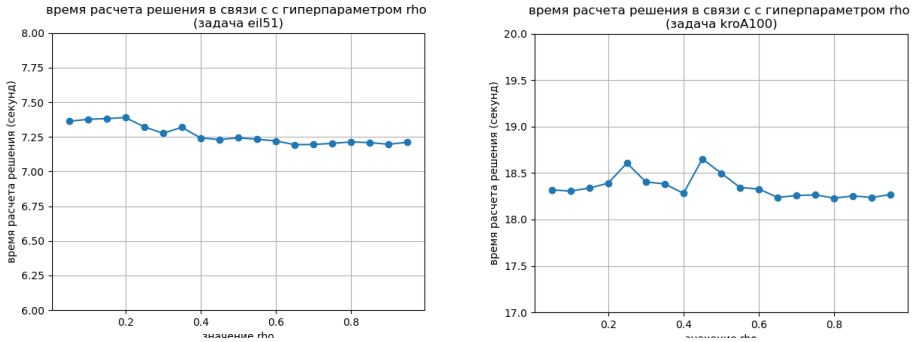


Рис. 2. Время расчета решения при разных значениях коэффициента rho

Обнаружено, что качество решения улучшается при использовании коэффициента испарения феромона (rho) в диапазоне от 0.4 до 0.6. Этот результат объясняется тем, что небольшие значения rho приводят к накоплению феромона, что увеличивает вероятность застраивания в локальных минимумах. С другой стороны, большие значения rho приводят к быстрому испарению феромона, что снижает его влияние на процесс поиска оптимального пути. Также отмечено, что изменения параметра rho практически не влияют на время вычисления решения.

Для изучения влияния количества итераций был проведен следующий подход: значения коэффициента испарения феромона и количество муравьев фиксировались, а количество итераций изменялось. Каждый набор гиперпараметров повторялся 10 раз для получения статистики. Коэффициент испарения феромона составил 0.5, количество муравьев составило 40. Эксперименты проводились с разными значениями количества итераций в интервале от 50 до 500 с шагом 25. На рис. 3 изображено изменение качества решения задачи TSP при различных значениях количества итераций. Рис. 4 демонстрирует время вычисления решения задачи TSP при различных значениях количества итераций.

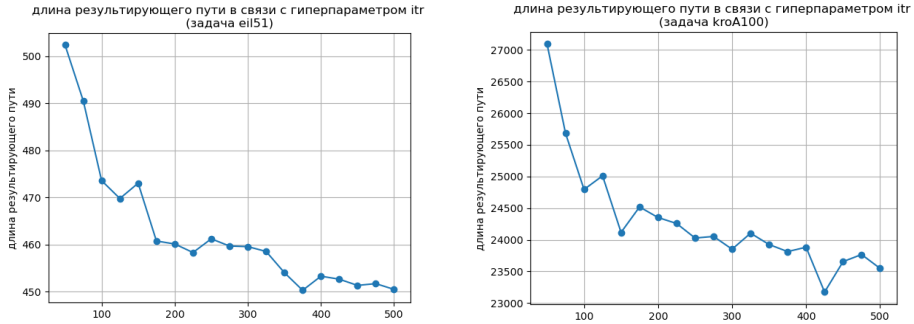


Рис. 3. Изменение качество решения при разных значениях параметра itr

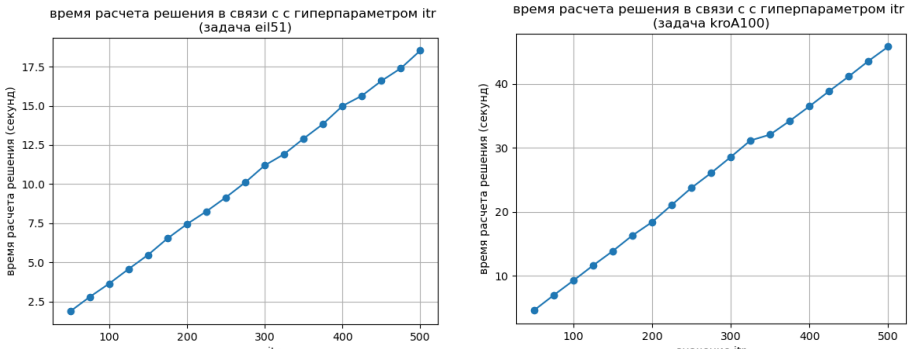


Рис. 4. Время расчета решения при разных значениях параметра itr

Обнаружено, что качество решения улучшается при использовании значения количества итераций (*itr*) в диапазоне от 350 до 450. Этот результат объясняется тем, что маленькие значения параметра недостаточно для обеспечения нахождения оптимального решения, а большие значения могут привести к большей вероятности застраивания в локальных минимумах. Также замечено, что увеличение значения параметра *itr* линейно приводит к увеличению времени вычисления решения.

Для изучения влияния количества муравьев был применен следующий подход: значения количества итераций и коэффициента испарения феромона фиксировались, а количество муравьев изменялось. Каждый набор гиперпараметров повторялся 10 раз для получения статистики. Количество итераций составило 100, коэффициент испарения феромона составил 0.5. Эксперименты проводились с разными значениями количества муравьев в диапазоне от 20 до 100 с шагом 10. На рис. 5 представлено изменение качества решения задачи TSP при различных значениях количества муравьев. Рис. 6 демонстрирует время вычисления решения задачи TSP при различных значениях количества муравьев.

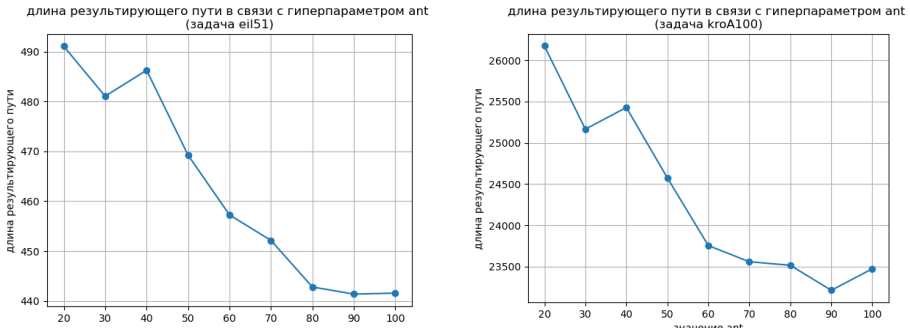


Рис. 5. Изменение качество решения при разных значениях параметра ant

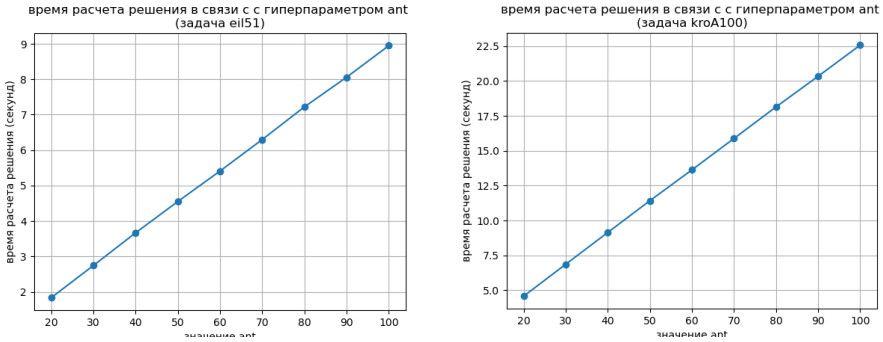


Рис. 6. Время расчета решения при разных значениях параметра ant

Обнаружено, что качество решения улучшается при использовании значения количества муравьев (ant) в диапазоне от 80 до 100. Этот результат объясняется тем, что маленькие значения параметра не достаточны для обеспечения нахождения оптимального решения, в то время как большие значения могут привести к большей вероятности застраивания в локальных минимумах. Также замечено, что увеличение параметра *ant* линейно приводит к увеличению времени вычисления решения.

Подбор оптимального набора гиперпараметров. Для определения оптимальных значений гиперпараметров, которые были рассмотрены в предыдущем разделе, необходимо проверить все их комбинации. Однако, учитывая вычислительную сложность такой задачи, целесообразно сократить размер пространства исследования путем выбора значений гиперпараметров из предварительно определенных диапазонов. Таким образом, были выбраны значения гиперпараметров: коэффициент испарения феромона (rho) в диапазоне [0.4, 0.6], количество итераций (itr) в диапазоне [350, 450], количество муравьев (ant) в диапазоне [80, 100]. Для исследования использовалась эталонная задача eil51. Каждый эксперимент с набором гиперпараметров был проведен 10 раз для получения статистики.

После изучения различных комбинаций гиперпараметров для решения задачи коммивояжера был найден оптимальный набор значений: количество итераций (itr) составляет 350, количество муравьев (ant) - 80, а коэффициент испарения феромона (rho) установлен на уровне 0.4.

Ограничения на применения предложенного метода. Пространство решений для предложенного метода можно оценить, как (n-1)! + n–1, где n – количество городов в задаче MTSP. Однако для традиционных методов объем пространства решений зависит от количества коммивояжеров, что приводит к уменьшению объема пространства при увеличении числа коммивояжеров. Следовательно, существуют случаи, когда использование традиционных методов становится целесообразным. Для определения этих случаев мы используем неравенство (2):

$$\left(\frac{n!}{m!}\right)^2 * \frac{m}{n(n-m)!} < (n-1)! + n-1 \tag{2}$$

$$\Rightarrow \frac{n!\,(n-2)!}{(n-2)!+1} < m!\,(m-1)!\,(n-m)!$$

Чтобы найти наибольшее значение m, удовлетворяющее этому неравенству, нужно задать значение n. Однако точно найти наибольшее m для заданного n может быть сложно из-за сложности факториалов. Численным методом было установлено, что для задачи eil51 количество коммивояжеров не должно превышать 16. Таким образом, можно определить количество коммивояжеров, при котором использование предложенного метода для решения задачи мульти коммивояжёра оправдано, используя неравенство (2).

Результаты и сравнения. Для сравнения используется алгоритм ACO-BMTSP [30], который, как и предложенный метод, базируется на муравьином алгоритме. Оба алгоритма были реализованы на языке программирования Python. Для исследования были подготовлены три эталонные задачи Eil51, KroA100 и KroA150 с 51, 100 и 150 городами последовательно. Для каждой из этих задач были установлены три сценария с 3, 5 и 10 агентами. Каждый сценарий запускался по 100 раз для сбора статистических данных. Критерии оценки результатов включают:

- время расчета решения;
- суммарную длину пройденных агентами путей;
- максимальную длину пути среди всех маршрутов, пройденных агентами.

В табл. 1 выражены преимущества предлагаемого метода в процентах по каждому из критериев оценки, которые рассчитывались следующим образом:

$$advantage_{\%} = \left(1 - \frac{criterion_{proposed}}{criterion_{ACO-BMTSP}}\right) * 100. \tag{3}$$

где $advantage_{\%}$ – преимущества в процентах, $criterion_{proposed}$ – среднее среди 100 экспериментов значения заданного критерия для предлагаемого метода, $criterion_{ACO\text{-}BMTSP}$ – среднее среди 100 экспериментов значения заданного критерия для алгоритма ACO-BMTSP.

Таблица 1

Средний выигрыш в процентах по времени (avg_{time}), суммы длины маршрутов (avg_{sum}), максимальный по длине маршрут среди всех агентов (avg_{max}) при использовании предложенного метода для сценариев eil51, KroA100 и KroA150

Задача	Eil51			KroA100			KroA150		
Кол. агентов	3	5	10	3	5	10	3	5	10
avg_{time} (%)	48.26	49.19	41.06	52.78	54.34	47.81	67.27	55.19	49.19
avg_{sum} (%)	6.07	4.84	2.36	12.57	11.58	12.89	19.15	18.62	16.42
avg_{max} (%)	9.22	10.9	9.39	14.22	12.78	13.31	20.25	18.67	17.18

Сравнительный анализ в итоге продемонстрировал, что применение предложенного метода обеспечивает следующие преимущества:

а) время расчета решения задачи MTSP сокращается в среднем на 52%;

б) общая длина пройденного пути уменьшается в среднем на 12%;

в) максимальная длина маршрута среди всех пройденных агентами (неравномерность нагрузки) сокращается в среднем на 14%.

Заключение. В данной работе проводилось исследование методов решения задачи мульти-коммивояжера, и был предложен новый метод, основанный на уменьшении размера пространства решений. Сравнительный анализ показал, что при использовании предложенного метода в среднем:

- Время расчета решения задачи мульти-коммивояжера сокращается на 52%.
- Общая длина пути уменьшается на 12%.
- Дисбаланс нагрузки уменьшается на 14%.

Также проводилось детальное исследование предложенного метода с точки зрения влияния его гиперпараметров (коэффициент испарения феромона, количество итераций, количество муравьев) на качество решения и время расчета. Оптимальные значения гиперпараметров были определены: коэффициент испарения феромона = 0.4, количество итераций = 350, количество муравьев = 80. Были также выявлены ситуации, когда целесообразно использовать традиционные методы решения задачи мульти-коммивояжера. Для задач с 51 городом и 16 или менее коммивояжерами предпочтительнее использовать предложенный метод.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Giordani S., Lujak M., Martinelli F.* Adistributed algorithm for the multi-robot task allocation problem, *Trends Applied Intelligent System*. Springer, 2010, pp. 721-730.
2. *Han-Lim C., Brunet L., How J.* Consensus-based decentralized auctions for robust task allocation, *IEEE Transactions on Robotics*, 2009, pp. 912-926.
3. *Ping-an G., Zi-xing C., Ling-liY*. Evolutionary computation approach to decentralized multi- robot task allocation, *International Conference on Natural Computation (ICNC)*, 2009, pp. 415-419
4. *Khamis A.M., Elmogy A.M., Karray F.O*. Complex task allocation in mobile surveillance systems, J. *Intell. Robot. Syst.,* 2011, 64, pp. 33-55.
5. *Gutin G., Punnen A.P.* The Traveling Salesman Problem and its Variations. – Springer, Boston, 2007.
6. *Braekers K., Ramaekers K. Nieuwenhuyse I.* The vehicle routing problem: State of the art classification and review, *Comput. Ind. Eng*., 99.
7. *Oncan T.* A survey of the generalized assignment problem and its applications, *Infor*, 2007, 45, pp. 123-141.
8. *Van Buer M.G., Woodru D.L., Olson R.T.* Solving the medium newspaper production/distribution problem, *European Journal of Operational Research*, 1999, 115, pp. 237-253.
9. *Carter A.E., Ragsdale C.T.* Scheduling pre-printed newspaper advertising inserts using genetic algorithms, *Omega*, 2002, pp. 415-421.
10. *Kewei Huang, Dingwei Wang.* Multiple traveling salesman problem and its application to hot rolling planning, *Application Research of Computers,* July 2007, 24 (7), pp. 43-46.

11. *Tang L., Liu J., Ronga A., Yang Z.* A multiple traveling salesman problem model for hot rolling schedule in Shanghai Baoshan Iron and Steel Complex, *European Journal of Operational Research*, 2000, 124 (2), pp. 267-282.
12. *Kim K.H., Park Y.M.* A crane scheduling method for port container terminals, *European Journal of Operational Research*, 2004, 156, pp. 752-768.
13. *Sheldon H. Jacobson, Laura A. McLay, Shane N. Hall, Darrall Henderson, Diane E. Vaughan*. Optimal search strategies using simultaneous generalized hill climbing algorithms, *Mathematical and Computer Modelling*, 2006, 43, pp. 1061-1073.
14. *Evelyn C. Brown, Cliff T. Ragsdale, Arthur E. Carter.* A grouping genetic algorithm for the multiple traveling salesperson problem, *International Journal of Information Technology & Decision Making*, 2007, Vol. 6, No. 2, pp. 333-347.
15. *Harrath Y., Salman A.F., Alqaddoumi A., Hasan H., Radhi A.* A novel hybrid approach for solving the multiple traveling salesmen problem, *Arab. J. Basic Appl. Sci*., 2019, pp. 103-112.
16. *Gomes D.E., Iglésias M.I.D., Proença A.P., Lima T.M., Gaspar P.D.* Applying a Genetic Algorithm to a m-TSP: Case Study of a Decision Support System for Optimizing a Beverage Logistics Vehicles Routing Problem. Electronics, 2021.
17. Akbay M.A., Kalayci C.B. A Variable Neighborhood Search Algorithm for Cost-Balanced Travelling Salesman Problem, *Proceedings of the Metaheuristics Summer School, Taormina, Italy, 2018*.
18. *Muñoz-Herrera S.; Suchan K.* Constrained Fitness Landscape Analysis of Capacitated Vehicle Routing Problems, *Entropy*, 2022, pp. 53.
19. *Xu H.L., Zhang C.M.* The research about balanced route MTSP based on hybrid algorithm, *Proceedings of the 2009 International Conference on Communication Software and Networks, Chengdu, China, 2009,* pp. 533-536.
20. *Khoufi I., Laouiti A., Adjih C.* A Survey of Recent Extended Variants of the Traveling Salesman and Vehicle Routing Problems for Unmanned Aerial Vehicles, *Drones*, 2019, pp. 66.
21. *de Castro Pereira S., Solteiro Pires E.J., de Moura Oliveira P.B.* A Hybrid Approach GABC–LS to Solve MTSP, *Proceedings of the Optimization, Learning Algorithms and Applications*. Springer International Publishing: Cham, Switzerland, 2022*,* pp. 520-532.
22. *Dorigo M., Birattari M., Stutzle T.* Ant colony optimization, *IEEE Comput. Intell. Mag*., 2006, pp. 28-39.
23. *Zhan S.C., Xu J., Wu J.* The optimal selection on the parameters of the ant colony algorithm, *Bull. Sci. Technol*, 2003, pp. 381-386.
24. *Dorigo M.* Optimization, Learning and Natural Algorithms. Ph.D. Thesis, Politecnico di Milano, Milano, Italy, 1992.
25. *Gambardella L.M., Dorigo M.* Ant-Q: A reinforcement learning approach to the traveling salesman problem, *Machine Learning Proceedings 1995.* Elsevier: Amsterdam, The Netherlands, 1995, pp. 252-260.
26. *Dorigo M., Gambardella L.M.* Ant colonies for the travelling salesman problem, *Biosystems*, 1997, pp. 73-81.
27. *Bullnheimer B., Hartl R., Strauss C.* A New Rank Based Version of the Ant System—A Computational Study, *Cent. Eur. J. Oper. Res*., 1999, pp. 25-38.
28. *Guntsch M., Middendorf M.* A population based approach for ACO, *Proceedings of the Workshops on Applications of Evolutionary Computation, Kinsale, Ireland, 3–4 April 2002.* Springer: Berlin/Heidelberg, Germany, 2002, pp. 72-81.
29. *Lah Ivo.* A new kind of numbers and its application in the actuarial mathematics, *Boletim do Instituto dos Actuários Portugueses,* 1954, pp. 7-15.
30. *de Castro Pereira S., Solteiro Pires E.J., de Moura Oliveira P.B.* Ant-Balanced Multiple Traveling Salesmen: ACO-BMTSP, *Algorithms*, 2023, pp. 37.

Статью рекомендовал к опубликованию д.т.н., профессор Н.Е. Сергеев.

Хуссейн Фирас Айманович – АО «НКБ Робототехники и систем управления»; e-mail: firas94mecha@gmail.com; г. Таганрог, Россия; тел.: 89996379357; м.н.с.

Houssein Fieas Aimanovich – Joint-Stock Company “Robotics and Control Systems” e-mail: firas94mecha@gmail.com; Taganrog, Russia; phone: +79996379357; junior researcher.

УДК 004.421.6 DOI 10.18522/2311-3103-2024-4-50-63

И.И. Левин, К.Н. Алексеев, А.А. Гуленок

ПРЕОБРАЗОВАНИЕ ПРОСТЕЙШИХ СОРТИРУЮЩИХ СЕТЕЙ К НЕЧЕТНО-ЧЕТНОЙ СЕТИ БАТЧЕРА

Все алгоритмы сортировки являются информационно-эквивалентными, поэтому выбор наиболее эффективного алгоритма обычно зависит от скорости его работы и от количества используемой памяти. При параллельной, аппаратной реализации на эффективность алгоритмов сортировки также влияют: степень утилизации аппаратных ресурсов; латентность результирующей вычислительной структуры; количество и разрядность сортируемых элементов. Задача сортировки или упорядочивания данных не формализована в виде математических преобразований, поэтому каждый из известных алгоритмов ее решения принято рассматривать как атомарную, независимую единицу. Переход от одного алгоритма решения задачи к другому возможен при описании задачи в виде информационного графа, вершины которого отражают элементарные выполняемые операции, а дуги – информационные зависимости между ними. Имея набор элементарных преобразований, можно влиять на функциональную регулярность связей информационного графа, латентность вычислительной структуры, коэффициент параллелизма и т.д. К преимуществам работы с информационным графом задачи также можно отнести сравнительную простоту используемого понятийного аппарата. Информационный граф задачи сортировки «пузырьком» представляет собой простейшую сортирующую сеть, построенную по принципу объединения ступеней «голова-хвост». В данной работе показана и обоснована функциональная избыточность подобных сортирующих сетей; приведены способы оптимизации числа операций и изменения порядка их следования. Основным результатом работы является методика преобразования сортирующих сетей в нечетно-четную сеть слияния Батчера. Разработана программа, автоматически выполняющая преобразование сортирующих сетей и позволяющая подстраивать топологию информационного графа под наиболее эффективный вид в зависимости от результирующей степени параллелизма вычислительной структуры. Обобщая полученные результаты, можно отметить, что автоматизированное приведение известных алгоритмов к «быстрым» может обеспечить получение оптимальной параллельно-конвейерной программы при заданных ограничениях, что позволит значительно ускорить процесс их разработки.

Алгоритмы сортировки; сортирующие сети; сеть слияния Батчера; информационный граф; ярусно-параллельная форма; преобразование алгоритмов.

I.I. Levin, K.N. Alekseev, A.A. Gulenok

TRANSFORMATION OF THE SIMPLEST SORTING NETWORKS TO AN ODD-EVEN BUTCHER NETWORK

All sorting algorithms are information-equivalent. Therefore, the choice of the most effective algorithm usually depends on its operation velocity and the capacity of used memory. At parallel, hardware implementation, the efficiency of sorting algorithms is also affected by the degree of utilization of hardware resources; the latency of the resulting computing structure; the number and digit capacity of the sorted elements. The problem of sorting or ordering data is not formalized in the form of mathematical transformations. Therefore, each of the known algorithms for solving it is considered an atomic, independent unit. The transition from one algorithm for solving the problem to another is possible at describing the problem in the form of an information graph, the vertices of which represent the elementary performed operations, and the arcs – the information dependencies between them. Having a set of elementary transformations, it is possible to influence the functional regularity of the information graph connections, the latency of the computing structure, the coefficient of parallelism, etc. The information graph of the "bubble" sorting problem is a simple sorting network, developed on the "head-tail" principle of combining steps. In this paper, the functional redundancy of such sorting networks is shown and justified; the methods to optimize the number of operations and change the order of their sequence are given. The main result of the paper is the method for converting sorting networks into an odd-even Butcher mergesort. A program has been developed that automatically performs the transformation of sorting networks and allows to adjust the information graph topology to the most effective form, depending on the resulting degree of parallelism of the computing structure. Summarizing the

obtained results, note that the automated reduction of known algorithms to "fast" ones can ensure the optimal parallel pipeline program under specified constraints, which will significantly accelerate the process of their development.

Sorting algorithms; sorting networks; Batcher mergesort; information graph; tiered-parallel form; algorithm transformation.

Введение. Операция сортировки является составной частью анализа, фильтрации и преобразования данных, поэтому широко используется в множестве алгоритмов решения прикладных задач. Известно большое число различных алгоритмов сортировки данных [1–3], начиная от академической «сортировки пузырьком» или «сортировки слиянием» и заканчивая наиболее производительными алгоритмами, такими как «team sort» [4] и «q-sort» [5]. Многообразие алгоритмов сортировки обусловлено различием в числе и типе выполняемых операций: непосредственное сравнение данных или сравнение с константными значениями; сравнение с краями диапазона или нахождение количества однотипных элементов и т.д. [1–3].

При выполнении сортировки на вычислительных системах традиционной архитектуры наилучшим является алгоритм с меньшим числом операций, оптимально использующий процессорное время и позволяющий минимизировать простои при передаче информации между системой хранения данных, оперативной памятью и кэшем процессора. Для ускорения сортировки больших массивов данных широко используются параллельные аппаратные реализации алгоритмов сортировки, например, конвейерная сортировка пузырьком, сортирующие сети Батчера, поразрядная сортировка, а также иные гибридные методы сортировки [6–9]. В этом случае выбор алгоритма сортировки сводится решению задачи многокритериальной оптимизации, на сложность которой влияют: степень распараллеливания, определяющая коэффициент редукции и латентность вычислительной структуры; заданное время обработки данных; доступный аппаратный ресурс; регулярность информационного графа задачи; требуемый объем памяти для хранения результатов промежуточных вычислений; тип, количество и разрядность сортируемых элементов [6].

Следует отметить, что все существующие алгоритмы сортировки информационно-эквивалентны, поскольку все они выполняют функцию упорядочивания элементов множества. В тоже время между различными алгоритмами сортировки существует семантический разрыв вследствие того, что математический аппарат для описания различных алгоритмов сортировки не формализован и не существует методологических основ для осуществления перехода от одного алгоритма к другому. Каждый из алгоритмов сортировки рассматривается как отдельный и независимый подход к решению задачи упорядочивания данных, из-за чего невозможно автоматизировать процесс получения более эффективных алгоритмов из исходного для определенного аппаратного ресурса.

Иным подходом к синтезу параллельных алгоритмов является анализ и преобразование информационных графов задачи [10–11]. Например, в работах [12–18] показаны способы преобразования информационных графов с ассоциативными и дистрибутивными алгебраическими операциями, за счет чего были получены графы, соответствующие другим известным алгоритмам решаемой задачи. Преимуществом работы с информационными графами по сравнению с математическими формулами является сравнительная простота используемого понятийного аппарата. Исследования показывают, что задача преобразования информационного графа обладает меньшей алгоритмической сложностью по сравнению последовательным выполнением математических преобразований, что может обеспечить возможность ее дальнейшей автоматизации.

В рамках данной статьи описаны правила преобразования информационных графов задачи сортировки, представляющих собой сортирующие сети, на базе которых создана методика оптимизации сортирующих сетей. Работоспособность методики проиллюстрирована на примере преобразования информационного графа простейшей сортирующей сети, соответствующей последовательному алгоритму сортировки «пузырьком», к нечетно-четной сортирующей сети Батчера (Batcher odd-even merge sort) [19], содержащей наименьшее число сортирующих элементов.

Сортирующие сети. Известно большое число разнообразных сортирующих сетей, отличающихся количеством сортирующих элементов SE, латентностью вычислительной структуры, регулярностью связей меду сортирующими элементами [1–3, 19]. Из-за этого, разные сортирующие сети являются оптимальными при разной степени параллелизма [12]. Характерной чертой простейших сортирующих сетей (рис. 1,а,б), является объединение сортирующих элементов SE в «ступени», выполняющие следующие функции (на рис. 1,а,б выделены пунктиром):

1) нахождение максимума: $X = X_n > \{X_1, X_2,\ X_3,\ \dots,\ X_{n-1}\}$, $X_n = max(A)$;
2) вставка элемента в отсортированное множество;
3) вставка элемента в частично упорядоченное множество;
4) перенос нескольких больших элементов ближе к старшим выходам графа.

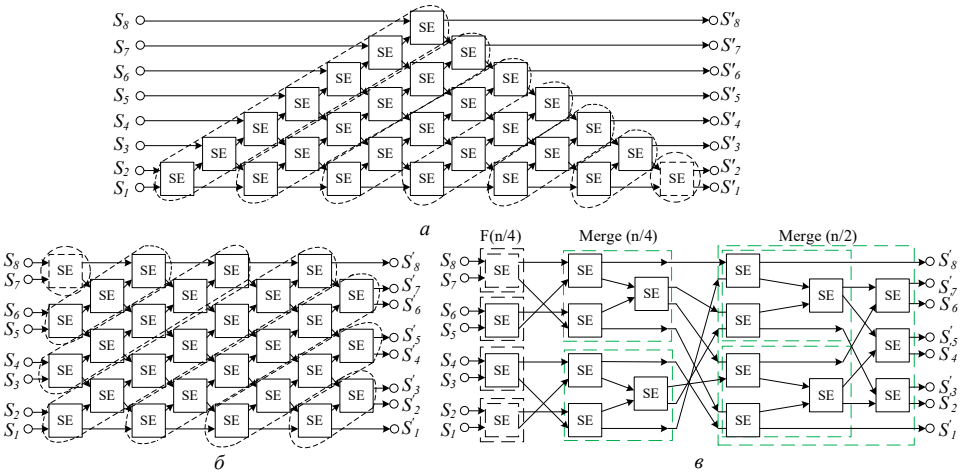


Рис. 1. Схемы простейших сортирующих сетей при n = 8, последовательная реализация которых соответствует последовательным алгоритмам: а – сортировки «пузырьком»; б – нечетно-четной сортировке; в – нечетно-четной сортирующей сети Батчера

Очевидно, чем меньше сортирующих элементов в схеме, тем большую степень масштабирования можно обеспечить в ограниченном аппаратном ресурсе. Кен Батчер создал две универсальные архитектуры сортирующих сетей с функционально-регулярными информационными графами и вычислительной сложностью около $O(n \cdot log(n))$, где n – мощность входного массива данных. Сеть, состоящая из наименьшего числа сортирующих элементов, называется «нечетно-четная сеть слияния Батчера» (рисунок 1-в). Основная идея при построении данной сети заключается в последовательном слиянии двух упорядоченных подмножеств по принципу: слияние нечетных элементов входных упорядоченных последовательностей; слияние четных элементов входных упорядоченных последовательностей; дальнейшее слияние промежуточных данных в одну общую упорядоченную последовательность с помощью $n/2 - 1$ операций.

Избыточность простейших сетей сортировки. Как отмечалось в работе [12], ступень сортирующей сети реализует сразу несколько различных функций, одной из которых является перенос нескольких больших элементов ближе к старшим выходам графа. Так как основным принципом работы ступени в простейших сортирующих сетях является именно нахождение максимума в неотсортированном множестве элементов, перемещение не максимальных элементов очевидно не несет никакого смысла. Также очевидно, что для сортировки элементов массива достаточно единожды переместить каждый элемент на новое место, для чего необходимо выполнить всего n операций перемещения элементов, что реализуется в последовательном алгоритме сортировки «выбором». Исходя из вышесказанного, можно сделать вывод, что простейшие сортирующие сети, количество сортирующих элементов в которых равно $n^2/2$ обладают избыточностью выполняемых операций сравнения и перемещения элементов.

Избавиться от подобной избыточности можно путем использования принципа «деления пополам», который соответствует принципу описания последовательных алгоритмов «разделяй и властвуй». Согласно данным принципам как раз и построена нечетно-четная сортирующая сеть Батчера, а также и некоторые «быстрые» алгоритмы иных задач: бинарный поиск, умножение Карацубы, быстрое дискретное преобразование Фурье и т.п. [20].

Для устранения избыточности выполняемых операций в первую очередь необходимо построить сортирующую сеть по принципу «деления пополам» (рис. 3), используя при этом принципы объединения ступеней, подробно рассмотренные в работе [12]. Заметим, что сортирующую сеть, соответствующую последовательному алгоритму «шейкерной» сортировки (рис. 2,а), можно логически разбить на подграфы: выделить два подграфа, которые представляют собой простейшие сортирующие сети меньшего размера *F(n/2)*, а также подграф сети слияния двух отсортированных последовательностей *Merge.* После аналогичного преобразования каждой из выделенных сортирующих сетей меньшего размера, была получена сортирующая сеть, состоящая из сетей слияния разного размера (рис. 2,б). Так как после перестановки ступеней в сортирующей сети число выполняемых операций не изменилось, можно сделать вывод о существующей избыточности выделенных сетей слияния.

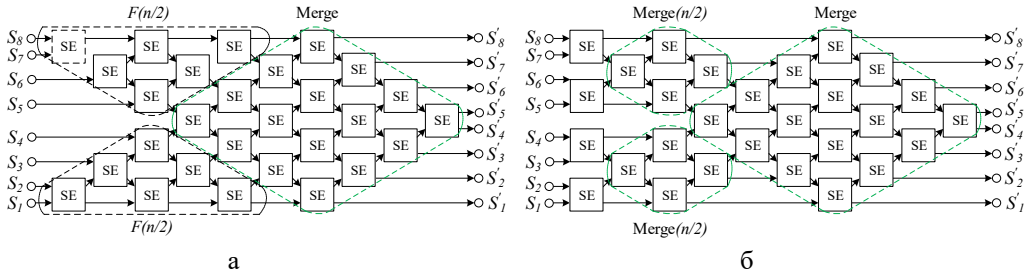


Рис. 2. Исходная (а) и преобразованная (б) схемы сортирующей сети, соответствующей последовательному алгоритму «шейкерной» сортировки

Было обнаружено, что сеть слияния можно функционально разделить на две составные части: сеть перестановки элементов, которая на рис. 3,а выделена пунктиром, и оставшаяся часть сортирующих элементов, отвечающая за непосредственное слияние последовательностей.

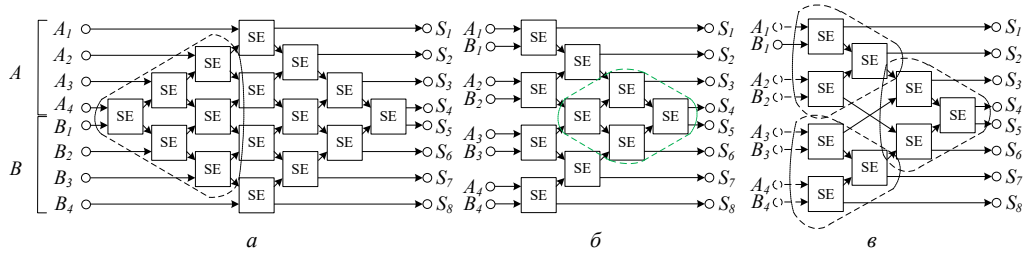


Рис. 3. а – сеть слияния, полученная из простейшей сортирующей сети при n = 8; б – сеть слияния после первого шага оптимизации; в – оптимизированная сеть слияния

Оптимизируем сеть слияния путем удаления функционально-избыточных операций. На первом шаге оптимизации удалим часть операций, отвечающую за перестановку данных (рис. 3,б), и изменим порядок поступления массивов A и B на функцию слияния. Заметим, что после этого часть сети слияния, выделенная пунктиром на рис. 3,б, представляет собой сеть слияния меньшего размера, которую можно оптимизировать тем же способом. На рис. 3,в приведена сеть слияния пирамидального вида, полученная после выполнения всех шагов оптимизации. Части сети слияния, выделенные пунктиром, являются простейшими сетями слияния Батчера.

На рис. 4 представлена сортирующая сеть для n = 8, полученная после оптимизации избыточных операций в сетях слияния. Хотя данная сеть отличается от сети Батчера, она имеет такое же число сортирующих элементов.

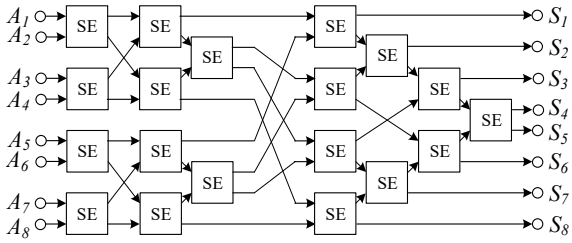


Рис. 4. Простейшая сортирующая сеть при n = 8 после оптимизации избыточных операций, отвечающих за перемещение элементов

На рис. 5 приведено сравнение структур сетей слияния, полученных после оптимизации простейшей сортирующей сети и из нечетно-четной сети Батчера, где видно, что при n = 2 и n = 4 графы сетей слияния изоморфны, а при n = 8 – состоят из одинакового числа операций. Число сортирующих элементов в сети слияния из нечетно-четной сортирующей сети Батчера определяется отношением: $N_{SE}(n) = n \cdot (\log_2 n - 1)/2 + 1$; а в пирамидальной сети слияния: $N_{SE}(n) = 3^{\log_2 n}$, в связи с чем можно говорить об имеющейся избыточности сетей слияния больших размерностей.

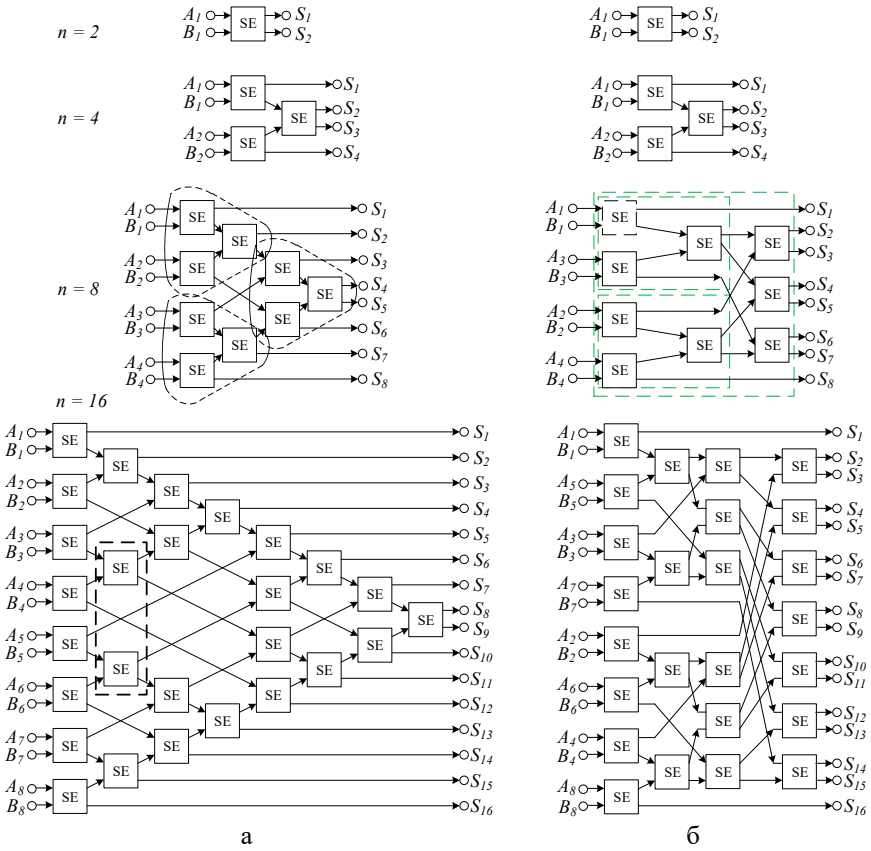


Рис. 5. Структуры сетей слияния при n=2, 4, 8 и 16, полученных:
а – из простейшей сети сортировки; б – из нечетно-четной сети Батчера

Покажем, что при $n = 16$ два сортирующих элемента, выделенные пунктиром на рис. 5,а пунктиром, являются функционально-избыточными. Очевидно, что $max(A_1, B_1)$ будет первым, максимальным элементом результирующего упорядоченного множества, тогда как $max(A_2, B_2)$ может занять одну из позиций в результирующем множестве в диапазоне от 2 до $n/2 + 1$; следующий максимальный выход $max(A_3, B_3)$ может занять позицию в диапазоне от 3 до $n/2 + 3$ и т.д. При этом, $min(A_1, B_1)$ может занять одну из позиций в результирующем множестве в диапазоне от 2 до $n/2 + 1$; $min(A_2, B_2)$ может занять позицию в диапазоне от 3 до $n/2 + 2$; следующий минимальный выход $min(A_3, B_3)$ может занять позицию в диапазоне от 5 до $n/2 + 3$ и т.д. Таким образом, после первого слоя SE сортируемые данные разбиваются на некоторое количество пересекающихся диапазонов результирующего множества. Верхний элемент SE, выделенный на рис. 5,а пунктиром, принимает для сравнения $min(A_3, B_3)$ и $max(A_4, B_4)$, а выход данного элемента участвует в формировании результирующих элементов $S_4 - S_{10}$. При этом известно, что $min(A_3, B_3)$ никогда не сможет занять позицию S_4, в связи с чем данное сравнение может быть оптимизировано.

Анализируя диапазоны позиций элементов в результирующем множестве после первого слоя SE, можно определить функционально-избыточные элементы в сетях слияния любой мощности. Так на рис. 6 приведена сеть слияния при $n = 32$, в которой черным цветом выделены сортирующие элементы, которые могут быть удалены.

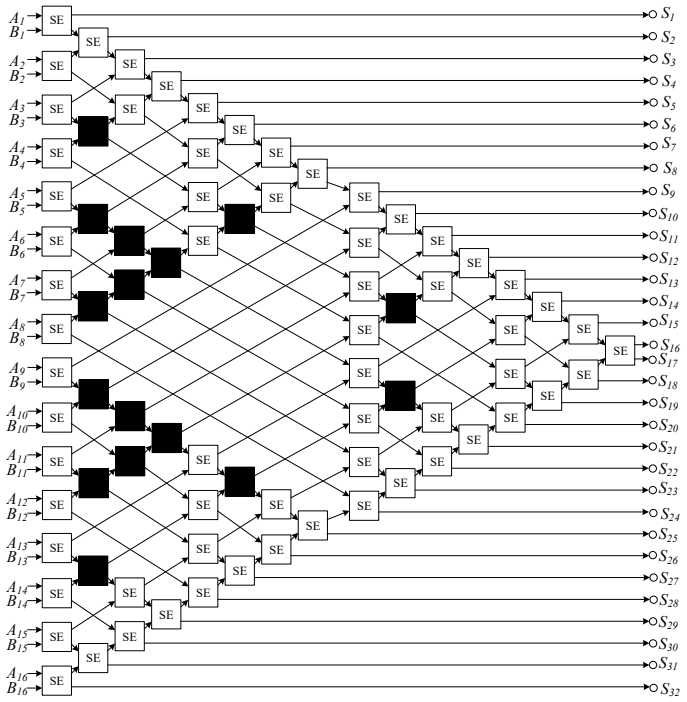


Рис. 6. Сеть слияния при $n = 32$ *с выделенными сортирующими элементами для дальнейшей оптимизации*

После удаления избыточных сортирующих элементов, их число будет соответствовать числу операций в нечетно-четной сети слияния Батчера. Однако, эксперименты показали, что сеть слияния после такого рода оптимизации перестает работать на всем множестве входных данных, из чего можно сделать вывод, что структура пирамидальной сети слияния не может быть оптимизирована без изменения порядка выполнения операций.

Изменение порядка ассоциативных операций. В сети слияния можно выделить подграфы, которые реализуют функцию вставки одного элемента в отсортированную последовательность (выделены пунктиром на рис. 7,а и рассмотрены отдельно на рис. 7,б,в).

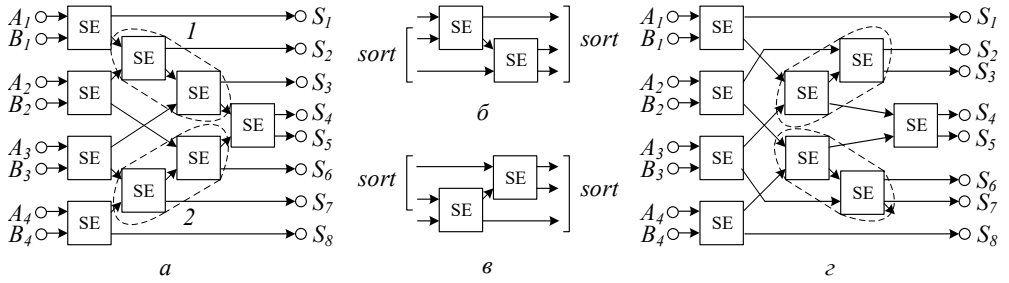


Рис. 7. а, б, в – выделение частей в сети слияния, реализующих операцию вставки элемента в отсортированную последовательность; г – сеть слияния после изменения порядка следования операций

Для подграфа №1 в качестве отсортированной части подаются элементы $max(A_2, B_2) > max(A_3, B_3)$, в качестве элемента для вставки – $min(A_1, B_1)$, а в результате будет получено упорядоченное множество элементов, содержащее два глобальных максимума. Аналогичная ситуация наблюдается при рассмотрении подграфа №2, с той разницей, что в результате будут получены два глобальных минимума. Так как результатом работы обеих рассмотренных структур будут являться упорядоченные множества элементов, можно утверждать, что они функционально-эквивалентны.

Наличие функциональной эквивалентности позволяет заменять одну структуру другой, фактически меняя структуру ступени: со структуры нахождения максимума на структуру нахождения минимума или наоборот. Так, изменив порядок следования операций в подграфах №1 и №2 на обратный, получим сеть слияния, представленную на рис. 7,г. Отметим, что граф полученной сети слияния имеет тот же порядок выполнения операций и идентичные связи между ними, что и сеть слияния нечетно-четной сортирующей сети Батчера, в связи с чем эти графы являются изоморфными.

Данное свойство можно обобщить для любой размерности ступеней сортирующей сети. Графы, приведенные на рис. 8,а,б, являются функционально-эквивалентными, поскольку выполняют одну и ту же функцию: вставку элемента в упорядоченное множество. Если изменить порядок следования операций, можно построить каскадную, пирамидальную схему, которая представлена на рис. 8,в. В данной схеме после сравнения A_2 и B, максимум перейдет для сравнения с A_1, а минимум – с A_3. Довольно очевидно, что данный граф также является функционально-эквивалентным графам на рис. 8,а,б. Рассмотренный принцип работает и для графов ступеней большей размерности (рис. 8,г,д,е,ж). При этом, если отдельно рассмотреть ступень, состоящую из трех сортирующих элементов и выделенную на рис. 8,д пунктиром, она сама по себе выполняет операцию вставки элемента в упорядоченное множество. Граф вставки элемента в отсортированный массив, приведенный на рис. 8,ж, реализует каскадную схему ставки при $n = 8$. Заметим, что данная схема построена по принципу бинарного поиска: первым выполняется сравнение со средним элементом упорядоченного подмножества, затем с серединами оставшихся частей и т.д.

Изменение порядка следования операций с последовательного «голова-хвост» на каскадный «деление пополам» при сохранении функциональной эквивалентности позволяет значительно уменьшить как число слоев в графе, так и латентность полученной на его основе вычислительной структуры.

Переход к нечетно-четной сети слияния Батчера. Рассмотрим простейшую сортирующую сеть на $n = 16$ элементов, соответствующую последовательному алгоритму сортировки «пузырьком». После перестроения сети согласно принципу «деления пополам» и оптимизации функционально-незначимых элементов получен граф, представленный на рис. 9,а. Для приведения данной сети к сети Батчера необходимо оптимизировать все выделенные пунктиром сети слияния.

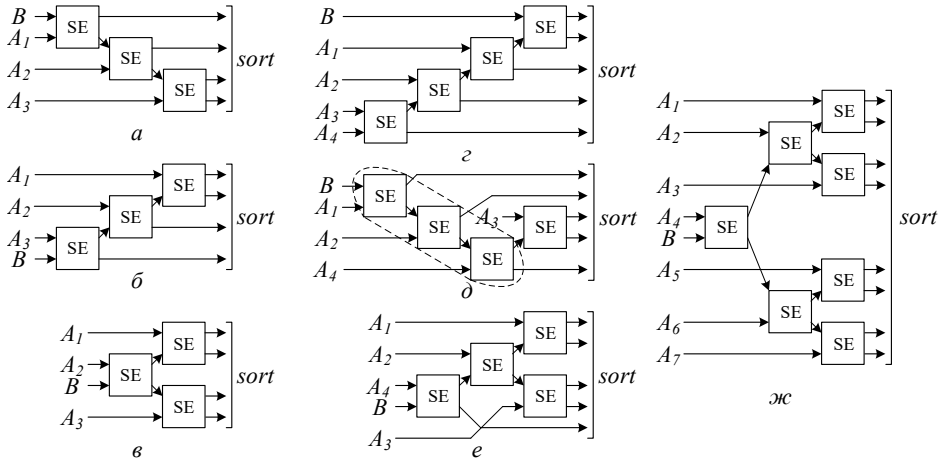


Рис. 8. Варианты графов вставки элемента в отсортированный массив при: а-в – $n = 4$; г-е – $n = 5$; ж – $n = 8$

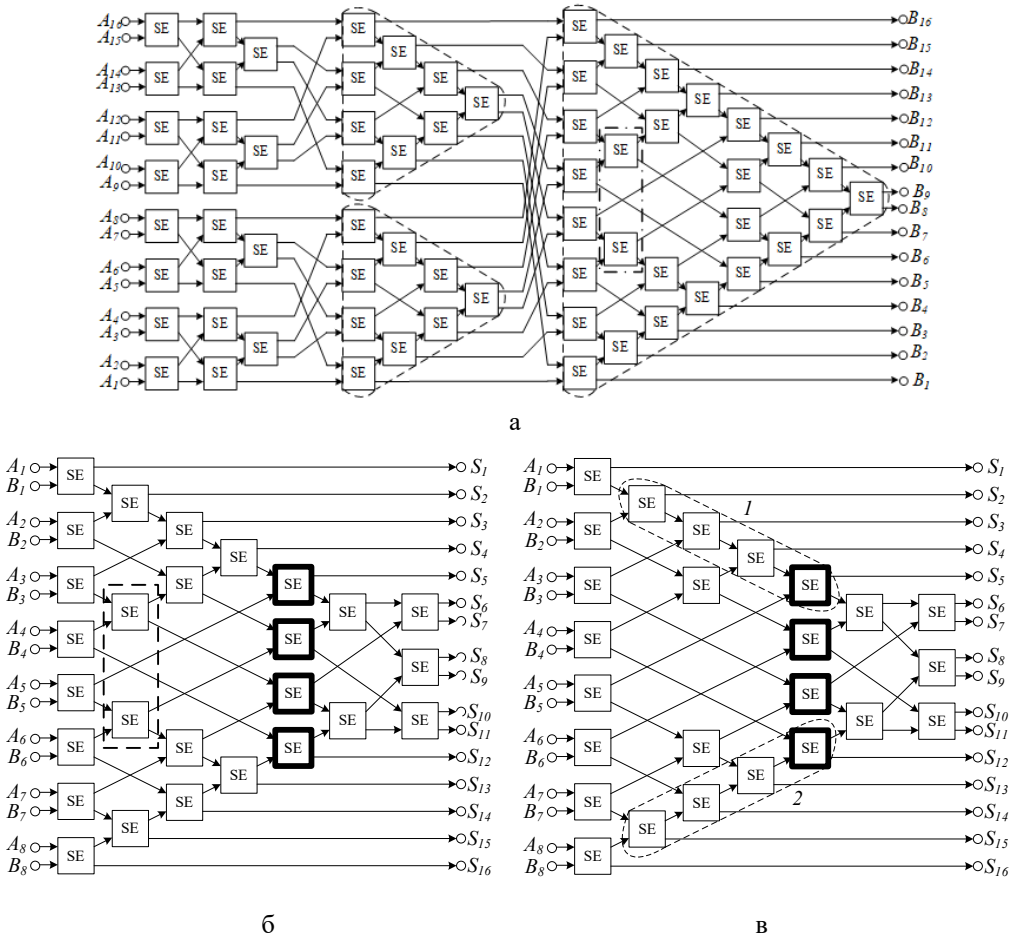


Рис. 9. а – сортирующая сеть при $n = 16$ после перестроения согласно принципу «деления пополам» и оптимизации функционально-незначимых элементов; б – структура сети слияния после преобразования подсети слияния средних элементов; в – после оптимизации числа сравнений

Преобразования стоит начинать с последнего, самого большого подграфа, описывающего сеть слияния двух последовательностей в отсортированное множество. Заметим, что конец данной сети тоже является сетью слияния меньшего размера и выполняет слияние центральных элементов последовательности. Выполним приведение данного подграфа к сети слияния Батчера путем изменения последовательности операций, как это было показано на рисунке 8, после чего оптимизируем функционально-избыточные сортирующие элементы, выделенные пунктиром на рис. 9,б. В данном случае после их удаления, сеть, структура которой приведена на рис, 9,в, продолжает выполнять функцию слияния. Таким образом, все наибольшие сети слияния, приведенные на рис. 9, являются функционально-эквивалентными.

Однако, для сетей слияния большей размерности это не является правилом, и сеть перестает выполнять заданную функцию, поэтому для сохранения функциональной эквивалентности требуется выполнить дальнейшие преобразования, а именно, изменить порядок ассоциативных операций таким образом, чтоб сортирующие элементы, выделенные жирной линией на рис. 9 переместились на место второго слоя сети слияния.

Изменим порядок ассоциативных операций, выделенных пунктиром на рис. 9,в. Для части сети слияния под номером 1, в качестве отсортированной части подаются элементы $max(A_2, B_2) > max(A_3, B_3) > max(A_5, B_5)$ и элемент $max(min(A_3, B_3), max(A_4, B_4), min(A_2, B_2))$. В качестве элемента для вставки поступает элемент $min(A_1, B_1)$. Последовательность преобразований для изменения порядка выполнения операций приведена на рис. 10. Аналогичные преобразования выполним для части сети слияния, представленной на рис. 9,б под номером 2. После выполнения всех преобразований, структура оптимизированной сети слияния, полученной из квадратичной сортирующей сети, станет изоморфна структуре сети слияния, выделенной из нечетно-четной сортирующей сети Батчера.

Для завершения преобразования сортирующей сети, представленной на рис. 9,в сеть Батчера необходимо последовательно изменить порядок ассоциативных операций во всех сетях слияния, выделенных пунктиром, после чего данная сеть станет изоморфна нечетно-четной сети слияния Батчера.

Анализ показал, что выполнение представленной последовательности действий позволяет привести любую простейшую сортирующую сеть к сети Батчера при любой мощности сортируемого массива данных.

Была разработана методика преобразования сортирующих сетей в нечетно-четную сеть слияния Батчера, согласно которой необходимо:

1. С помощью правил перестановки ступеней [12], привести сортирующую сеть к виду, из которого можно однозначно выделить сортирующие сети меньшей мощности и сеть слияния.
2. Повторить пункт 1 для всех полученных сортирующих сетей меньшей мощности. Рекурсивно исполнять данный пункт, пока не выполнится условие: все сортирующие сети выполняют сортировку 2 элементов исходной последовательности. Данные преобразования соответствуют построению сортирующей сети согласно принципу «деление пополам».
3. Выделить самую крупную сеть слияния.
4. Оптимизировать все выделенные сети слияния путем удаления избыточных операций, отвечающих за перемещение элементов.
5. Если размерность выделенной сети слияния больше четырех, выделить все сети слияния вдвое меньшей мощности и перейти к пункту 4. Иначе, перейти к пункту 6.
6. В полученной сети выделить и удалить функционально-избыточные сортирующие элементы, которые не участвуют в формировании результата.
7. Выделить последнюю и предпоследнюю сети слияния, не являющуюся сетью Батчера.

8. Выполнить изменение порядка следования операций начиная с краев сети слияния таким образом, чтоб первый слой SE сети слияния центральных элементов множества стал вторым слоем сети слияния центральных элементов большего размера.

9. Выполнять пункты методики 7–8 до тех пор, пока граф сортирующей сети не станет изоморфным графу сети нечетно-четного слияния Батчера.

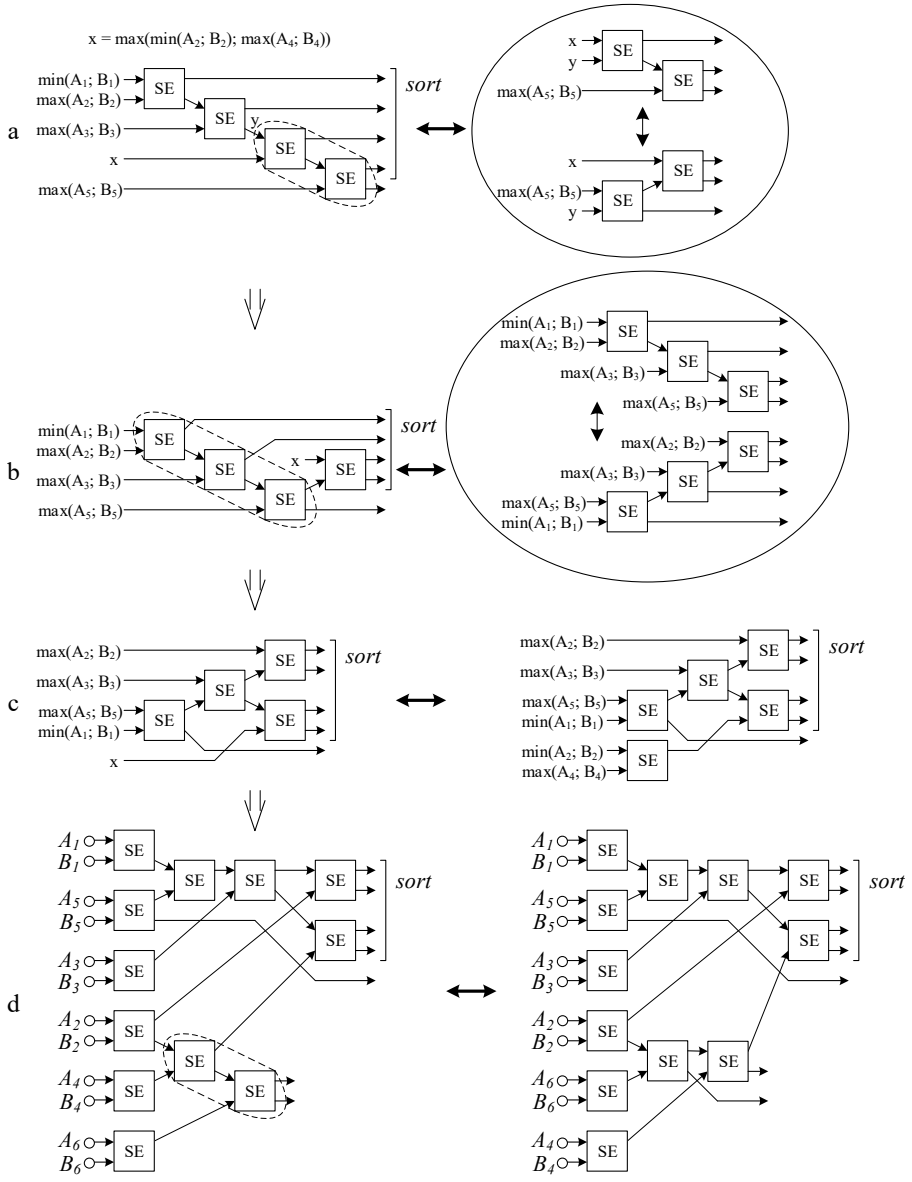


Рис. 10. Последовательность преобразований частей сети слияния, которая реализует вставку элемента в отсортированную последовательность

Апробация. Представленная методика преобразования сортирующих сетей в нечетно-четную сеть слияния Батчера была реализована в виде программы, написанной на языке высокого уровня C#. На вход программы поступает информационный граф сортирующей сети, соответствующей последовательному алгоритму сортировки «пузырьком». На выходе программа формирует информационный граф, изоморфный сети нечетно-четного слияния Батчера. Результат работы программы для $n = 32$ приведен на рис. 11.

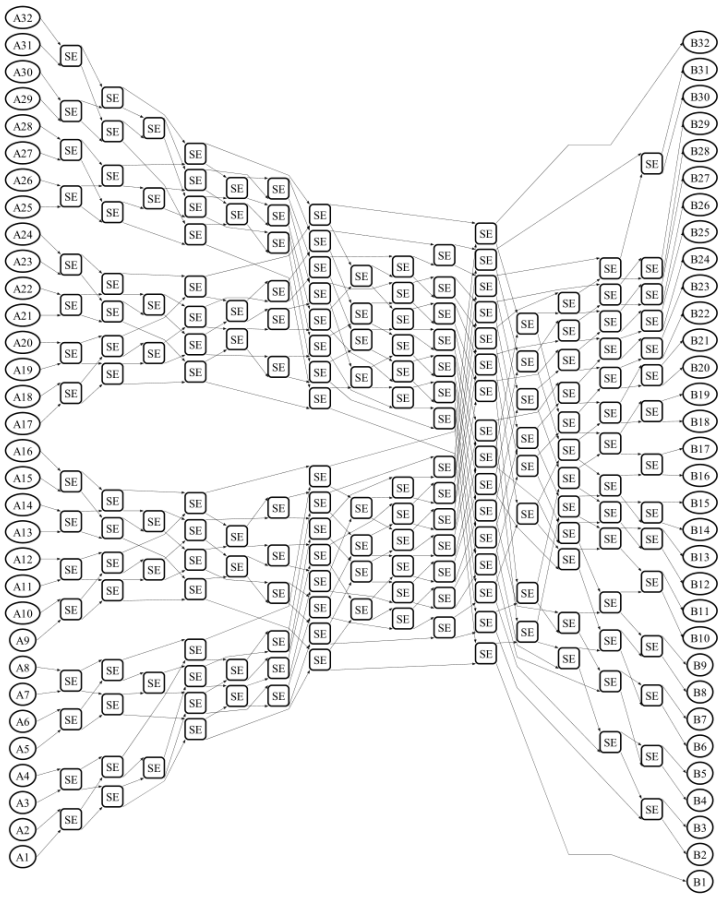


Рис. 11. Результат работы программы приведения информационных графов сортирующих сетей к сети нечетно-четного слияния Батчера

Разработанная программа была протестирована на сортирующих сетях большей размерности: для $n = 64, 128, 256$. Стоит отдельно отметить, что все операции выполняются над графом автоматически, без участия человека.

Заключение. В рамках данной работы рассмотрены способы оптимизации функционально-избыточных операций в сортирующих сетях, а также способы изменения порядка следования ассоциативных операций. Основным результатом является методика преобразования сортирующих сетей в нечетно-четную сеть слияния Батчера. Реализация разработанной методики в виде программного продукта позволяет подстраивать топологию информационного графа под наиболее эффективный вид в зависимости от результирующей степени параллелизма вычислительной структуры. Автоматизация процесса преобразования сортирующих сетей позволяет увеличить эффективность используемого оборудования без особых временных затрат на разработку и отладку кода параллельной программы при изменении темпа поступления и обработки данных.

Стоит отметить, что оптимизация функционально-избыточных элементов стала возможна только после построения сортирующей сети по принципу «деления пополам». Обобщая данный подход, можно утверждать, что алгоритмы, построенные согласно принципу «деления пополам» способны значительно уменьшить общее число выполняемых операций, что наблюдается при переходе от сортирующей сети, соответствующей последовательному алгоритму сортировки «пузырьком», к нечетно-четной сети слияния Батчера.

Вместе с этим, существуют иные подходы, иначе реализующие парадигму «разделяй и властвуй»: предобработка входных данных с последующим выполнением основных функций. По таким принципам построена битоническая сортировка Батчера, «q-sort», «ведерная» сортировка и поразрядная сортировка. Теоретически, автоматизация перехода от сортирующих сетей к данным алгоритмам сортировки позволит выбирать оптимальный алгоритм сортировки для имеющегося аппаратного ресурса и заданного времени решения задачи без участия разработчика, что позволит значительно ускорить процесс разработки параллельно-конвейерных программ.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Кнут Д.Э.* Искусство программирования. Т. 3. Сортировка и поиск. – 2-е изд.: пер. с англ. – М.: Изд. дом «Вильямс», 2001.
2. *Скиена С.С.* Алгоритмы. Руководство по разработке. – 3-е изд.: пер. с англ. – СПб.: БХВ-Петербург, 2022. – 848 с.
3. Sorting algorithm. – Режим доступа: https://en.wikipedia.org/wiki/Sorting_algorithm (дата обращения: 31.07.2024).
4. Timsort. – Режим доступа: https://en.wikipedia.org/wiki/Timsort (дата обращения: 31.07.2024).
5. Quicksort. – Режим доступа: https://en.wikipedia.org/wiki/Quicksort (дата обращения: 31.07.2024).
6. *Алексеев К.Н.* Методы и средства создания параллельно-конвейерных программ для решения задач реального времени на реконфигурируемых вычислительных системах: дисс. … канд. техн. наук, по специальности 05.13.11 “Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей” / научный руководитель: д.т.н., проф. Левин И.И. – Таганрог, 2020. – 213 с.
7. *Алексеев К.Н., Левин И.И., Сорокин Д.А.* Структурно-процедурная реализация на реконфигурируемых вычислительных системах кодирования Хаффмана в реальном масштабе времени // Вестник компьютерных и информационных технологий. – 2018. – № 9. – С. 3-10. – ISSN 1810-7206. – DOI: 10.14489/vkit.2018.09.pp.003-010. – http://www.vkit.ru/index.php//current-issue-rus/751-003-010.
8. *Алексеев К.Н., Сорокин Д.А., Семерникова Е.Е.* Структурно-процедурная реализация кодирования алгоритмом Хаффмана в задачах управления // Матер. 10-й Всероссийской мультиконференции по проблемам управления (МКПУ-2017), 11-16 сентября 2017 г. – Ростов-на-Дону: Изд-во ЮФУ, 2017. – Т. 3. – С. 126-127.
9. *Dudnikov E.A., Alekseev K.N., Sorokin D.A.* High-Speed Huffman encoding of dense data flows on FPGA // German International Journal of Modern Science / Deutsche internationale Zeitschrift für zeitgenössische Wissenschaft. – 2021. – No. 9, Vol. 1. – P. 36-40. – https://dizzw.com/wp-content/uploads/2021/05/Deutsche-internationale-Zeitschrift-für-zeitgenössische-Wissenschaft-№9-part-1-2021-37-41.pdf.
10. *Каляев А.В., Левин И.И.* Модульно-наращиваемые многопроцессорные системы со структурно-процедурной организацией вычислений. – М.: Янус-К, 2003. – 380 с.
11. *Каляев И.А., Левин И.И., Семерников Е.А., Шмойлов В.И.* Реконфигурируемые мультиконвейерные вычислительные структуры. – 2-е изд., перераб. и доп. / под общ. ред. И.А. Каляева. – Ростов-на-Дону: Изд-во ЮНЦ РАН, 2009. – 344 с. – ISBN 978-5-902982-61-6.
12. *Левин И.И., Алексеев К.Н.* Преобразование сортирующих сетей для разной степени параллелизма // Известия ЮФУ. Технические науки. – 2023. – № 5 (235). – С. 104-118. – DOI: 10.18522/2311-3103-2023-5-104-118.
13. *Писаренко И.В., Алексеев К.Н., Мельников А.К.* Ресурсонезависимое представление сортирующих сетей на языке программирования Set@l // Вестник компьютерных и информационных технологий. – 2019. – № 11 (185). – С. 53-60. – DOI: 10.14489/vkit.2019.11.pp.053-060. http://vkit.ru/index. php/ current-issue-rus/857-053-060.
14. *Писаренко И.В., Касаркин А.В., Алексеев К.Н.* Рекурсивное описание графов с ассоциативными операциями на языке программирования Set@l // Всероссийская научная конференция "Суперкомпьютерные технологии (СКТ) 2020", 05-10 октября 2020 г. Крым, Алушта, Россия. – Ростов-на-Дону; Таганрог: Изд-во ЮФУ, 2020. – С. 79-83.
15. *Михайлов Д.В.* Методы создания параллельно-конвейерных программ для задач с последовательным информационным графом: дисс. ... канд. техн. наук по специальности 05.13.11 “Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей” / научный руководитель: д.т.н., проф. Левин И.И. – Таганрог, 2022. – 213 с.
16. *Михайлов Д.В., Левин И.И., Дордопуло А.И., Писаренко И.В.* Представление графов с ассоциативными операциями на языке программирования SET@L // Известия ЮФУ. Технические науки. – 2020. – № 3 (213). – С. 98-109. – DOI: 10.18522/2311-3103-2020-3-98-111.

17. *Михайлов Д.В.* Преобразование некоторых видов последовательных информационных графов в параллельно-конвейерную форму // Известия ЮФУ. Технические науки. – 2020. – №7 (217). – С. 78-93. – DOI: 10.18522/2311-3103-2020-7-78-93.
18. *Михайлов Д.В.* Преобразование последовательного информационного графа метода прогонки в параллельную форму // Известия ЮФУ. Технические науки. – 2021. – № 7 (224). – С. 177-188. – DOI: 10.18522/2311-3103-2021-7-177-188.
19. Batcher odd–even mergesort. – Режим доступа: https://en.wikipedia.org/wiki/Batcher_odd%E2%80%93even_mergesort (дата обращения: 31.07.2024)
20. *Richard E.* Blahut Fast Algorithms for Digital Signal Processing. Cambridge University Press, 05.2011. – DOI: https://doi.org/10.1017/CBO9780511760921.

REFERENCES

1. *Knut D.E.* Iskusstvo programmirovaniya. T. 3. Sortirovka i poisk [The Art of Computer Programming. Vol. 3. Sorting and Search]. 2nd ed.: transl. from engl. Moscow: Izd. dom «Vil'yams», 2001.
2. *Skiena S.S.* Algoritmy. Rukovodstvo po razrabotke [Algorithms. Development Guide]. 3rd ed.: transl. from engl. Saint Petersburg: BKhV-Peterburg, 2022, 848 p.
3. Sorting algorithm. Available at: https://en.wikipedia.org/wiki/Sorting_algorithm (accessed 31 July 2024).
4. Timsort. Available at: https://en.wikipedia.org/wiki/Timsort (accessed 31 July 2024).
5. Quicksort. Available at: https://en.wikipedia.org/wiki/Quicksort (accessed 31 July 2024).
6. *Alekseev K.N.* Metody i sredstva sozdaniya parallel'no-konveyernykh programm dlya resheniya zadach real'nogo vremeni na rekonfiguriruemykh vychislitel'nykh sistemakh: diss. ... kand. tekhn. nauk, po spetsial'nosti 05.13.11 “Matematicheskoe i programmnoe obespechenie vychislitel'nykh mashin, kompleksov i komp'yuternykh setey” [Methods and tools for creating parallel-pipeline programs for solving real-time problems on reconfigurable computing systems: cand. of eng. sc. diss., in specialty 05.13.11 “Mathematical and software support for computing machines, complexes and computer networks”], scientific supervisor: dr. of eng. sc., professor Levin I.I. Taganrog, 2020, 213 p.
7. *Alekseev K.N., Levin I.I., Sorokin D.A.* Strukturno-protsedurnaya realizatsiya na rekonfiguriruemykh vychislitel'nykh sistemakh kodirovaniya Khaffmana v real'nom masshtabe vremeni [Structural-procedural implementation on reconfigurable computing systems of Huffman coding in real time], *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Bulletin of computer and information technologies], 2018, No. 9, pp. 3-10. ISSN 1810-7206. DOI: 10.14489/vkit.2018.09.pp.003-010. Available at: http://www.vkit.ru/index.php//current-issue-rus/751-003-010.
8. *Alekseev K.N., Sorokin D.A., Semernikova E.E.* Strukturno-protsedurnaya realizatsiya kodirovaniya algoritmom KHaffmana v zadachakh upravleniya [Structural and procedural implementation of coding by the Huffman algorithm in control problems], *Mater. 10-y Vserossiyskoy mul'tikonferentsii po problemam upravleniya (MKPU-2017), 11-16 sentyabrya 2017 g.* [Proceedings of the 10th All-Russian Multiconference on Control Problems (MCCP-2017), September 11-16, 2017]. Rostov-on-Don: Izd-vo YuFU, 2017, Vol. 3, pp. 126-127.
9. *Dudnikov E.A., Alekseev K.N., Sorokin D.A.* High-Speed Huffman encoding of dense data flows on FPGA, *German International Journal of Modern Science / Deutsche internationale Zeitschrift für zeitgenössische Wissenschaft*, 2021, No. 9, Vol. 1, pp. 36-40. Available at: https://dizzw.com/wp-content/uploads/2021/05/Deutsche-internationale-Zeitschrift-für-zeitgenössische-Wissenschaft-№9-part-1-2021-37-41.pdf.
10. *Kalyaev A.V., Levin I.I.* Modul'no-narashchivaemye mnogoprotsessornye sistemy so strukturno-protsedurnoy organizatsiey vychisleniy [Modularly scalable multiprocessor systems with structural-procedural organization of computations]. Moscow: Yanus-K, 2003, 380 p.
11. *Kalyaev I.A., Levin I.I., Semernikov E.A., Shmoylov V.I.* Rekonfiguriruemye mul'tikonveyernye vychislitel'nye struktury [Reconfigurable multipipeline computing structures]. 2nd ed., under the general ed. of I.A. Kalyaeva. Rostov-on-Don: Izd-vo YuNTS RAN, 2009, 344 p. ISBN 978-5-902982-61-6.
12. *Levin I.I., Alekseev K.N.* Preobrazovanie sortiruyushchikh setey dlya raznoy stepeni parallelizma [Transformation of sorting networks for different degrees of parallelism], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2023, No.№ 5 (235), pp. 104-118. DOI: 10.18522/2311-3103-2023-5-104-118.
13. *Pisarenko I.V., Alekseev K.N., Mel'nikov A.K.* Resursonezavisimoe predstavlenie sortiruyushchikh setey na yazyke programmirovaniya Set@l [Resource-independent representation of sorting networks in the Set@l programming language], *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Bulletin of Computer and Information Technologies], 2019, No. 11 (185), pp. 53-60. DOI: 10.14489/vkit.2019.11. pp.053-060. Available at: http://vkit.ru/index. php/ current-issue-rus/857-053-060.

14. *Pisarenko I.V., Kasarkin A.V., Alekseev K.N.* Rekursivnoe opisanie grafov s assotsiativnymi operatsiyami na yazyke programmirovaniya Set@l [Recursive Description of Graphs with Associative Operations in the Set@l Programming Language], *Vserossiyskaya nauchnaya konferentsiya "Superkomp'yuternye tekhnologii (SKT) 2020", 05-10 oktyabrya 2020 g. Krym, Alushta, Rossiya* [All-Russian Scientific Conference "Supercomputer Technologies (SCT) 2020", October 5-10, 2020. Crimea, Alushta, Russia]. Rostov-on-Don; Taganrog: Izd-vo YuFU, 2020, pp. 79-83.
15. *Mikhaylov D.V.* Metody sozdaniya parallel'no-konveyernykh programm dlya zadach s posledovatel'nym informatsionnym grafom: diss. ... kand. tekhn. nauk po spetsial'nosti 05.13.11 “Matematicheskoe i programmnoe obespechenie vychislitel'nykh mashin, kompleksov i komp'yuternykh setey” [Methods for creating parallel-pipeline programs for problems with a sequential information graph:... cand. of eng. sc. diss. in specialty 05.13.11 “Mathematical and software support for computing machines, complexes and computer networks”], scientific supervisor: dr. of eng. sc., professor Levin I.I. Taganrog, 2022, 213 p.
16. *Mikhaylov D.V.* Metody sozdaniya parallel'no-konveyernykh programm dlya zadach s posledovatel'nym informatsionnym grafom: diss. ... kand. tekhn. nauk po spetsial'nosti 05.13.11 “Matematicheskoe i programmnoe obespechenie vychislitel'nykh mashin, kompleksov i komp'yuternykh setey” [Methods for creating parallel-pipeline programs for problems with a sequential information graph:... cand. of eng. sc. diss. in specialty 05.13.11 “Mathematical and software support for computing machines, complexes and computer networks”], scientific supervisor: dr. of eng. sc., professor Levin I.I. Taganrog, 2022, 213 p.
17. *Mikhaylov D.V.* Preobrazovanie nekotorykh vidov posledovatel'nykh informatsionnykh grafov v parallel'no-konveyernuyu formu [Transformation of Some Types of Sequential Information Graphs into a Parallel-Pipeline Form], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2020, No. 7 (217), pp. 78-93. DOI: 10.18522/2311-3103-2020-7-78-93.
18. *Mikhaylov D.V.* Preobrazovanie posledovatel'nogo informatsionnogo grafa metoda progonki v parallel'nuyu formu [Transformation of the sequential information graph of the sweep method into a parallel form] *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 7 (224), pp. 177-188. DOI: 10.18522/2311-3103-2021-7-177-188.
21. Batcher odd–even mergesort. Available at: https://en.wikipedia.org/wiki/Batcher_odd%E2%80%93even_mergesort (accessed 31 July 2024).
19. *Richard E.* Blahut Fast Algorithms for Digital Signal Processing. Cambridge University Press, 05.2011. DOI: https://doi.org/10.1017/CBO9780511760921.

Статью рекомендовал к опубликованию д.т.н., профессор А.В. Боженюк.

Левин Илья Израилевич – Южный федеральный университет; e-mail: iilevin@sfedu.ru; г. Таганрог, Россия; тел.: 88634612111; зав. кафедрой; д.т.н.; профессор.

Алексеев Кирилл Николаевич – e-mail: alexseev91@mail.ru; тел.: 89283536268; к.т.н.; доцент.

Гуленок Андрей Александрович – Научно-исследовательский центр суперЭВМ и нейрокомпьютеров; e-mail: gulenok@superevm.ru; г. Таганрог, Россия; тел.: 88634612111; начальник сектора; к.т.н.

Levin Ilya Izrailevich – Southern Federal University; e -mail: iilevin@sfedu.ru; Taganrog, Russia; phone: +78634612111; head of department; dr. of eng. sc.; professor.

Alekseev Kirill Nikolayevich – e-mail: alexseev91@mail.ru; phone: +79283536268; cand. of eng. sc.; associate professor.

Gulenok Andrei Aleksandrovich – Supercomputers and Neurocomputers Research Center; e-mail: gulenok@superevm.ru; Taganrog, Russia; phone: +78634612111; head of sector; cand. of eng. sc.

УДК 621.315.611 DOI 10.18522/2311-3103-2024-4-64-79

Н.К. Полуянович, О.В. Качелаев, Т.Э. Фалькон

ОЦЕНКА ВЛИЯНИЯ ГИПЕРПАРАМЕТРОВ НЕЙРОСЕТИ НА ТОЧНОСТЬ ПРОГНОЗИРОВАНИЯ ЭНЕРГОПОТРЕБЛЕНИЯ

Работа посвящена проблеме повышения точности краткосрочного прогнозирования потребления электроэнергии с помощью инструментов глубокого машинного обучения. Исследовано влияние задаваемых гиберпараметров нейронной сети (НС) на погрешность прогнозирования электропотребления, таких как: размер пакета данных – Bs; количество слоев НС – j; функции активации нейронов – Fa; оптимизаторы – O. Определены оптимальные гиперпараметры НС-модели для прогнозирования электропотребления (ЭП) для потребителя аддитивного и циклического типа. Проведенный анализ влияния размера пакета (BatchSize) на точность прогноза показал повышение эффективности обучения НС с ростом пакета входных данных. Проведенный анализ влияние количества слоев показал, что с увеличением количества слоев НС время обучения сокращается, а ее предсказания становятся точнее. Проведенное исследование различных оптимизаторов на скорость обучения показало, что наилучшие результаты демонстрируют оптимизаторы "Adam" и "RMSprop". Показано, что от выбора функции активации зависит, то насколько быстро будет обучаться НС и насколько точными будут ее прогнозы. Использование различных методов регуляризации позволяет НС достичь лучших результатов на практике, улучшая их способность к обобщению и увеличивая точность предсказаний. Показано, что для достижения минимальной погрешности прогнозирования, следует индивидуально настраивать параметры сети для каждого потребителя, учитывая значительные различия в характере энергопотребления. Проведено обучение и тестирование созданной сети с подобранными параметрами на обучающей и тестовой выборке, содержащей данные об электропотреблении за 2 года (17520 часов). Анализ входных данных по электропотреблению показал, что оптимальными параметрами прогностической нейросетевой модели в ручном режиме являются: размер пакета 250 (подобрано эмпирически), 5 слоев, функция активации "ReLU", оптимизатор "Adam". Рассмотрены различные способы подбора гиперпараметров (вручную и посредством генетического алгоритма (ГА)).

Прогнозирование электропотребления; нейронная сеть; атипичный характер электропотребления; надежность систем энергоснабжения.

N.K. Poluyanovich, O.V. Kachelaev, T.H. Falcón

ASSESSMENT OF THE INFLUENCE OF NEURAL NETWORK HYPERPARAMETERS ON THE ACCURACY OF FORECASTING ENERGY CONSUMPTION

The work is devoted to the problem of improving the accuracy of short-term forecasting of electricity consumption using deep machine learning tools. The influence of the specified neural network NN hyperparameters on the error in predicting power consumption, such as: data packet size – Bs; number of NN layers – j; neuron activation functions – Fa; optimizers – O. The optimal hyperparameters of the NN model for predicting electrical consumption (EC) for consumers of additive and cyclic types have been determined. The analysis of the impact of the batch size on the accuracy of the forecast showed an increase in the effectiveness of NN training with the growth of the input data package. The analysis of the influence of the number of layers showed that with an increase in the number of layers of the NN, the learning time decreases and its predictions become more accurate. A study of various optimizers for learning speed has shown that the best results are demonstrated by the optimizers "Adam" and "RMSProp". It is shown that the choice of the activation function determines how quickly the NN will be trained and how accurate its forecasts will be. The use of various regularization methods allows NS to achieve better results in practice, improving their generalization ability and increasing the accuracy of predictions. It is shown that in order to achieve the minimum forecasting error, it is necessary to individually configure the network parameters for each consumer, taking into account significant differences in the nature of energy consumption. The training and testing of the created network with selected parameters was carried out on a training and test sample containing data on electricity consumption for 2 years (17520 hours). The analysis of input data on power consumption showed that the optimal parameters of

the predictive neural network model in manual mode are: package size 250 (selected empirically), 5 layers, activation function "ReLU", optimizer "Adam". Various ways of selecting hyperparameters (manually and by means of genetic algorithm (GA)) are considered.

Cyber-physical system; neural network; atypical nature of power consumption; reliability of power supply systems.

Введение. Текущая ситуация в сфере электроснабжения характеризуется значительным физическим износом оборудования, высокими показателями удельных расходов топлива и потерь в сетях. Для решения задачи поддержания баланса мощностей между генерирующей стороной и потреблением электрической энергии используется прогнозирование ресурса линий электропередач [1–3], состояния ее электроизоляционных свойств [4], влияния электромагнитных полей [5–7], термофлуктуационных воздействий [8–10] и многое другое. Актуальность данной работы обусловлена требованиями законодательства к прогнозированию электропотребления для решения задачи поддержания баланса мощностей между генерирующей стороной и потреблением электрической энергии [11]. Наиболее распространен метод прогноза, основанный на удельных нормах расхода электроэнергии с экспертной корректировкой (корреляционно-регрессионный анализ). В качестве модели для аппроксимации графика нагрузки и его прогноза могут успешно использоваться также полиномиальные разложения. Если необходим прогноз на длительный интервал времени можно использовать авторегрессионные методы. Недостатками этих методов являются: сложность вычислений; применение требует участия аналитиков; применяются только к определенному типу прогноза; – недостаточная точность прогноза; чувствительны к входным данным. Новые подходы к прогнозированию электрической нагрузки базируются на нейросетевом моделировании, позволяющем минимизировать вмешательство в работу системы управления на этапе сбора данных об исследуемом объекте.

Проблема прогнозирования электропотребления заключается в том, что необходимо учесть огромное количество факторов, имеющих влияние на изменение энергопотребления. По оценкам российских специалистов, каждая оплошность в ежегодном прогнозе электропотребления на 1% – это 4 млрд. долларов дополнительных инвестиций на возведение генерирующих мощностей.

Целью прогнозирования электропотребления является обеспечение баланса между выработанной и потребленной электроэнергией.

Постановка задачи. Разработка оптимальной структуры системы прогнозирования для потребителей с различным характером электропотребления т.е. не повторяющимся значением (активная мощность P, реактивная мощность Q) за рассматриваемые периоды времени, а также математической модели прогнозируемого объема электропотребления, включающей в себя кроме целевых параметров электросети учет воздействующих факторов: социально-экономических; технологических и метеорологических.

1. Обзор литературных источников. В настоящее время создано большое количество архитектур нейронных сетей (НС) и методов их обучения. Наиболее распространенной является связка – многослойный персептрон (МП) и алгоритм обратного распространения ошибки [15], но точные значения градиента, в данном методе, не всегда доступны и, кроме того, решение задачи оптимизации в условиях большой размерности требует значительных вычислительных, а, следовательно, и временных затрат.

В работе [16] предлагается модель краткосрочного прогнозирования суточных графиков активной мощности с учетом прогнозных данных естественной освещенности. Указывается на необходимость проведения дополнительные исследования по совершенствованию структуры и подбору экспериментально настраиваемых параметров использованной LSTM сети с целью уменьшения ошибки прогноза.

Другим направлением исследований являются ансамблевые методы [17], чаще других используются рекуррентные сети [18] (Recurrent neural networks, RNN), поскольку они созданы для обработки последовательных данных, в частности временных рядов.

Сравнение глубоких нейронных сетей с классическим многослойным перцептроном, ARIMA и сезонной моделью Хольта-Уинтерса для прогноза электропотребления промышленных предприятий приведено в работе [19]. Недостатками глубоких нейронных сетей являются необходимость обучения на очень большой выборке данных, высо-

кая вычислительная трудоемкость обучения [20], кроме того, высокая трудоемкость выбора архитектуры сети. При этом существует проблема переиспользования обученной сети. Например, если для прогноза электропотребления модель в качестве входа использовала метеорологические данные (температура, скорость ветра, давление, облачность), то без них модель уже не будет работать. И может намного хуже работать в другой климатической зоне.

В исследовании [21] было выяснено, что математическая модель почасового электропотребления должна учитывать помимо стандартных факторов времени, температуры наружного воздуха и освещенности факторы отключения на электрических сетях; производственный режим потребителей электроэнергии; наличие центрального отопления и горячего (холодного) водоснабжения в отдельных населенных пунктах; инерционности изменения режима электропотребления в связи с перепадами температуры воздуха. В связи с учетом вышеперечисленных факторов при краткосрочном прогнозировании потребления электроэнергии необходимо рассмотреть вопросы выбора оптимальной конфигурации нейронной сети, которая позволяет достичь максимальной точности прогноза электропотребления.

Также вышеперечисленные методы прогнозирования потребления электроэнергии основаны на устаревших методах машинного обучения, таких как метод опорных векторов, метод главных компонент, персептроны различной конфигурации и т.п. В настоящее время НС и генетические алгоритмы перешли из перспективных методов решения задач в доступные и широко применяемые методы.

2. Структура НС-модели. Для решения задачи НС прогноза аддитивного потребления электроэнергии разработана модель нейронной сети, представленной на рис. 1. Многофакторная НС-модель содержит во входном слое: X_1–X_8 – входные нейроны для влияющих факторов (дата, метеоусловия, социально-экономические и др.), нейроны X_9–Xnn – являются входами для данных о потребленной мощности за предыдущие nn часов. Ретроспективные данные, подаваемые на вход НС-модели, дают возможность реализовать метод скользящего окна (СО), что позволяет уменьшить ошибку прогнозирования.

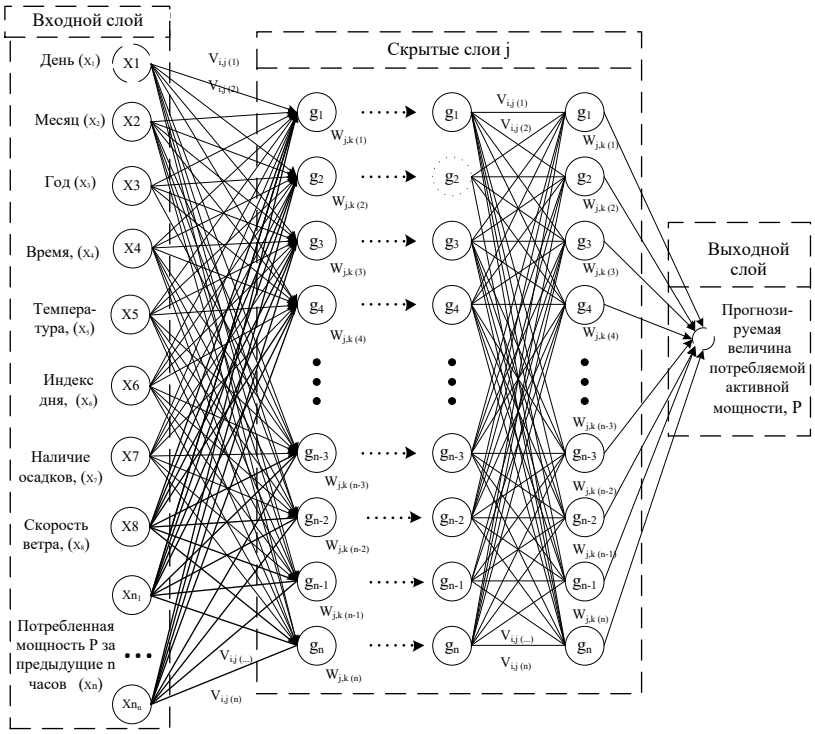


Рис. 1. Структура НС-модели прогнозирования

Предлагаемая топология сети с произвольным количеством слоев и нейронов на слой позволяет применять генетический алгоритм для подбора гиперпараметров НС-модели для конкретного потребителя. А именно: количество скрытых слоев j, количество нейронов N на слой, функция активации нейронов Fa, оптимизатор O. Выходной слой содержит 1 нейрон, соответствующий прогнозным значениям потребленной мощности P. Тип линейной НС с обучением при помощи алгоритма обратного распространения ошибки выбран по причине простоты реализации, и приемлемой точности работы с аддитивными данными о потребленной мощности.

3. Исследование влияния гиперпараметров. Алгоритмы обучения глубоких нейронных сетей, предназначенных для краткосрочного прогнозирования электропотребления, включают оптимизацию в различных контекстах. Так, основная задача оптимизации в машинном обучении заключается в подборе параметров нейронной сети, значительно уменьшающих величину функции стоимости J(n). Часть параметров, которые называются свободными параметрами нейронной сети, определяются в процессе обучения на обучающей выборке данных. Однако существует набор параметров, называемых гиперпараметрами, которые невозможно определить и уточнить в процессе обучения. В список основных параметров архитектур для нейронных сетей входят следующие:

- количество скрытых слоев сети;
- количество нейронов в каждом скрытом слое;
- функция активации нейронов скрытых и выходного слоев;
- оптимизаторы функции потерь.

Показателем точности прогноза выбран средний модуль процента ошибки (MAPE):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^* - y_i}{y_i}\right| \cdot 100\%,$$

где n – число часов в выборке; y^*_i – прогноз электропотребления в i-й час; y_i – истинное значение электропотребления в i-й час.

Выбор MAPE обусловлен тем, что именно этот показатель используется предприятиями как главная метрика точности прогноза электропотребления при работе на оптовом и розничном рынках электроэнергии и мощности.

3.1. Влияние размера пакета (BatchSize). Скорость обучения нейросети зависит от различных факторов, которые влияют на эффективность процесса обучения и время, необходимое для достижения конвергенции модели. На рис. 2 приведена зависимость эффективности обучения (отклонения прогнозирования НС) от задаваемого при обучении размера пакета (Batch size).

Рис. 2. Зависимость отклонения прогнозирования НС, от задаваемого размера пакета

Эффективность обучения (Данный показатель является безразмерным) НС рассчитывается следующим образом:

$$E = \frac{\Delta l}{t} * 100,$$

где E – эффективность обучения, Δl – дельта ошибки. $\Delta l = l_{нач}/l_{кон}$, t – время затраченное на обучение.

Анализ графика на рис. 2 показал, что чем больше пакет входных данных, тем эффективнее обучается НС. Однако, чем больше пакет входных данных, тем больше требуется оперативной памяти для процесса обучения. Поэтому все последующие тесты будут проводится с размером пакета входных данных равным 250, что является оптимальным размером для 8Гб оперативной памяти.

3.2. Влияние количества слоев. Для случая аддитивных входных параметров (данные МАКР) исследовано влияние количества слоев на скорость обучения НС приведено на рис. 3.

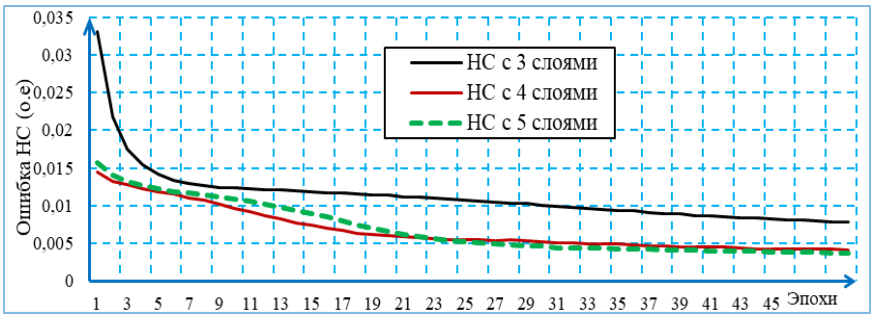


Рис. 3. Влияние количества слоев на скорость обучения НС

Наблюдается понижение ошибки НС прогнозирования с разным количеством слоев. Результаты погрешностей НС-модели с разным количеством слоев, в случае аддитивного характера электропотребления, представлены в табл. 1.

Таблица 1

Влияние количества слоев на точность предсказаний НС

Количество слоев в НС	Количество эпох	Среднее отклонение предсказаний НС, %
3	47	17,10
4	24	13,72
5	24	0,79

Таким образом, анализируя рис. 3 и погрешность НС (табл. 3) видим, что с большим количеством слоев НС обучается быстрее и ее предсказания точнее. Большое значение ошибки является результатом быстрого обучения НС (цель – получение информации о влиянии на отклонение прогнозирования количества скрытых слоев в структуре нейросети).

3.3. Влияние оптимизатора. Оптимизатор – это метод достижения лучших результатов, помощь в ускорении обучения. Другими словами, это алгоритм, используемый для незначительного изменения параметров, таких как веса и скорость обучения, чтобы модель работала правильно и быстро. Для обучения прогностической НС-модели, наиболее распространены следующие оптимизаторы:

Стохастический градиентный спуск (SGD) – это мини-пакетный градиентный спуск стремится сбалансировать устойчивость стохастического градиентного спуска и эффективность пакетного градиентного спуска. Это наиболее распространенная реализация градиентного спуска, используемая в области глубокого обучения. SGD имеет ряд преимуществ перед стандартным градиентным спуском, таких как более быстрая конвергенция и меньшие требования к памяти, особенно для больших наборов данных. Он также более устойчив к зашумленным и нестационарным данным и может выходить за пределы локальных минимумов. Однако для конвергенции может потребоваться больше итераций, чем для градиентного спуска, и скорость обучения должна быть тщательно настроена для обеспечения конвергенции.

Momentum – это метод оптимизации, используемый в машинном обучении и глубоком обучении для ускорения обучения нейронных сетей. Он основан на идее добавления части предыдущего обновления к текущему обновлению весов в процессе оптимизации. При импульсной оптимизации градиент функции затрат вычисляется относительно каждого веса в нейронной сети. В целом, momentum – это мощный метод оптимизации, который может помочь ускорить обучение глубоких нейронных сетей и повысить их производительность.

Среднеквадратичное распространение корня (RMSprop) – это алгоритм оптимизации, используемый в машинном обучении и глубоком обучении для оптимизации обучения нейронных сетей. В целом, RMSProp – это мощный алгоритм оптимизации, который может помочь ускорить обучение глубоких нейронных сетей и повысить их производительность, особенно в ситуациях, когда целевая функция нестационарна.

Адаптивная оценка момента (Adam) – один из самых эффективных алгоритмов оптимизации в обучении нейронных сетей. Adam сочетает в себе идеи RMSProp и оптимизатора импульса. Вместо того чтобы адаптировать скорость обучения параметров на основе среднего первого момента (среднего значения), как в RMSProp, Adam также использует среднее значение вторых моментов градиентов. Adam также включает этап коррекции смещения для корректировки скользящих средних, поскольку в начале процесса оптимизации они смещены к нулю. Это помогает улучшить производительность алгоритма оптимизации на ранних стадиях обучения.

$$\begin{cases} V_{dW} = \beta_1 V_{dW} + (1-\beta_1) dW; \\ S_{dW} = \beta_2 S_{dW} + (1-\beta_2) dW^2; \\ V_{dW}^{corr} = V_{dw}(1-\beta_1^t); \\ S_{dW}^{corr} = S_{dw}(1-\beta_2^t); \\ W = W - \alpha V_{dW}^{corr}\left(\sqrt{S_{dW}^{corr}+\varepsilon}\right)^{-1}, \end{cases} \quad (1)$$

где V_{dw} – матрица, характеризующая инерционные свойства параметров ИНС, по сути, матрица скорости изменения параметров; β_1 – параметр, задающий баланс между учетом предыдущего направления градиента и направления градиента, полученного на очередной эпохе обучения и на очередном пакете, обычно значение этого параметра близко к 1 (~0,9); dW – матрица градиентов, задающих направление повышения ошибки, $\partial E/\partial W$; β_2 – параметр, задающий баланс между учетом предыдущей энергии изменения направления градиента и направления градиента, полученного на очередной эпохе обучения и на очередном пакете, обычно значение этого параметра близко к 1 (~0,999); S_{dw} – матрица, характеризующая степень («энергию», так как градиент возведен в квадрат) изменения параметров ИНС, без учета направления изменения; t – номер пакета при обучении; W – матрица весов; α – величина шага обучения; ε – близкое к нулю положительное число для предотвращения деления на ноль [22].

Сравнение скорости падения ошибки НС, в зависимости от оптимизатора применяемого в процессе обучения приведено на рис. 4.

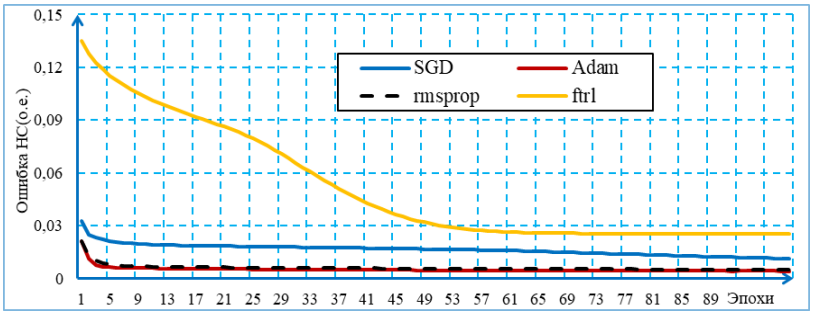


Рис. 4. Сравнение скорости обучения оптимизатора Adam и SGD

Таким образом, анализируя скорость обучения различных оптимизаторов, замечено, что наилучшие результаты демонстрируют оптимизаторы “Adam” и “RMSprop”.

Влияние функции активации. Для нейронов нейронной сети можно использовать любую функцию активации (2), однако есть те, которые принято считать наиболее оптимальными [22].

$$\begin{cases} f(x) = \dfrac{1}{1+e^{-x}}, & Y \in [-1;1] \quad (Sigmoid) \\ Tanh(x) = \dfrac{e^{x}-e^{-x}}{e^{x}+e^{-x}}, & Y \in [-1;1] \quad (Tansig) \\ f(x) = \max(0,x), & Y \in [0;\ +\infty)\ (ReLU) \\ f(x) = \mathrm{x}, & Y \in (-\infty;\ +\infty) \quad (Linear) \\ f(x)_i = \dfrac{e^{x_i}}{\sum_{j=1}^{K} e^{x_i}}, & Y \in [0;\ +\infty) \quad (Softmax) \end{cases} \tag{2}$$

Для выбора оптимальной функции активации нейронной сети проведено сравнение скорости падения ошибки НС, в зависимости от функции активации, рис. 5.

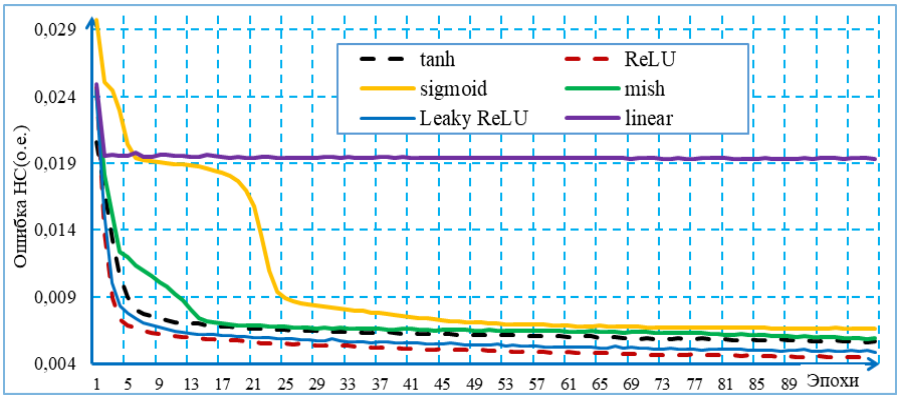


Рис. 5. Сравнение ошибки с разными функциями активации нейронов

Анализируя рис. 5 наблюдаем, что функция активации “ReLU” превосходит остальные функции активации. Так же хорошо себя показали функции активации “Leaky ReLU” и “tanh”.

Таким образом, от выбора функции активации зависит, то насколько быстро будет обучаться нейронная сеть и насколько точными будут ее прогнозы. Так же при выборе функции активации важно нормировать входные данные таким образом, чтобы область определения функции активации совпадала с областью определения данных. Если входные данные будут иметь отрицательные значения, то функция активации должна иметь отрицательные значения в своей области определения, иначе НС будет терять часть информации.

Методы регуляризации. Для борьбы с переобучением можно использовать методы регуляризации [23]. Преимущество регуляризации заключается в том, что она позволяет уменьшить размер весов модели и сделать их более равномерными, тем самым повышая устойчивость модели к малым изменениям данных.

При выборе метода регуляризации важно учитывать, что слишком сильная регуляризация может привести к недообучению, тогда как недостаточная регуляризация может привести к переобучению. Регуляризация нейронных сетей – это мощный инструмент, который помогает контролировать переобучение и повышать обобщающие способности моделей. Путем применения различных методов регуляризации и их оптимального настройки, можно достичь более стабильных и надежных результатов в задачах машинного обучения.

L1-регуляризацию часто называют регрессией лассо или Манхэттенской. Суть данного метода заключается в добавлении к функции потерь суммы абсолютных значений весовых коэффициентов.

$$L_{L1}(Y,Y',W) = L(Y,Y') + \lambda \sum_{i=1}^{n} |w_i|,$$

где $L_{L1}(Y,Y',W)$ – функция потерь с L1-регуляризацией; L(Y,Y') – одна из рассмотренных ранее функций потерь; λ – коэффициент регуляризации (штраф); w_i – i-ый весовой коэффициент.

L2-регуляризация. L2, или гребневая, регуляризация, как и L1-регуляризация, вводит в функцию потерь штраф за большие весовые коэффициенты. Но при этом используется L2-норма — сумма квадратов весовых коэффициентов. В результате функция потерь будет иметь следующий вид.

$$L_{L2}(Y,Y',W) = L(Y,Y') + \lambda \sum_{i=1}^{n} w_i^2$$

L2-регуляризация вводит штраф за большие веса, тем самым усиливает влияние важных признаков и снижает, но не исключает, влияние случайного шума на конечный результат.

Регуляризация нейросетей с помощью dropout. Одной из самых эффективных и широко применяемых техник регуляризации нейронных сетей является dropout. Dropout – это метод, который случайным образом удаляет (отключает) некоторые нейроны в процессе обучения, чтобы уменьшить переобучение. Dropout является эффективным методом борьбы с переобучением и может значительно улучшить обобщающую способность нейросетей. Преимуществом использования dropout является: – уменьшение переобучения; –улучшение обобщающей способности; – эффективность. Dropout может быть вычислен параллельно и не требовать дополнительные вычислительные ресурсы. Dropout – это мощный инструмент для улучшения обобщающей способности и предотвращения переобучения нейросетей.

Таким образом, регуляризация играет важную роль в улучшении обобщающей способности нейронных сетей. Она помогает избежать переобучения и сделать модель более устойчивой к новым данным. Использование различных методов регуляризации позволяет нейронным сетям достичь лучших результатов на практике, улучшая их способность к обобщению и увеличивая точность предсказаний.

4. Экспериментальные исследования

4.1. Многофакторное моделирование электропотребления. Для обучения и тестирования алгоритмов краткосрочного прогнозирования электропотребления, необходимо произвести выборку и описание входных параметров. Кроме целевых данных (основных), которыми являются прогнозные значения потребленной активной P(t), и реактивной Q(t) мощности, необходимы следующие значимые факторы:

- Социально-экономические (Время, (t – час, сутки, неделя, месяц); Дата (D – день, M – месяц, Y – год); Статус дня (S; 1– рабочий, 0 – выходной))
- Метеорологические (Температура среды (T); Наличие осадков (R; 1 – осадки есть, 0 – осадков нет); Скорость ветра (W))

Обучение и тестирование НС проводилось для потребителей с различным характером электропотребления (циклическим, и аддитивным). Размер выборки составил данные о потреблённой ЭЭ и влияющих факторах (табл. 3) за два года по каждому из потребителей. Частота снятия данных P(t), и Q(t) составляет не менее, чем раз в 1 час.

4.2. Настройка НС. Алгоритмы обучения глубоких нейронных сетей, предназначенных для краткосрочного прогнозирования электропотребления, включают оптимизацию в различных контекстах. Так, основная задача оптимизации в машинном обучении заключается в подборе параметров нейронной сети, значительно уменьшающих величину функции стоимости J(n). Часть параметров, которые называются свободными параметрами нейронной сети, определяются в процессе обучения на обучающей выборке данных.

Однако существует набор параметров, называемых гиперпараметрами, которые невозможно определить и уточнить в процессе обучения. В список основных параметров архитектур для нейронных сетей входят следующие: – количество скрытых слоев сети; – количество нейронов в каждом скрытом слое; – функция активации нейронов скрытых и выходного слоев. Некоторые варианты (комбинаций параметров НС) для потребителей с различным характером электропотребления, приведены в табл. 4.

Таблица 4

Комбинации параметров НС

Номер комбинации параметров НС	Количество слоев НС, j	Число нейронов в слое, n	Скорость обучения, Lr	Функция активации, Fa	Тип ошибки
Комбинации параметров НС-модели для потребителя МАКР					
1	1	[10]	0.01	[Tansig]	MSE
2	1	[10]	0.01	[Tansig]	SSE
3	1	[10]	0.0001	[Tansig]	MSE
Комбинации параметров НС-модели для потребителя ЮЗЭС					
1	2	[10;10]	0.01	[Tansig, Tansig]	MSE
2	1	[25]		[Tansig]	MSE
3	1	[30]		[Tansig]	MSE
Комбинации параметров НС-модели для потребителя МАКР со скользящим окном					
1	1	[10]	0.0001	[Tansig]	MSE
2	1	[10]	0.001	[Tansig]	MSE
3	1	[10]	0.00001	[Tansig]	MSE

4.3. Тестирование нейросети

4.3.1. Оценка влияния гиперпараметров НС на погрешность прогноза. Рассматривались такие гиперпараметры НС как: тип распространения ошибки, скорость обучения и обучающая функция. Погрешность прогнозирования ЭП с различными вариантами гиперпараметров НС, приведены в табл. 5.

Таблица 5

Оценка влияния гиперпараметров НС на погрешность прогноза

Параметр	Значение	Погрешность прогнозирования %	
		Средняя по часам	Средняя за период
Обучающая функция	trainbr	31,79	15,79
	trainlm	259,79	244,38
Тип ошибки	MSE	30,28	6,14
	SSE	26,31	14,38
Скорость обучения, о.е.	0,00001	25,19	15,95
	0,0001	30,28	6,14
	0,01	31,12	20,91

Проведена оценка влияния гиперпараметров НС на погрешность прогноза. Рассчитаны погрешности прогнозирования ЭП при различных комбинациях параметров НС-модели приведены в табл. 6.

Таблица 6

Результаты прогнозирования электропотребления

Способ измерения погрешности	Номер комбинации параметров НС-модели	Погрешность прогнозирования %
Средний по часам	1	30.76
	2	26.80
	3	26.31
Средний за период	1	12.81
	2	15.59
	3	14.38

Проведено исследование для двух различных типов потребителей ЭЭ (аддитивного, и имеющего циклический характер). Прогнозирование потребителя электроэнергии с аддитивным характером ЭП (МАКР) [15], для трех вариантов настройки НС-модели (табл. 4), приведено на рис. 6.

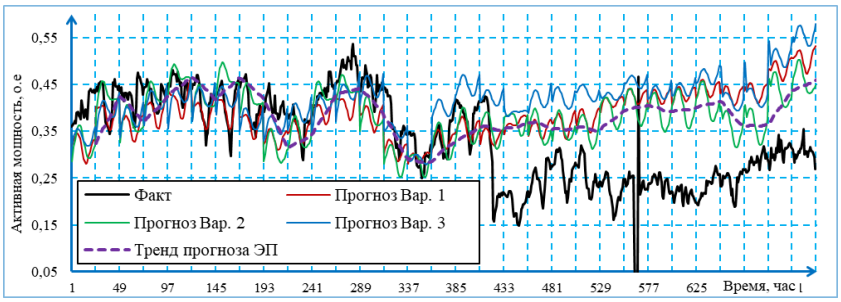


Рис. 6. График сравнения фактических и прогнозных значений (Вариант 1-3)

Анализ полученных зависимостей (рис. 6) прогнозирования электропотребления, с использованием НС-модели типа «многослойный персептрон» со скользящим окном показал, что в момент резкого скачка фактического ЭП (18 числа) (более чем в 2-2.5 раза) погрешность используемой НС-модели существенно увеличилась, что говорит о некорректной работе данной НС-модели. НС-модель небольшого размера (табл. 4) не обеспечивает требуемой точности. Использование метода СО при прогнозировании ЭП должно уменьшить погрешность прогнозирования на подобных участках кривой электропотребления.

4.3.2. НС-модель со скользящим окном. Для выбора оптимальной нейросети проведено исследование влияния использования метода скользящего окна на точность прогнозирования электропотребления аддитивного характера. Параметры некоторых НС-моделей, обученных с методом СО приведены в табл. 4. Прогнозирование потребителя электроэнергии с аддитивным характером ЭП с использованием СО, для трех вариантов настройки НС-модели (табл. 4), приведено на рис. 7.

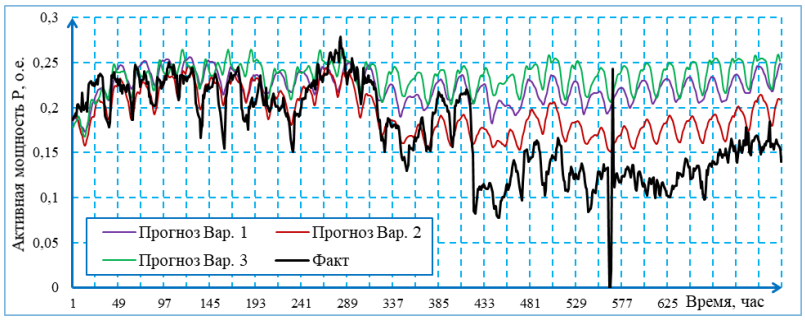


Рис. 7. Графики сравнения фактических и прогнозных значений ЭП при использовании скользящего окна

Анализ графиков прогнозирования ЭП при помощи НС-модели со скользящим окном показал, что 2 вариант комбинаций параметров НС-модели дает наиболее приближенную к фактической кривую электропотребления. Это говорит о целесообразности использования данного метода и возможности подбора гпиперпараметров НС-модели. Результаты расчетов погрешности прогнозирования электропотребления при использовании скользящего окна сведены в табл. 7.

Таблица 7

Результаты прогнозирования электропотребления со скользящим окном

Способ измерения погрешности	Номер комбинации параметров НС-модели	Погрешность прогнозирования, %
Средний по часам	1	39.85
	2	47,69
	3	23.18
Средний за период	1	27.33
	2	33.42
	3	10.79

Расчеты погрешностей (табл. 9) показали, что:

- Расчет погрешности в виде среднего отклонения по каждому часу имеет смысл применять при оперативном или краткосрочном прогнозировании. Использование метода скользящего окна показало снижение ошибки прогнозирования в среднем на 2% при использовании подобного способа расчета.
- Расчет погрешности в виде среднего значения сумм потребленной и спрогнозированной электроэнергии за прогнозный период имеет смысл использовать при среднесрочном и долгосрочном прогнозировании. Использование метода скользящего окна показало снижение ошибки прогнозирования в среднем на 17% при использовании подобного способа расчета.

Таким образом, результаты проведенных экспериментальных исследований позволили установить оптимальную комбинацию параметров для аддитивного потребителя электроэнергии: – кол-во скрытых слоев (1); количество нейронов на слой [10]; функция активации Tansig, а в выходном слое Purelin; тип распространения ошибки (MSE); скорость обучения (0,001); обучающая функция (TrainBR - Байесовская регуляризация), использование метода СО.

4.3.3. Прогнозирования, с циклическим характером ЭП. Графики прогнозирования, с циклическим характером ЭП (ЮЗЭС), для трех вариантов настройки НС-модели (табл. 4), приведены на рис. 8-10, где на оси времени откладываются интервалы равные 0.5 часа.

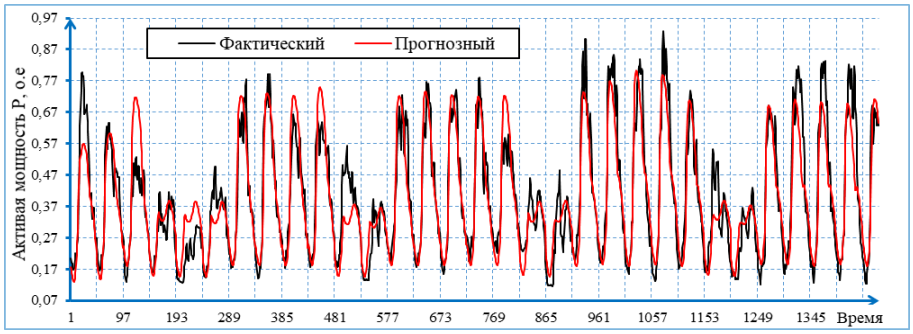


Рис. 8. График сравнения фактических и прогнозных значений (Вариант 1)

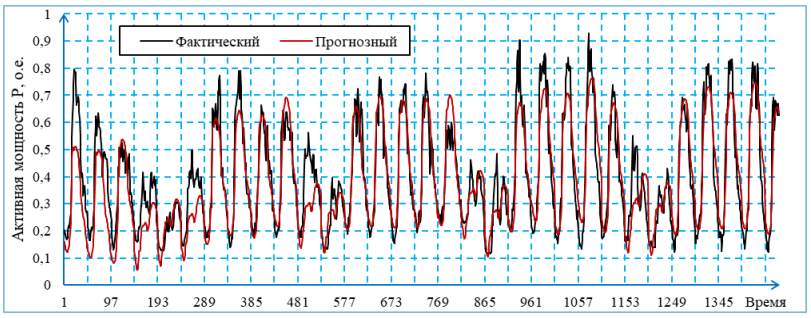


Рис. 9. График сравнения фактических и прогнозных значений (Вариант 2)

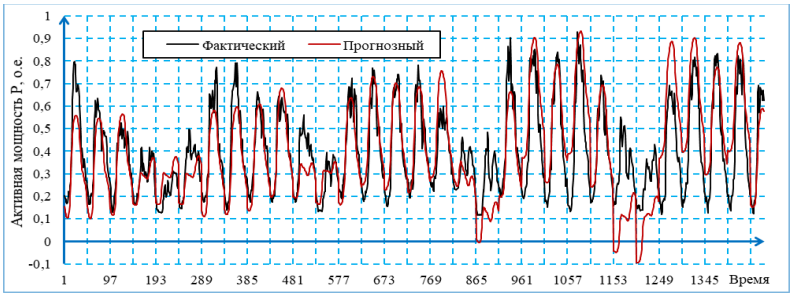


Рис. 10. График сравнения фактических и прогнозных значений (Вариант 3)

Результаты погрешностей при различных комбинациях параметров НС-модели приведены в табл. 8

Таблица 8

Результаты прогнозирования электропотребления

Способ измерения погрешности	Номер комбинации параметров НС-модели	Ошибка прогнозирования ε, %
Средняя по часам	1	15,84
	2	18,17
	3	30,30
Средняя за период	1	1,36
	2	3,81
	3	0,04

Проведена оценка влияния гиперпараметров НС на погрешность прогноза. Рассматривались такие гиперпараметры как: тип распространения ошибки, скорость обучения и обучающая функция. Погрешность прогнозирования с различными гиперпараметрами приведены в табл. 9.

Таблица 9

Оценка влияния гиперпараметров НС на погрешность прогноза

Параметр НС	Значение параметра	Погрешность прогнозирования %	
		Средняя по часам	Средняя за период
Тип ошибки	MSE	18,13	4,81
	SSE	18,48	6,29
Скорость обучения, о.е.	0,00001	17,44	1,16
	0,0001	18,7	0,76
	0,01	19,23	5,92
Обучающая функция	trainbr	16,31	8,87
	trainlm	152,72	81,95

Таким образом, по результатам проведенных экспериментальных исследований выявлена оптимальная комбинация параметров НС-модели для прогнозирования электропотребления с циклическим характером: кол-во скрытых слоев (2); количество нейронов на слой [10 10]; функция активации Tansig-Tansig, а в выходно слое Purelin); тип распространения ошибки (MSE); – скорость обучения (0,0001); обучающая функция (TrainBR - Байесовская регуляризация).

Заключение. Проведено исследование влияния различных параметров и структуры нейросети на точность прогнозирования энергопотребления с целью нахождения оптимальных ее настроек (наименьшая погрешность). Определены наиболее соответствующие гиперпараметры НС-модели для прогнозирования ЭП для потребителей аддитивного и циклического типа. При прогнозировании электропотребления с циклическим характером оптимальные параметры НС: кол-во скрытых слоев (2); количество нейронов на слой [10 10]; функция активации Tansig-Tansig, а в выходно слое Purelin); тип распространения ошибки (MSE); скорость обучения (0,0001); обучающая функция (TrainBR - Байесовская регуляризация). В случае аддитивного потребителя электроэнергии подобраны оптимальные параметры: – кол-во скрытых слоев (1); количество нейронов на слой [10]; функция активации Tansig, а в выходном слое Purelin; тип распространения ошибки (MSE); скорость обучения (0,001); обучающая функция (TrainBR - Байесовская регуляризация), предпочтительно использование метода СО. Выявлено, что для обеспечения наименьшей погрешности прогнозирования, необходима индивидуальная настройка параметров нейросети для каждого потребителя в отдельности с учётом его характера энергопотребления.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Полуянович Н.К., Дубяго М.Н.* Алгоритм обучения искусственной нейронной сети факторного прогнозирования ресурса изоляционных материалов силовых кабельных линий // Известия ЮФУ. Технические науки. – 2021. – № 2. – С. 59-73.
2. *Полуянович Н.К., Дубяго М.Н.* Оценка воздействующих факторов и прогнозирование электропотребления в региональной энергосистеме с учетом режима ее эксплуатации // Известия ЮФУ. Технические науки. – 2022. – № 2 (226). – С. 31-46.
3. *Полуянович Н.К., Дубяго М.Н., Бурьков Д.В.* Термофлуктуационная теория разрушения и оценка долговечности электрической изоляции СПЭ кабелей // Известия ЮФУ. Технические науки. – 2021. – № 6. – С. 180-194.
4. *Poluyanovich N.K., Dubyago M.N.* Book Chapter. Structural Changes During Electrical Aging of Insulation Materials of Cable Networks // Studies in Systems, Decision and Control. – 2023. – 477. – P. 245-256. – https://doi.org/10.1007/978-3-031-33159-6_19.
5. *Poluyanovich N.K., Sharykin A.A., Dubyago M.N., Kachelaev O.V.* Conference Paper. Analysis of the Electro-magnetic Field on Cable Systems with Insulation Made of Polymer Materials // Proceedings - 2023 5th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency, SUMMA 2023. – 2023. – P. 977-982.
6. *Shiyan Ren; Siwei Guo; Xiaojun Liu; Qingxin Liu.* Shielding Effectiveness of Double-layer Magnetic Shield of Current Comparator Under Radial Disturbing Magnetic Field // IEEE Transactions on Magnetics. – 2016. – Vol. PP, Issue 99. – P. 1.
7. IEE Guide for Field Testing of Shielded Power Cable Systems Using Very Low Frequency (VLF) (less than 1 Hz). IEE.400.2. – 2013.
8. *Ukil A., Braendle H., Krippner P.* Distributed temperature sensing: review of technology and applications // Sensors Journal, IEEE. – 2012. – Vol. 12, No. 5. – P. 885-892.
9. *Diaz-Aguilo M., De Leon F.* Introducing Mutual Heating Effects in the Lad-der-Type Soil Model for Dynamic Thermal Rating of Underground Cables // IEEE Tran. on Power Delivery. – 2015. – Vol. 30 (4). – P. 1958-1964.
10. *Poluyanovich N.K., Dubyago M.N., Azarov N.V., Ogrenichev A.V., Sharykin A.A.* Multifactor Model for Forecasting Thermal Processes in the Insulating Materials of Cable Lines // Conference proceedings // 2020 International Multi-Conference on Indus-trial Engineering and Modern Technologies, FarEastCon. – 2020. – P. 9271367. – ISBN: 978-172816951-4. – DOI: 10.1109/FarEastCon50210.2020.9271367.
11. Постановление Правительства РФ от 9 сентября 2023 г. N 1473 "Об утверждении ком-плексной государственной программы Российской Федерации "Энергосбережение и повышение энергетической эффективности" // Размещен в СПС "Консультант Плюс". – https://base.garant.ru/407632842/.
12. *Дубяго М.Н., Полуянович Н.К.* Совершенствование методов диагностики и прогнозирования электроизоляционных материалов систем энергоснабжения: монография. – Ростов-на-Дону; Таганрог: Изд-во ЮФУ, 2019. – 192 с.

13. *Sergeev N.N., Matrenin P.V.* Enhancing Efficiency of Ensemble Machine Learning Models for Short-Term Load Forecasting through Feature Selection // 2022 IEEE 23rd International Conference of Young Professionals in Electron Devices and Materials (EDM): Proc., Altai, 30 June – 04 July 2022. – Altai: IEEE, 2022. – P. 368-371.
14. *Rafi S.H., Nahid-Al-Masood, Deeba S.R., Hossain E.* A short-term power load forecasting method using integrated CNN and LSTM network // IEEE Access. – 2021. – Vol. 9. – P. 32436-32448.
15. *Deng Z., Wang B., Xu Y. [et al.].* Multi-scale convolutional neural network with time-cognition for multi-step short- term power load forecasting // IEEE Access. – 2019. – Vol. 7. – P. 88058-88071.
16. *Вялкова С.А., Надтока И.И., Корнюкова О.А.* Применение нейронных сетей для прогнозирования электропотребления мегаполиса // Russian Internet Journal of Industrial Engineering. – 2023. – Vol. 10, No. 4.
17. *Поляхов Н.Д., Приходько И.А., Ван Ефэн*. Прогнозирование электропотребления на основе метода опорных векторов с использованием эволюционных алгоритмов оптимизации // Современные проблемы науки и образования. – 2013. – № 2.
18. *Qiu X., Zhang L., Ren Y., Suganthan P., Amaratunga G.* Ensemble deep learning for regression and time series forecasting // Symposium on Computational Intelligence in Ensemble Learning. – IEEE, 2014. – P. 21-26. – DOI: 10.1109/CIEL.2014.7015739.
19. *Kong W., Dong Z., Jia Y., Hill D., Xu Y., Zhang Y.* Short-term residential load forecasting based on LSTM recurrent neural network // IEEE Transactions on Smart Grid. – 2019. – Vol. 10, No. 1. – P. 841-851. – DOI: 10.1109/TSG.2017.2753802.
20. *Ryu S., Noh J., Kim H.* Deep neural network based demand side short term load forecasting // Energies. – 2016. – Vol. 10, No. 1. – P. 1-20. – DOI: 10.3390/en10010003.
21. *Chen K., Wang Q., He Z., Hu J., He J.* Short-Term Load Forecasting with Deep Residual Networks // IEEE Trans. Smart Grid. – 2019. – Vol. 10, No. 4. – P. 3943-3952. – DOI: 10.1109/TSG.2018.2844307.
22. *Хомутов С.О., Серебряков Н.А.* Создание нейросетевой математической модели краткосрочного прогнозирования электропотребления электротехнического комплекса участка районных электрических сетей 6–35 кВ // Транспортные системы и технологии. – 2020. – Т. 6, № 1. – С. 80-91. – DOI: 10.17816/ transsyst20206180-91.
23. Основные функции активации в нейронных сетях // Статьи по программированию. – URL: https://www.yourtodo.ru/posts/osnovnyie-funktsii-aktivatsii-v-nejronnyih-setyah/ (04.07.2024).
24. Приемы повышения сходимости нейронных сетей. Регуляризация // MQL5; трейдинг, автоматические торговые системы, тестирование стратегий и технические индикаторы на MetaTrader. – URL: https://www.mql5.com/ru/neurobook/index/about_ai/improvement/regularization/ (04.07.2024).

REFERENCES

1. *Poluyanovich N.K., Dubyago M.N.* Algoritm obucheniya iskusstvennoy neyronnoy seti faktornogo prognozirovaniya resursa izolyatsionnykh materialov silovykh kabel'nykh liniy [Algorithm for processing an artificial neural network for factor forecasting of the resource of insulating materials of power cable lines], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 2, pp. 59-73.
2. *Poluyanovich N.K., Dubyago M.N.* Otsenka vozdeystvuyushchikh faktorov i prognozirovanie elektropotrebleniya v regional'noy energosisteme s uchetom rezhima ee ekspluatatsii [Assessment of the current factors and forecasting of power consumption in the regional power system, taking into account its operating mode], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 2 (226), pp. 31-46.
3. *Poluyanovich N.K., Dubyago M.N., Bur'kov D.V.* Termofluktuatsionnaya teoriya razrusheniya i otsenka dolgovechnosti elektricheskoy izolyatsii SPE kabeley [Thermal fluctuation theory of destruction and durability assessment-news of electrical insulation of SPE cables], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 6, pp. 180-194.
4. *Poluyanovich N.K., Dubyago M.N.* Book Chapter. Structural Changes During Electrical Aging of Insulation Materials of Cable Networks, *Studies in Systems, Decision and Control,* 2023, 477, pp 245-256. Available at: https://doi.org/10.1007/978-3-031-33159-6_19.
5. *Poluyanovich N.K., Sharykin A.A., Dubyago M.N., Kachelaev O.V*. Conference Paper. Analysis of the Electro-magnetic Field on Cable Systems with Insulation Made of Polymer Materials, *Proceedings - 2023 5th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency, SUMMA 2023,* 2023, pp. 977-982.
6. *Shiyan Ren; Siwei Guo; Xiaojun Liu; Qingxin Liu.* Shielding Effectiveness of Double-layer Magnetic Shield of Current Comparator Under Radial Disturbing Magnetic Field, *IEEE Transactions on Magnetics,* 2016, Vol. PP, Issue 99, pp. 1.
7. IEE Guide for Field Testing of Shielded Power Cable Systems Using Very Low Frequency (VLF) (less than 1 Hz). IEE.400.2, 2013.

8. *Ukil A*., *Braendle H., Krippner P.* Distributed temperature sensing: review of technology and applications, *Sensors Journal, IEEE,* 2012, Vol. 12, No. 5, pp. 885-892.
9. *Diaz-Aguilo M., De Leon F*. Introducing Mutual Heating Effects in the Lad-der-Type Soil Model for Dynamic Thermal Rating of Underground Cables, *IEEE Tran. on Power Delivery,* 2015, Vol. 30 (4), pp. 1958-1964.
10. *Poluyanovich N.K., Dubyago M.N., Azarov N.V., Ogrenichev A.V., Sharykin A.A.* Multifactor Model for Forecasting Thermal Processes in the Insulating Materials of Cable Lines // Conference proceedings // 2020 International Multi-Conference on Indus-trial Engineering and Modern Technologies, FarEastCon. – 2020. – P. 9271367. – ISBN: 978-172816951-4. – DOI: 10.1109/FarEastCon50210.2020.9271367.
11. Postanovlenie Pravitel'stva RF ot 9 sentyabrya 2023 g. N 1473 "Ob utverzhdenii kompleksnoy gosudarstvennoy programmy Rossiyskoy Federatsii "Energosberezhenie i povyshenie energeticheskoy effektivnosti" [Decree of the Government of the Russian Federation No. 1473 dated September 9, 2023 "On approval of the comprehensive State program of the Russian Federation "Energy Saving and energy Efficiency improvement"], *Razmeshchen v SPS "Konsul'tant Plyus"* [Posted in the SPS Consultant Plus]. Available at: https://base.garant.ru/407632842/.
12. *Dubyago M.N., Poluyanovich N.K.* Sovershenstvovanie metodov diagnostiki i prognozirovaniya elektroizolyatsionnykh materialov sistem energosnabzheniya: monografiya [Improvement of methods of diagnostics and forecasting of electrical insulating materials of power supply systems: monograph]. Rostov-on-Don; Taganrog: Izd-vo YuFU, 2019, 192 p.
13. *Sergeev N.N., Matrenin P.V.* Enhancing Efficiency of Ensemble Machine Learning Models for Short-Term Load Forecasting through Feature Selection, *2022 IEEE 23rd International Conference of Young Professionals in Electron Devices and Materials (EDM): Proc., Altai, 30 June – 04 July 2022*. Altai: IEEE, 2022, pp. 368-371.
14. *Rafi S.H., Nahid-Al-Masood, Deeba S.R., Hossain E.* A short-term power load forecasting method using integrated CNN and LSTM network, *IEEE Access,* 2021, Vol. 9, pp. 32436-32448.
15. *Deng Z., Wang B., Xu Y. [et al.]*. Multi-scale convolutional neural network with time-cognition for multi-step short- term power load forecasting, *IEEE Access,* 2019, Vol. 7, pp. 88058-88071.
16. *Vyalkova S.A., Nadtoka I.I., Kornyukova O.A*. Primenenie neyronnykh setey dlya prognozirovaniya elektropotrebleniya megapolisa [Application of neural networks for forecasting the power consumption of a megalopolis], *Russian Internet Journal of Industrial Engineering,* 2023, Vol. 10, No. 4.
17. *Polyakhov N.D., Prikhod'ko I.A., Van Efen.* Prognozirovanie elektropotrebleniya na osnove metoda opornykh vektorov s ispol'zovaniem evolyutsionnykh algoritmov optimizatsii [Forecasting of power consumption based on the method of support vectors using evolutionary optimization algorithms], *Sovremennye problemy nauki i obrazovaniya* [Modern problems of science and education], 2013, No. 2.
18. *Qiu X., Zhang L., Ren Y., Suganthan P., Amaratunga G.* Ensemble deep learning for regression and time series forecasting, *Symposium on Computational Intelligence in Ensemble Learning*. IEEE, 2014, pp. 21-26. DOI: 10.1109/CIEL.2014.7015739.
19. *Kong W., Dong Z., Jia Y., Hill D., Xu Y., Zhang Y.* Short-term residential load forecasting based on LSTM recurrent neural network, *IEEE Transactions on Smart Grid,* 2019, Vol. 10, No. 1, pp. 841-851. DOI: 10.1109/TSG.2017.2753802.
20. *Ryu S., Noh J., Kim H.* Deep neural network based demand side short term load forecasting, *Energies,* 2016, Vol. 10, No. 1, pp. 1-20. DOI: 10.3390/en10010003.
21. *Chen K., Wang Q., He Z., Hu J., He J.* Short-Term Load Forecasting with Deep Residual Networks, *IEEE Trans. Smart Grid,* 2019, Vol. 10, No. 4, pp. 3943-3952. DOI: 10.1109/TSG.2018.2844307.
22. *Khomutov S.O., Serebryakov N.A.* Sozdanie neyrosetevoy matematicheskoy modeli kratkosrochnogo prognozirovaniya elektropotrebleniya elektrotekhnicheskogo kompleksa uchastka rayonnykh elektricheskikh setey 6–35 kV [Creation of a neural network mathematical model for short-term forecasting of electrical consumption of an electrotechnical complex of a section of district electric networks 6-35 kV], *Transportnye sistemy i tekhnologii* [Transport systems and technologies], 2020, Vol. 6, No. 1, pp. 80-91. DOI: 10.17816/ transsyst20206180-91.
23. Osnovnye funktsii aktivatsii v neyronnykh setyakh [Basic activation functions in neural networks], *Stat'i po programmirovaniyu* [Articles on programming]. Available at: https://www.yourtodo.ru/posts/osnovnyie-funktsii-aktivatsii-v-nejronnyih-setyah/ (04 July 2024).
24. Priemy povysheniya skhodimosti neyronnykh setey. Regulyarizatsiya [Techniques for increasing the convergence of neural networks. Regularization], *MQL5; treyding, avtomaticheskie torgovye sistemy, testirovanie strategiy i tekhnicheskie indikatory na MetaTrader* [MQL5; trading, automatic trading systems, strategy testing and technical indicators on MetaTrader]. Available at: https://www.mql5.com/ru/neurobook/index/about_ai/improvement/regularization/ (04 July 2024).

Статью рекомендовал к опубликованию д.т.н., профессор Н.Е. Сергеев.

Полуянович Николай Константинович – Южный федеральный университет; e-mail: nik1-58@mail.ru; г. Таганрог, Россия; тел.: 89508609983; кафедра электротехники и мехатроники; к.т.н.; доцент.

Качелаев Олег Вадимович. – e-mail: 22.olezhka@mail.ru; тел.: 89281163960; кафедра техносферной безопасности и химии; аспирант.

Фалькон Талия Эрнандез – Университет Камагуэя "Игнасио Аграмонте"; e-mail: thaliahernandezf@gmail.com; Камагуэй, Куба; phone: +5355439090; Электротехнический факультет. Инженерно-электромеханический факультет; технический ассистент преподавателя.

Poluyanovich Nikolay Konstantinovich. – Southern Federal University; e-mail: nik1-58@mail.ru; Taganrog, Russia; phone: +79508609983; the department of Electric Technics and Mechatronics; cand. of eng. sc.; associate professor.

Kachelaev Oleg Vadimovich – e-mail: 22.olezhka@mail.ru; phone: +79281163960; the department of Techno-sphere Safety and Chemistry; postgraduate student.

Falcón Talia Hernandez – The University of Camagüey "Ignacio Agramonte"; e-mail: thaliahernandezf@gmail.com; Camagüey, Cube; phone: +5355439090; Electrical department. Electromechanical engineering faculty; teacher technical assistant.

Раздел II. Анализ данных и моделирование

УДК 681.3.062 DOI 10.18522/2311-3103-2024-4-80-91

С.И. Клевцов

КЛАССИФИКАЦИЯ СТЕПЕНИ ИЗМЕНЕНИЯ ПАРАМЕТРА В РЕАЛЬНОМ ВРЕМЕНИ НА ОСНОВЕ АНАЛИЗА ОБЛАКА ТОЧЕК ВРЕМЕННОГО РЯДА

Задача построения модели оценки работоспособности технического объекта имеет множество применений в области контроля различных опасных ситуаций. Необходимость опережающего мониторинга состояния технического объекта для предотвращения и управления протеканием нештатных ситуаций в целях их ликвидации с минимальными последствиями делает постановку и выполнение этой задачи актуальной и своевременной. Для выполнения оценки состояния технического объекта целесообразно использовать простые модели, позволяющие получить результат в режиме реального времени без существенной нагрузки на микроконтроллерную систему управления. В работе рассматривается построение модели классификации динамики изменения параметра технического объекта, что позволит прогнозировать изменение его состояния в процессе оценки степени работоспособности объекта. Используются данные, отражающих изменение параметров в реальном времени и представленные в виде временных рядов значений параметров. Изменение параметра объекта во времени фиксируется с помощью временного окна, которое двигается вдоль временного ряда, вырезая из множества исходных данных подмножество с неизменным количеством временных отсчетов. Для классификации динамики изменения параметра предложено использовать представление точек временного окна в виде графика Пуанкаре, который фактически является особым типом графика повторения или видом диаграммы рассеяния. В качестве критерия используется коэффициент сжатия эллипса (эллиптичность), который охватывает облако точек, формируемое при построении диаграммы рассеяния, для временного ряда технического параметра. Разработана методика обучения и использования модели, включая формирование классов состояний динамики параметра объекта и вычисление критериев. Проведена апробация модели. Модель обеспечивает реализацию процедур выявления в реальном времени возможности возникновения нештатной ситуации на ранней стадии ее развития с помощью микропроцессорного модуля, расположенного на нижнем уровне системы мониторинга объекта.

Идентификация; состояние; оценка; технический объект; параметр; микроконтроллер; классификация.

S.I. Klevtsov

CLASSIFICATION OF THE DEGREE OF PARAMETER CHANGE IN REAL TIME BASED ON TIME SERIES POINT CLOUD ANALYSIS

The task of building a model for assessing the performance of a technical object has many applications in the field of controlling various hazardous situations. The need for advanced monitoring of the technical object state to prevent and control the course of abnormal situations in order to eliminate them with minimal consequences makes the statement and fulfillment of this task relevant and timely. To perform the assessment of the state of the technical object it is advisable to use simple models that allow to obtain the result in real time without significant load on the microcontroller control system. The paper considers the construction of a model for classifying the dynamics of change in the parameter of a technical object, which will allow you to predict the change in its state in the process of assessing the degree of serviceability of the object. The data reflecting the change of parameters in real time and presented in the form of time series of parameter values are used. The change of the object parameter in time is fixed with the help of a time window, which moves along the time series, cutting out of the set of initial data a subset with an unchanged number of time samples. To classify the dynamics of parameter variation, it is

proposed to use a representation of the time window points in the form of a Poincaré plot, which is actually a special type of repetition plot or a type of scatter plot. The ellipse compression factor (ellipticity) is used as a criterion, which encompasses the point cloud formed during the construction of the scatter diagram, for the time series of the technical parameter. A methodology for training and using the model, including the formation of classes of states of the dynamics of the object parameter and the calculation of criteria, is developed. The model has been tested. The model provides the realization of procedures for real-time detection of the possibility of an abnormal situation at an early stage of its development with the help of a microprocessor module located at the lower level of the object monitoring system.

Identification; condition; evaluation; technical object; parameter; microcontroller classification.

Введение. Задача идентификации состояний объектов на основе анализа параметров имеет множество применений в области контроля различных опасных ситуаций. Система классификации для идентификации состояний объекта может представлять собой набор подмножеств значений параметров, которые не пересекаются [1, 2]. Тогда для классификации состояния объекта необходимо отнесение совокупности параметров, снятых датчиками с анализируемого объекта, к одному из нескольких попарно не пересекающихся подмножеств значений параметров. Датчики, являющиеся компонентами системы мониторинга объекта, передают сигналы для обработки в микроконтроллерном модуле. В зависимости от типа датчика либо выполняется оцифровка аналогового сигнала и, далее, обработка данных, если датчик аналоговый, или сразу осуществляется обработка в случае цифрового датчика [3–5].

В процессе обработки выполняется оценка состояния объекта с использованием системы классификации. Оценка выполняется в реальном времени. Часто вычислительные возможности микроконтроллера могут быть достаточно скромными. Возможно, что микроконтроллерный модуль решает широкий спектр задач и на каждую задачу выделяется ограниченная вычислительная мощность. Поэтому к алгоритму оценки состояния на основе классификации предъявляются требования низкой сложности алгоритма [3, 6, 7].

Идентификация состояния объекта в реальном времени требует фиксации значений параметров в конкретный промежуток времени и соотнесение этой совокупности значений параметров к определенным классам. Как правило, совокупность параметров представляется в виде вектора. Компонентами вектора являются значения параметров, фиксируемых датчиками [8–11]. Таким образом, имеется N-мерное пространство, которое часто называют пространством входов, размерность пространства равно количеству компонент вектора.

Выбранный или разработанный алгоритм классификации должен на основе раннее полученной предварительной информации в виде вектора N-мерного пространства определить, к какому классу необходимо отнести данную совокупность параметров, представленную в виде компонент вектора или, иначе, к какому классу состояний отнести текущее параметрическое представление объекта [12, 13]. При этом предполагается, что подмножества значений параметров, характеризующих конкретные состояния объекта, не пересекаются. То есть, каждое подмножество значений параметров задает конкретный класс состояния объекта.

Схема классификатора состояний объекта обладает, по крайней мере, следующими общими характерными чертами [11–15]:

Значения параметров, которые фиксируются датчиками через равные промежутки времени, представляют собой входные данные. Состояние объекта оценивается на основе этих данных. Кроме того, данные снимаются в реальном времени и представляют временные ряды значений параметров:

1. Данные обрабатываются микроконтроллером системы мониторинга и управления объектом в реальном времени. В связи с возможным ограничением на выделяемые вычислительные мощности контроллера, алгоритм обработки и оценки состояния должен быть достаточно простым.

2. Процедура определения состояния объекта должна быть выполнена в рамках фиксированного промежутка времени, который определяется характеристиками функционирования объекта (например, инерционность) и особенностями решаемых им задач.

3. Схема классификации должна базироваться на анализе нескольких параметров при оценке состояния объекта, которые представлены временными рядами значений.

Однако, согласно данной схемы, оценка состояния объекта осуществляется в текущий момент времени без учета динамики изменения параметров. В связи с этим существует реальная опасность несвоевременного реагирования на быстрые изменения параметров, поскольку в схеме не реализован механизм прогнозирования изменения состояния. Схема не учитывает опасные тенденции в поведения параметров, что сужает область ее применения.

Необходимо расширение схемы в направлении реализации модели определения тенденций изменения параметров для реализации задачи прогнозирования состояния объекта. Тогда значительно повышается вероятность определения опасных изменений до их фиксации и, как следствие, значительно увеличиваются возможности системы мониторинга и управления объектом по нейтрализации нештатных ситуаций.

Постановка задачи. В работе рассматривается построение модели классификации динамики изменения параметра технического объекта, что позволит прогнозировать изменение его состояния в процессе оценки степени работоспособности объекта известными методами [2–15], которые основаны на анализе текущих данных, отражающих изменение параметров в реальном времени. Рассматриваются данные, представленные в виде временных рядов значений параметров.

Предположим, что среди группы параметров технического объекта можно выделить критичный параметр, от изменения которого в наибольшей степени изменяются другие параметры из группы, причем на основании изменения его значений можно судить об изменениях других параметров в группе и, в целом, о состоянии функционирования объекта. Следовательно, этот параметр можно рассматривать как идентифицирующий группу.

При построении модели классификации на основе временного ряда такого критичного параметра предполагалось, что шаг дискретизации при съеме данных можно установить небольшим, таким, что изменение параметра в течение нескольких последовательных шагов будет незначительным. Это позволит повысить точность классификации.

Пусть значения критичного параметра технического объекта $y(t)$ измеряются микроконтроллерным модулем в дискретные моменты времени с постоянным шагом h. В результате получим массив $Y = \{y_i\}_{i=0}^{n}$ измеренных значений параметра $y_i = y(t_i)$ в точках $t_0, t_1, ..., t_n$; $t_i = t_{i-1} + h$.

Для оценки изменения параметра объекта во времени можно воспользоваться временным окном, которое будет двигаться вдоль временного ряда, вырезая из множества исходных данных подмножество с неизменным количеством временных отсчетов. Таким образом, подмножество будет характеризоваться постоянным количеством входящих в него элементов и положением на временной оси формирующего это множество временного окна, например, начальная и конечная временные точки для текущей версии подмножества.

Определим для параметра y, представленного временным рядом, временное окно T. Для произвольного временного окна T_i получим массив $Y_{y_{T_i}} = \{y_k\}_{k=i}^{i+K}$ измеренных значений параметра в точках $y_k = y(t_k)$, где $t_k = t_i, ..., t_{i+K}$; $t_k = t_{k-1} + h$; $\{t_k\}_{k=i}^{i+K} \in T_i$; $Y_{y_{T_i}} \in Y$; $K \ll n$; $T_i = \{t_i, t_{i+1}, ..., t_{i+K}\}, i = 1, 2, ..., n$. Необходимо построить модель оценки изменения состояния объекта на основе ограниченного набора значений критичного параметра, определяемого временным окном.

Кроме того, параметр, используемый для оценки, может характеризоваться высоким уровнем шума. Фильтрация исходных данных потребует дополнительного времени, что негативным образом может повлиять на скорость и своевременность принятия решений по защите от нештатных ситуаций, особенно в условиях ограниченных вычислительных

ресурсов микроконтроллера системы мониторинга и управления объектом [16–18]. Поэтому стоит рассмотреть модель, работающую с «сырыми» данными без предварительной обработки.

Модель классификации степени изменения параметра в реальном времени на основе анализа облака точек временного ряда. Для оценки динамики изменения критичного параметра объекта можно использовать временной ряд без предварительной обработки в виде совокупности значений, выбранных из временного окна и представленных особым способом. Для этого предлагается использовать представление точек временного окна в виде графика Пуанкаре, который фактически является особым типом графика повторения или видом диаграммы рассеяния [19–24]. График Пуанкаре представляет собой точечное графическое отображение N значений временной последовательности x_k при k = 1, …, N на двумерном поле, в котором ординатой точки является значение x_{k+1}, а абсциссой – предшествующее значение x_k.

Определим для критичного параметра *y* соотношение, определяющее степень изменения параметра в рамках временного окна T_i:

$$s_i = F\left(Y_{y_{T_i}}, R_y, S_y, T_i\right), \tag{1}$$

где $Y_{y_{T_i}}$ – массив значений критичного параметра *y* во временном окне T_i, далее по тексту Y_i;

R_y – критерий оценки динамики изменения параметра *y*;

S_y – вектор классов, характеризующий динамики изменения параметра *y*;

s_i – значение класса для временного окна T_i, $s_i \in S_y$

В соотвествии с (1) значение класса, определяющего динамику изменения критичного параметра во временном окне T_i зависит от величины временного окна и массива значений критичного параметра *y*, фиксированных в этом окне. Для оценки динамики изменения параметра предварительно необходимо определить элементы вектора классов S_y и критерий R_y оценки динамики изменения параметра *y*.

В качестве временного окна T_i выбирается минимальный промежуток времени, в течение которого для данного типа объекта характерен переход от одного среднего уровня снимаемого параметра к другому. Временное окно будет различным для каждого параметра технического объекта и зависит от требований решаемой задачи.

Элементы вектора классов S_y зависят от особенностей влияния критичного параметра на состояние объекта. Формирование классов также зависит от функционального назначения объекта. Два класса, например, «работоспособен» и «не работоспособен» или «стабильное движение» и «нестабильное движение», являются минимально необходимым количеством классов вектора S_y. Возможно расширение количества классов исходя из градации уровней выполнения объектом своих функций.

Облако точек, формируемое при построении диаграммы рассеяния, для временного ряда технического параметра, как правило, имеет форму эллипса [25]. В зависимости от динамики изменения параметра форма эллипса меняется. Поэтому предлагается в качестве критерий R_y использовать характеристики эллипса, связанные с его геометрической конфигурацией. В связи с тем, что скорость изменения параметра представляется на диаграмме рассеяния в виде сжатия или растяжения геометрической области облака точек диаграммы, в качестве критерия целесообразно установить коэффициент сжатия эллипса (эллиптичность) $k = \frac{b}{a}$, где *a* и *b* – большая и малая полуоси эллипса [25].

Используя диаграммы рассеяния, построим процедуру отслеживания динамики изменения параметра.

Для критичного параметра:

Выбираем участки временного ряда параметра, которые характеризуют каждое из заданных состояний параметра или классов $s_1, s_2, \ldots, s_m, \ldots, s_L$, где L – количество классов. в результате получим массив участков временного ряда, которые соответствует классу состояния s_1, массив, соответствующий классу s_2, и т.д.

Строим диаграмму рассеяния для каждого участка ряда, соответствующего классу s_1. В результате получим массив диаграмм рассеяния для класса s_1. Ту же процедуру реализуем для каждого класса s_m.

На точках каждой диаграммы рассеяния строим эллипс следующим образом:

Используя метод наименьших квадратов, стоим линейную аппроксимацию (тренд) точек диаграммы рассеяния в виде $y_{i+1} = d_1 \cdot y_i + d_0$. Таким образом, определяем направления одной из осей эллипса (большая ось).

Определяем центр масс точек диаграммы рассеяния по формуле:

$$\bar{G}_{Y_i} = \frac{\sum_{k=i}^{i+K-1} y_k r_k}{\sum_{k=i}^{i+K-1} r_k}; \quad \bar{G}_{Y_i+1} = \frac{\sum_{k=i+1}^{i+K} y_k r_k}{\sum_{k=i+1}^{i+K} r_k}, \tag{2}$$

где $\bar{G}_{Y_i}, \bar{G}_{Y_i+1}$ – координаты центра масс точек на области определения диаграммы рассеяния;

y_k, r_k – значение j-го элемента временного ряда значений сигнала и его вес.

Если центр масс точек диаграммы рассеяния, вычисленный в соответствии с (2), не лежит на тренде, перемещаем тренд параллельно исходному расположения так, чтобы точка центра масс находилась на нем.

Строим перпендикулярную линию относительно тренда, содержащую точку центра масс. Таким образом, определяем направления другой оси эллипса (малая ось).

Принимаем решение формировать эллипс на точках диаграммы рассеяния так, чтобы эллипс включал в себя не менее P всех точек диаграммы рассеяния, например, $P = 0{,}95$.

Для этого перенесем центр координат в точку $D\left(\bar{G}_{Y_i}, \bar{G}_{Y_i+1}\right)$. Ось абсцисс направим по направлению большой оси эллипса, ось ординат – по направлению малой оси эллипса. Согласно результатам моделирования [25, 26], для задач анализа параметров технических объектов, облако точек, принадлежащих временному окну, как правило, располагается в первом квадранте декартовой системы координат. Новая система координат повернута относительно базовой системы координат на угол поворота φ против часовой стрелки. Угол поворота $\varphi < \pi/2$.

Тогда новая система координат $0' y'_i y'_{i+1}$ (рис. 1) построена с помощью параллельного переноса базовой системы координат $0 y_i y_{i+1}$ в соответствии с соотношениями:

$$y_i = d_1 + y'_i; y_{i+1} = d_0 + y'_{i+1}, \tag{3}$$

а также ее последующего поворота на угол φ:

$$y_i = y'_i \cdot \cos\varphi - y'_{i+1} \sin\varphi; \; y_{i+1} = y'_i \sin\varphi + y'_{i+1} \cos\varphi, \tag{4}$$

$$\cos\varphi = \sqrt{\frac{1}{1+d_1^2}}, \sin\varphi = a \cdot \sqrt{\frac{1}{1+d_1^2}}.$$

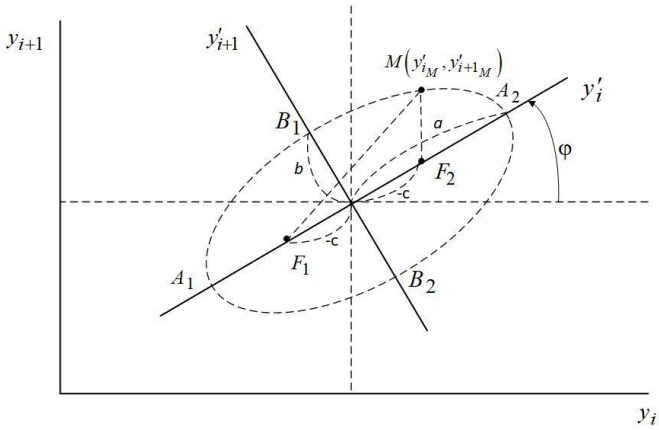


Рис. 1. Построение системы координат $0'y'_i y'_{i+1}$

Выразив из уравнений (3) и (4) y'_i и y'_{i+1}, получим соотношения, выражающие координаты $\left(y'_i, y'_{i+1}\right)$ в системе координат $0'y'_i y'_{i+1}$ относительно системы координат $0y_i y_{i+1}$:

$$y'_i = \left(y_i - d_1\right)\cdot\cos\varphi + \left(y_{i+1} - d_0\right)\cdot\sin\varphi, \tag{5}$$

$$y'_{i+1} = -\left(y_i - d_1\right)\cdot\sin\varphi + \left(y_{i+1} - d_0\right)\cdot\cos\varphi. \tag{6}$$

Соотношения (5) и (6) дают возможность определить координаты точек временного окна в новой системе координат для построения эллипса и оценки состояния параметра технического объекта.

Процесс построения эллипса итерационный и заключается в подборе большой и малой полуосей *a* и *b* соответственно таким образом, чтобы доля *P* точек из текущего участка T_i временного ряда параметра находилась внутри эллипса. Эллипсом называется множество всех точек плоскости, сумма расстояний от каждой из которых до двух данных точек этой плоскости, называемых фокусами, есть величина постоянная, большая чем расстояние *c* между фокусами []. Расстояние c между фокусами и точкой начала координат в системе координат $0'y'_i y'_{i+1}$ равно $c = \pm\sqrt{a^2 - b^2}$.

Согласно определению, для произвольной точки $M\left(y'_i, y'_{i+1}\right)$, лежащей на кривой, определяющей контур эллипса с полуосями *a* и *b* (*a* > *b*) и фокальным расстоянием c, определяется выражением $\frac{\left(y'_i\right)^2}{a^2} + \frac{\left(y'_{i+1}\right)^2}{b^2} = 1$.

Тогда точка $M\left(y'_i, y'_{i+1}\right)$ лежит внутри контура, ограниченного эллипсом, если $\frac{\left(y'_i\right)^2}{a^2} + \frac{\left(y'_{i+1}\right)^2}{b^2} \le 1$, и вне эллипса, если $\frac{\left(y'_i\right)^2}{a^2} + \frac{\left(y'_{i+1}\right)^2}{b^2} > 1$.

Задав начальные значения *a* и *b*, исходя из расположения крайних точек по осям $0'y'_i$ и $0'y'_{i+1}$ на диаграмме рассеяния, и используя вышеприведенные соотношения, можно построить эллипс, содержащий заданную долю *P* точек.

На основе полученных значений полуосей *a* и *b* можно вычислить коэффициент сжатия эллипса $k = \frac{b}{a}$, который является в нашем случае критерием для оценки скорости изменения параметра.

Таким образом, процедура обучения системы классификации скорости изменения параметра, а фактически, определения значений критериев, соответствующих классам s_1, s_2, …, s_m, …, s_L, схематично выглядит следующим образом:

1. Выбираем участок временного ряда T_i изменения параметра, соответствующий классу s_m .
2. Формируем диаграмму рассеяния в системе координат $0y_iy_{i+1}$.
3. На основе обработки данных диаграммы рассеяния строим новую систему координат $0'y_i'y_{i+1}'$.
4. В новой системе координат на основе облака точек, принадлежащих выбранному участку временного ряда, строим эллипс, который включает в себя не менее P всех точек.
5. Коэффициент сжатия эллипса $k=\frac{b}{a}$ дает частное значение критерия R_{y_m} для того, чтобы отнести изменение параметра на участке T_i к классу s_m.
6. После обработки всего массива участков $\{T_i\}$ обучающей выборки, состоящей из полученных в результате экспериментов временных рядов, получим для каждого класса $s_m \in \{s_j\}_{j=1}^{j=L}$ значение критерия R_{y_m} в виде диапазона чисел, сформированных на основе его частных значений, полученных на стадии обучения.

Процедура классификации скорости изменения параметра объекта частично включает этапы обучения:

1. Фиксируем участок временного ряда T_i изменения параметра.
2. Формируем диаграмму рассеяния в системе координат $0y_iy_{i+1}$.
3. Строим новую систему координат $0'y_i'y_{i+1}'$.
4. В новой системе координат на основе облака точек, принадлежащих выбранному участку временного ряда, строим эллипс, который включает в себя не менее P всех точек.
5. По вычисленному коэффициенту сжатия эллипса $k=\frac{b}{a}$ определяем класс, к которому относится изменение параметра на участке T_i.

Моделирование процесса классификации степени изменения параметра в реальном времени. Для апробации модели в качестве параметра технического объекта будем рассматривать ускорение автомобиля по оси Y [25, 26]. Это поперечная составляющая ускорения, перпендикулярная по отношению к направлению движения автомобиля. Результаты экспериментов представлены в виде временных рядов.

Обработка обучающей выборки временных рядов позволила определить ориентировочные значения критерия для классификации скорости изменения параметра автомобиля в процессе движения. Следует отметить, что класс параметра непосредственно связан с состоянием объекта.

Всего выделено три класса и определены значения критерия R_{y_m}:

Первый класс s_1: $R_{y_1}=0,42\le k\le 0,81$. Характеризует состояние устойчивого движения - движение с постоянным ускорением.

Второй класс s_2: $R_{y_2}=0,33\le k\le 0,41$. Характеризует движение с незначительно меняющимся ускорением, когда имеются предпосылки для нестабильного движения.

Третий класс s_3: $R_{y_3}=0,11\le k\le 0,32$. Характеризует состояние нестабильного движения – движение с непостоянным ускорением.

Для тестирования модели использовались временные ряды параметра, не входящие в обучающую выборку.

В качестве примера рассмотрим временной ряд изменение ускорения по оси Y при повороте автомобиля. Представленный на рис. 2 временной ряд не входил в обучающую выборку временных массив.

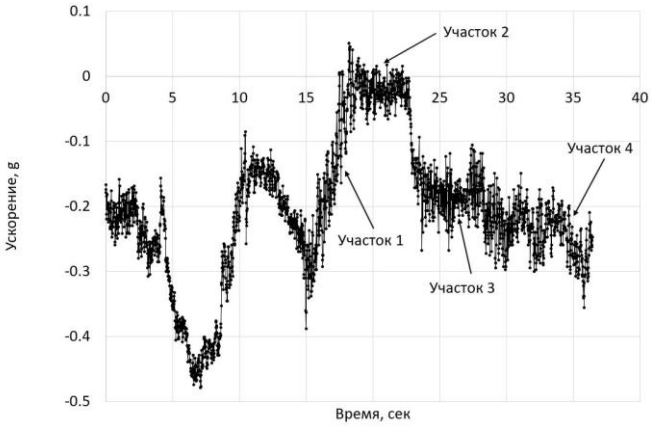


Рис. 2. Ускорение по оси Y

Для тестирования работы модели из временного ряда (рис. 2) были выбраны четыре фрагмента. Для каждого из них была сформирована диаграмма рассеяния в системе координат $0y_iy_{i+1}$. Затем в новой системе координат $0'y'_iy'_{i+1}$ на основе облака точек, принадлежащих выбранному участку временного ряда, построен эллипс, который включает в себя не менее P= 95% всех точек (рис. 3).

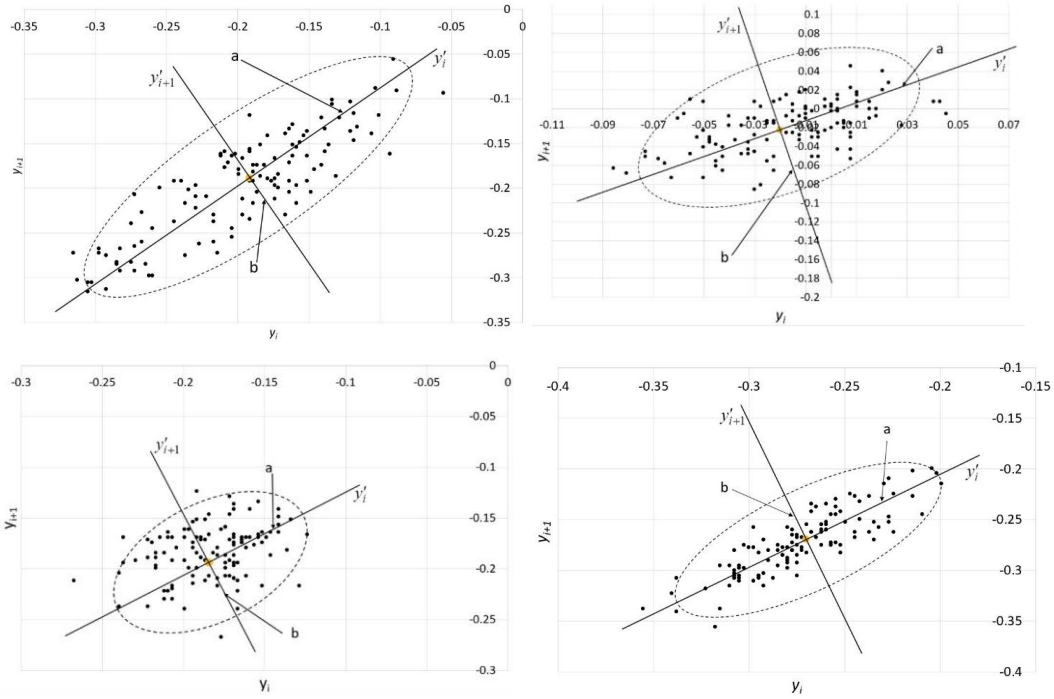


Рис. 3. Результаты обработки данных для определения класса изменения контролируемого параметра объекта на участках 1 (а), 2 (б), 3 (в) и 4 (г)

По вычисленному коэффициенту сжатия эллипса $k = \frac{b}{a}$ определен класс, к которому относится изменение параметра на каждом из четырех участков:

♦ Участок 1: класс s_3: $k = 0,299 \in R_{y_3}$, состояние нестабильного движения - движение с непостоянным ускорением.

♦ Участок 2: класс s_1: $k = 0,455 \in R_{y_1}$, состояние устойчивого движения - движение с постоянным ускорением.

♦ Участок 3: класс s_1: $k = 0,550 \in R_{y_1}$, состояние устойчивого движения - движение с постоянным ускорением.

♦ Участок 4: класс s_2: $k = 0,341 \in R_{y_2}$, движение с незначительно меняющимся ускорением. Имеются предпосылки для нестабильного движения.

Экспертная оценка степени изменения параметра объекта при движении соответствует полученной в результате процедуры классификации.

Заключение. Таким образом, разработана модель классификации динамики изменения критичного параметра, который определяет состояние работоспособности технического объекта. Модель базируется на анализе поведения временных рядов параметра в процессе функционирования объекта. Для оценки динамики изменения критичного параметра объекта используется временной ряд без предварительной обработки в виде совокупности значений, выбранных из временного окна и представленных в виде диаграммы рассеяния.

Критерием оценки является коэффициент сжатия эллипса (эллиптичность), который строится на облаке точек диаграммы рассеяния. Определение классов, характеризующих степень изменения параметра и, следовательно, состояние объекта, зависит от рассматриваемого критичного параметра и, конечно, от объекта.

Модель позволит обеспечить реализацию процедур выявления в реальном времени возможности возникновения нештатной ситуации на ранней стадии ее развития с помощью микропроцессорного модуля, расположенного на нижнем уровне системы мониторинга объекта.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Буков В.Н.* Адаптивные прогнозирующие системы управления полетом. – М.: Наука, Гл. ред. физ.-мат. лит., 1981. – 232 с.
2. *Клевцова А.Б., Клевцов Г.С.* Модели параметрической экспресс-оценки состояния технического объекта // Известия ЮФУ. Технические науки. – 2008. – № 11 (88). – С. 15-19.
3. *Васильев В.В., Грездов Г.И., Симак Л.А. и др.* Моделирование динамических систем: Аспекты мониторинга и обработки сигналов / под ред. В.В. Васильева. – К.: НАН Украины, 2002. – 344 с.
4. *Steblev Yu.I., Susarev S.V., Bykov D.E.* The principles of designing automated systems for diagnostic monitoring of the engineering structures of hazardous production objects // Russian Journal of Nondestructive Testing. – April 2015. – Vol. 51, Issue 4. – P. 185-197.
5. *Detlev W. Gross*. Partial Discharge Measurement and Monitoring on Rotating Machines // IEEE Int. Sym. On Elect. Insul, Boston MAUSA, April 7-10, 2002. – P. 33-41.
6. *Vovk S.P., Ginis L.A.* Modelling and forecasting of transitions between levels of hierarchies in Difficult formalized systems // European Researcher. – 2012. – Vol. (20), No. 5-1. – P. 541-545.
7. *Клевцов С.И., Клевцова А.Б., Буринов С.В.* Модель параметрической качественной иерархической оценки состояния технической системы // Инженерный вестник Дона. – 2015. – № 3. – URL: ivdon.ru/ru/magazine/archive/n3y2015/3088.
8. *Matuszewski J.* Application of clustering methods for recognition of technical objects // Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET) – 2010 International Conference. – 2010. – P. 39-40.
9. *Lihua Sun, Yingjun Guo, Haichao Ran.* A New Method of Early Real-Time Fault Diagnosis for Technical Process // Electrical and Control Engineering (ICECE), 2010 International Conference, Wuhan, China, 2010. – P. 4912-4915.

10. *Васильев В.В.* Современные проблемы компьютерного мониторинга в энергетике // Известия ТРТУ. – 2001. – № 3. – С. 99-120.
11. *Клевцова А.Б.* Интегральная оценка состояния объекта мониторинга // Известия ТРТУ. – 2004. – № 2 (37). – С. 58-65.
12. *Клевцов С.И.* Предварительная оценка состояния совокупности параметров технического объекта с использованием интеллектуального микропроцессорного модуля // Известия ЮФУ. Технические науки. – 2010. – № 5 (106). – С. 43-48.
13. *Клевцов С.И.* Прогнозирование измерения состояния параметров технического объекта с помощью интеллектуального микропроцессорного модуля // Проблемы разработки перспективных микро- и наноэлектронных систем – 2010: Сб. трудов. – М.: ИППМ РАН, 2010. – С. 619-622.
14. *Krivosheev I.A., Rozhkov K.E., Simonov N.B.* Complex Diagnostic Index for Technical Condition Assessment for GTE // International Conference on Industrial Engineering, ICIE, 2017. – 2017. – 206. – P. 176-181.
15. *Jerzy Hoja, Grzegorz Lentka.* A family of new generation miniaturized impedance an analyzers for technical object diagnostics // Metrology and Measurement Systems. – 2013. – Vol. XX, No. 1.
16. *Пьявченко О.Н., Горелова Г.В., Боженюк А.В., Клевцов С.И., Клевцова А.Б.* Методы и алгоритмы моделирования развития сложных ситуаций. – Таганрог: Изд-во ТРТУ, 2003. – 157 с.
17. *Stanisław Duer.* Diagnostic system with an artificial neural network in diagnostics of an analogue technical object // Neural Computing and Applications. – February 2010. – Vol. 19, Issue 1. – P. 55-60.
18. *Klevtsov Sergei I.* Identification of the state of technical objects based on analyzing a limited set of parameters // 2016 International Siberian Conference on Control and Communications (SIBCON): Proceedings. National Research University Higher School of Economics. Russia, Moscow, May 12-14, 2016. – URL: http://ieeexplore.ieee.org/document/7491752/.
19. *Новоселов О.Н.* Идентификация и анализ динамических систем: монография. – 3-е изд., перераб.и доп. – М.: ГОУ ВПО МГУЛ, 2010. – 424 с.
20. *Гуфельд И.Л., Гаврилов В.А., Корольков А.В., Новоселов О.Н.* Эндогенная активность Земли и декомпрессионная модель сейсмического шума // Доклады РАН. – 2008. – Т. 423, № 6. – С. 811-814.
21. *Орлов В.Н.* Руководство по электрокардиографии. – М.: Медицина, 1984. – 526 с. – DOI: 10.1109/BMEiCon.2013.6687679.
22. *Kannakorn Intharakham, Kesorn Suwanprasert.* Complexity of Autonomic control during Cerebrovascular Reactivity // Proceedings of the 6th Biomedical Engineering International Conference (BMEiCON2013), October 2013.
23. *Carmen González , Erik W. Jensen, Pedro L. Gambús, Montserrat Vallverdú.* Poincaré plot analysis of cerebral blood flow signals: Feature extraction and classification methods for apnea detection. Published PLoS ONE: December 7, 2018. – P. 43-52. – https://doi.org/10.1371/journal.pone.0208642.
24. *Jan Monieta*. Selection of Diagnostic Symptoms and Injection Subsystems of Marine Reciprocating Internal Combustion Engines // Appl. Sci. – 2019. – 9 (8). – 1540. – https://doi.org/10.3390/app9081540.
25. *Клевцов С.И.* Определение характера изменений параметра на основе анализа динамики формы совокупности его значений в реальном времени // Известия ЮФУ. Технические науки. – 2020. – № 3. – С. 46-55.
26. *Клевцов С.И.* Определение момента скачкообразного изменения быстропеременной физической величины в реальном времени с использованием диаграмм Пуанкаре // Известия ЮФУ. Технические науки. – 2012. – № 5. – С. 108-113.

REFERENCES

1. *Bukov V.N.* Adaptivnye prognoziruyushchie sistemy upravleniya poletom [Adaptive predictive flight control systems]. Moscow: Nauka, Gl. red. fiz.-mat. lit., 1981, 232 p.
2. *Klevtsova A.B., Klevtsov G.S.* Modeli parametricheskoy ekspress-otsenki sostoyaniya tekhnicheskogo ob"ekta [Models of parametric express-assessment of the state of a technical object], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2008, No. 11 (88), pp. 15-19.
3. *Vasil'ev V.V., Grezdov G.I., Simak L.A. i dr.* Modelirovanie dinamicheskikh sistem: Aspekty monitoringa i obrabotki signalov [Modeling of dynamic systems: Aspects of monitoring and signal processing], ed. by V.V. Vasil'eva. Kiev: NAN Ukrainy, 2002, 344 p.
4. *Steblev Yu.I., Susarev S.V., Bykov D.E.* The principles of designing automated systems for diagnostic monitoring of the engineering structures of hazardous production objects, *Russian Journal of Nondestructive Testing,* April 2015, Vol. 51, Issue 4, pp. 185-197.
5. *Detlev W. Gross*. Partial Discharge Measurement and Monitoring on Rotating Machines, *IEEE Int. Sym. On Elect. Insul, Boston MAUSA, April 7-10, 2002,* pp. 33-41.

6. *Vovk S.P., Ginis L.A.* Modelling and forecasting of transitions between levels of hierarchies in Difficult formalized systems, *European Researcher,* 2012, Vol. (20), No. 5-1, pp. 541-545.
7. *Klevtsov S.I., Klevtsova A.B., Burinov S.V.* Model' parametricheskoy kachestvennoy ierarkhicheskoy otsenki sostoyaniya tekhnicheskoy sistemy [Model of parametric qualitative hierarchical assessment of the state of a technical system], Inzhenernyy vestnik Dona [Engineering Bulletin of the Don], 2015, No. 3. Available at: ivdon.ru/ru/magazine/archive/n3y2015/3088.
8. *Matuszewski J.* Application of clustering methods for recognition of technical objects, *Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET) – 2010 International Conference,* 2010, pp. 39-40.
9. *Lihua Sun, Yingjun Guo, Haichao Ran.* A New Method of Early Real-Time Fault Diagnosis for Technical Process, *Electrical and Control Engineering (ICECE), 2010 International Conference, Wuhan, China, 2010,* pp. 4912-4915.
10. *Vasil'ev V.V.* Sovremennye problemy komp'yuternogo monitoringa v energetike [Modern problems of computer monitoring in power engineering], *Izvestiya TRTU* [Izvestiya TSURE], 2001, No. 3, pp. 99-120.
11. *Klevtsova A.B.* Integral'naya otsenka sostoyaniya ob"ekta monitoringa [Integral assessment of the state of the monitoring object], *Izvestiya TRTU* [Izvestiya TSURE], 2004, No. 2 (37), pp. 58-65.
12. *Klevtsov S.I.* Predvaritel'naya otsenka sostoyaniya sovokupnosti parametrov tekhnicheskogo ob"ekta s ispol'zovaniem intellektual'nogo mikroprotsessornogo modulya [Preliminary assessment of the state of a set of parameters of a technical object using an intelligent microprocessor module], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2010, No. 5 (106), pp. 43-48.
13. *Klevtsov S.I.* Prognozirovanie izmereniya sostoyaniya parametrov tekhnicheskogo ob"ekta s pomoshch'yu intellektual'nogo mikroprotsessornogo modulya [Forecasting the measurement of the state of the parameters of a technical object using an intelligent microprocessor module], *Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh sistem – 2010: Sb. trudov* [Problems of development of promising micro- and nanoelectronic systems - 2010: Collection of works]. Moscow: IPPM RAN, 2010, pp. 619-622.
14. *Krivosheev I.A., Rozhkov K.E., Simonov N.B.* Complex Diagnostic Index for Technical Condition Assessment for GTE, *International Conference on Industrial Engineering, ICIE, 2017,* 2017, 206, pp. 176-181.
15. *Jerzy Hoja, Grzegorz Lentka.* A family of new generation miniaturized impedance an analyzers for technical object diagnostics, *Metrology and Measurement Systems,* 2013, Vol. XX, No. 1.
16. *P'yavchenko O.N., Gorelova G.V., Bozhenyuk A.V., Klevtsov S.I., Klevtsova A.B.* Metody i algoritmy modelirovaniya razvitiya slozhnykh situatsiy [Methods and algorithms for modeling the development of complex situations]. Taganrog: Izd-vo TRTU, 2003, 157 p.
17. *Stanisław Duer.* Diagnostic system with an artificial neural network in diagnostics of an analogue technical object, *Neural Computing and Applications*. February 2010, Vol. 19, Issue 1, pp. 55-60.
18. *Klevtsov Sergei I.* Identification of the state of technical objects based on analyzing a limited set of parameters, *2016 International Siberian Conference on Control and Communications (SIBCON): Proceedings. National Research University Higher School of Economics. Russia, Moscow, May 12-14, 2016*. Available at: http://ieeexplore.ieee.org/document/7491752/.
19. *Novoselov O.N.* Identifikatsiya i analiz dinamicheskikh sistem: monografiya [Identification and analysis of dynamic systems: monograph]. 3rd ed. Moscow: GOU VPO MGUL, 2010, 424 p.
20. *Gufel'd I.L., Gavrilov V.A., Korol'kov A.V., Novoselov O.N.* Endogennaya aktivnost' Zemli i dekompressionnaya model' seysmicheskogo shuma [Endogenous activity of the Earth and decompression model of seismic noise], *Doklady RAN* [Reports of the Russian Academy of Sciences], 2008, Vol. 423, No. 6, pp. 811-814.
21. *Orlov V.N.* Rukovodstvo po elektrokardiografii [Handbook of Electrocardiography]. Moscow: Meditsina, 1984, 526 p. DOI: 10.1109/BMEiCon.2013.6687679.
22. *Kannakorn Intharakham, Kesorn Suwanprasert.* Complexity of Autonomic control during Cerebrovascular Reactivity, *Proceedings of the 6th Biomedical Engineering International Conference (BMEiCON2013), October 2013*.
23. *Carmen González , Erik W. Jensen, Pedro L. Gambús, Montserrat Vallverdú.* Poincaré plot analysis of cerebral blood flow signals: Feature extraction and classification methods for apnea detection. Published PLoS ONE: December 7, 2018, pp. 43-52. Available at: https://doi.org/10.1371/journal.pone.0208642.
24. *Jan Monieta*. Selection of Diagnostic Symptoms and Injection Subsystems of Marine Reciprocating Internal Combustion Engines, *Appl. Sci*., 2019, 9 (8), 1540. Available at: https://doi.org/10.3390/app9081540.

25. *Klevtsov S.I.* Opredelenie kharaktera izmeneniy parametra na osnove analiza dinamiki formy sovokupnosti ego znacheniy v real'nom vremeni [Determining the nature of parameter changes based on the analysis of the dynamics of the shape of a set of its values in real time], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences]. – 2020. – № 3. – S. 46-55.
26. *Klevtsov S.I.* Opredelenie momenta skachkoobraznogo izmeneniya bystroperemennoy fizicheskoy velichiny v real'nom vremeni s ispol'zovaniem diagramm Puankare [Determination of the moment of a sudden change in a rapidly varying physical quantity in real time using Poincare diagrams], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 5, pp. 108-113.

Статью рекомендовал к опубликованию д.т.н., профессор Ю.А. Кравченко.

Клевцов Сергей Иванович – Южный федеральный университет; e-mail: sergkmps@mail.ru; г. Таганрог, Россия; тел.. 88634328025; к.т.н.; доцент.

Klevtsov Sergey Ivanovich – Southern Federal University; e-mail: sergkmps@mail.ru; Taganrog, Russia; phone: +78634328025; cand. of eng. sc.; associate professor.

УДК 004.021 DOI 10.18522/2311-3103-2024-4-91-100

А.А. Егорчев, Д.М. Пашин, Н.А. Сарамбаев, А.Ф. Фахрутдинов

АНАЛИЗ СИСТЕМ ОПРЕДЕЛЕНИЯ И КЛАССИФИКАЦИИ ЭМОЦИЙ ЧЕЛОВЕКА ПО ДАННЫМ ЗВУКОВОГО ПОТОКА

В современной быстро меняющейся и требовательной рабочей среде способность быстро и точно оценить эмоциональное состояние сотрудника имеет решающее значение для защиты человеческих жизней и снижения материальных рисков. Эмоциональное благополучие играет важную роль в обеспечении безопасности на рабочем месте, производительности труда и общего психического здоровья. Поэтому разработка эффективных инструментов для мониторинга негативных эмоций и реагирования на них является актуальной задачей современности. Целью данного исследования является разработка алгоритма, способного классифицировать эмоции, используя аудиоданные, записанные смартфоном пользователя. Такой инструмент особенно полезен, если интегрирован в более широкую систему мониторинга здоровья, позволяющую оценивать показатели здоровья человека в режиме реального времени с помощью неинвазивных методов. В этой статье представлено новое решение, которое использует акустические сигналы, улавливаемые микрофоном смартфона, для обнаружения и классификации эмоций пользователя. Используя сверточные нейронные сети (CNNS), тип алгоритма глубокого обучения, известного своей эффективностью при обработке аудио- и визуальных данных, предлагаемая система может определять эмоциональное состояние пользователя. Модель CNN обучена распознавать признаки в аудиоданных, соответствующие различным эмоциональным проявлениям, фокусируясь на обнаружении негативных эмоций, таких как, гнев или печаль. Результаты исследования демонстрируют эффективность системы: частота ошибок при определении негативных эмоций составляет 19,5% для ложноположительных результатов (ошибки I рода) и 20,1% для ложноотрицательных результатов (ошибки II рода). Эти показатели указывают на ее потенциал для практического применения в реальных условиях. Внедряя это решение в существующие системы биомедицинского мониторинга, организации могут расширить свои возможности по мониторингу эмоционального благополучия сотрудников, потенциально предотвращая негативные последствия, такие как несчастные случаи на производстве или кризисы психического здоровья. Интеграция распознавания эмоций с помощью смартфонов в системы мониторинга состояния здоровья представляет собой значительный прогресс в области неинвазивного биомедицинского мониторинга, использующего повсеместное присутствие смартфонов и возможности машинного обучения.

Система неинвазивного мониторинга; машинное обучение; биомедицинский мониторинг; смартфонная сенсорика; анализ акустического сигнала; распознавание эмоций.

A.A. Egorchev, D.M. Pashin, N.A. Sarambaev, A.F. Fakhrutdinov

EMOTION DETECTION AND CLASSIFICATION SYSTEM BASED ON SOUND FLOW DATA

In today's rapidly changing and demanding work environment, the ability to quickly and accurately assess an employee's emotional state is crucial to protecting human lives and reducing material risks. Emotional well-being plays an important role in workplace safety, productivity, and overall mental health. Therefore, the development of effective tools for monitoring negative emotions and responding to them is an urgent task of our time. The purpose of this study is to develop an algorithm capable of classifying emotions using audio data recorded by a user's smartphone. Such a tool is especially useful if integrated into a broader health monitoring system that allows you to evaluate human health indicators in real time using non-invasive methods. This article presents a new solution that uses acoustic signals picked up by a smartphone microphone to detect and classify user emotions. Using convolutional neural networks (CNNS), a type of deep learning algorithm known for its effectiveness in processing audio and visual data, the proposed system can determine the user's emotional state. The CNN model is trained to recognize patterns in audio data corresponding to various emotional manifestations, focusing on detecting negative emotions such as anger or sadness. The results of the study demonstrate the effectiveness of the system: the error rate in determining negative emotions is 19.5% for false positive results (errors of the first kind) and 20.1% for false negative results (errors of the second kind). These indicators indicate its potential for practical application in real conditions. By integrating this solution into existing biomedical monitoring systems, organizations can expand their ability to monitor the emotional well-being of employees, potentially preventing negative consequences such as industrial accidents or mental health crises. The integration of emotion recognition using smartphones into health monitoring systems represents significant progress in the field of non-invasive biomedical monitoring, using the ubiquitous presence of smartphones and machine learning capabilities.

Noninvasive monitoring system; machine learning; biomedical monitoring; smartphone sensors; acoustic signal analysis; emotion recognition.

Введение. В современном развивающемся мире имеется огромное количество рабочих должностей для разного уровня специалистов. Очевидно, что стресс оказывает негативное влияние на деятельность работника, а в конечном счете на производительность компании. В бизнесе и менеджменте эмоциональный интеллект играет важную роль в формировании лидерства и взаимоотношений с коллегами и подчиненными. В образовании эмоции позволяют создавать позитивную образовательную среду и эмоционально поддерживать учащихся. В творческих профессиях эмоции напрямую влияют на производимый продукт. Более того негативные эмоции могут быть предвестниками конфликтов. Следовательно, существует необходимость в разработке автоматизированной системы, которая позволит своевременно выявлять людей с негативными эмоциями для снижения негативного их влияния на рабочую среду, а также оказания своевременной помощи, например, для предотвращения суицидов. В данной работе предложено решение на основе анализа алгоритмов определения эмоций из категории алгоритмов искусственного интеллекта, который служит для распознавания одной из трех эмоций: положительной, нейтральной или негативной в речевом сигнале. Данное решение предназначено для применения в рамках большой системы мониторинга состояния здоровья человека на предприятиях.

Основная часть. Тема классификации эмоций является довольно популярной и имеет множество работ, авторы которых предлагают свои подходы к определению. В основном решения базируются на использовании алгоритмов машинного обучения. Машинное обучении может использоваться для распознавания эмоций в аудиоданных следующим образом:

- Построение табличной структуры данных: строится таблица данных со значениями эмоций для каждой аудиозаписи.
- Предварительная обработка: используя инструменты OpenSMILE [1], извлекаются статистические признаки из аудиосигнала

- Обучение модели: с использованием CNN и построенной таблицы, модель обучается на распознавание эмоций.
- Тестирование модели: модель тестируется на способность распознавания эмоций с использованием тестовой выборки.
- Использование модели: обученная модель может использоваться для распознавания эмоций в новых аудиосигналах.

В решении Siddhant Mulajkar [2] автор предлагает модель распознавания стресса, основанную на глубоком обучении с использованием речевых сигналов. Предложенный алгоритм сначала извлекает мел-кепстральные коэффициенты из предварительно обработанных речевых данных, а после предсказывает состояние стресса, деля данные признаки на две группы (стресс/отсутствие стресса), опираясь на данных сверточной нейронной сети (CNN). Для построения данного решения использовался только один датасет (RAVDESS), так как весь алгоритм по извлечению признаков и построению табличной структуры данных строился, опираясь на идентификаторы аудиофайлов. Следовательно, для других датасетов, с другими идентификаторами этот алгоритм не подходит. Датасет RAVDESS наиболее распространенный и часто используется другими авторами. Опираясь на идентификаторы имен файлов, датасет разделён на табличную структуру данных, в отдельных столбцах которого: путь к файлу, источник, номер голоса, пол, интенсивность, повторение, эмоция. В датасете имеются записи как мужских, так и женских голосов, а так как они различаются по тембру и высоте голоса, автором было предложено решение классифицировать эмоции только для мужских голосов. После обучения модели, на основании тестовой выборки представлена матрица ошибок по 5-ти эмоциям (рис. 1). Данное решение было протестировано и получена матрица ошибок (рис. 2), не совпадающая с матрицей, заявленной автором. От данного решения было решено отказаться, так как для него подходит только один датасет. Для использования этого алгоритма оказалось необходимым полностью переписать решение, чтобы можно было использовать признаки с других датасетов.

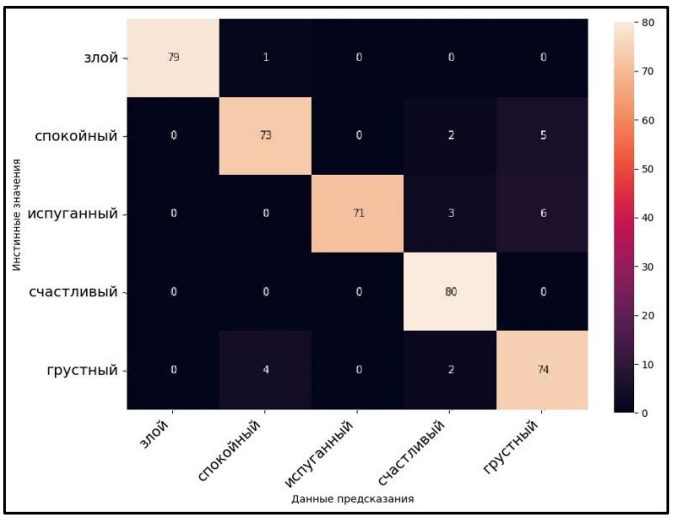


Рис. 1. Матрица ошибок для определения эмоций для лиц мужского пола для решения [2]

Решение на платформе Kaggle от Eu Jin Lok [3], состоит из пяти частей (разбиение на табличную структуру данных, вычисление признаков, построение модели, проверка на записанных данных, проверка на измененных данных). Данное решение написано на языке Python и в нём автором представлен классификатор эмоций по голосу человека. Решение было предложено с целью упростить работу колл-центров. На рис. 3 продемонстрирована матрица ошибок для данного решения.

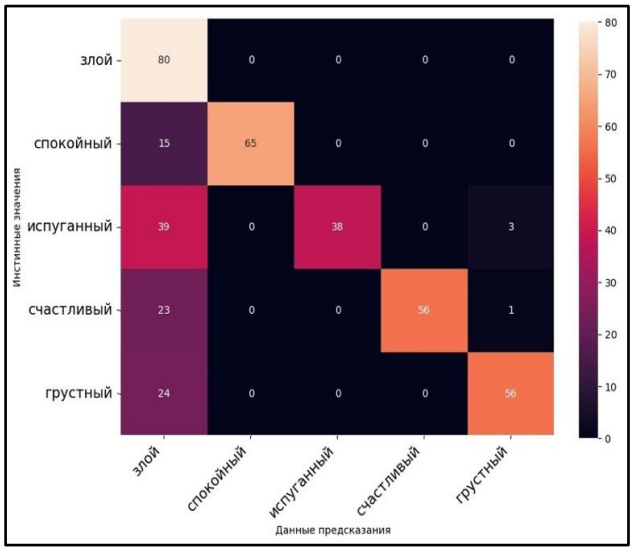


Рис. 2. Матрица ошибок для определения эмоций для лиц женского пола для решения [2]

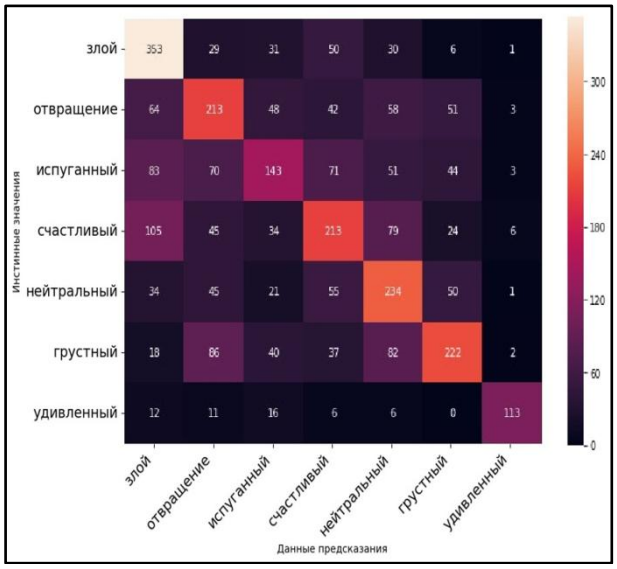


Рис. 3. Матрица ошибок для определения эмоций для лиц мужского и женского пола для решения [3]

Датасет в машинном обучении – это набор связанных между собой данных. Датасет может включать в себя текст, изображения, аудио и другие типы данных. Он может быть собран из различных источников, таких как базы данных, веб-сайты, социальные сети и т.д. Датасет может быть использован для обучения модели, а также для оценки ее качества и настройки параметров. Машинное обучение нацелено на создание систем, способных получать знания из данных, способных с помощью обучения улучшать показатели своей работы [4].

Для проверки готовых решений и обучения модели, был произведен поиск данных. Критерии, по которым были отобраны датасеты:

- Четкость аудиозаписей (без шума).
- Сбалансированность по количеству голосов (акцент, национальность).

- Объем датасета.
- Язык.

Были отобраны наиболее распространенные датасеты, которые находятся в открытом доступе:

- SAVEE [5]. Используемый датасет, составленный на базе записей четырех носителей английского языка, которые являются аспирантами и исследователями Университета Суррея в возрасте от 27 до 31 года, включает в себя семь видов эмоций: злость, отвращение, испуг, страх, счастье, печаль, удивление и нейтральное состояние.
- RAVDESS [6]. Данный датасет содержит 1440 звуковых файлов голосов 12 мужчин и 12 женщин, где для каждого человека имеется 60 записей. Файлы записаны для семи эмоциональных состояний человека: грусть, злость, радость, спокойствие, страх, удивление и отвращение. В каждой записи выражение воспроизводится на одном из двух уровнях эмоциональной интенсивности: нормальный и сильный.
- CREMA-D [7]. Этот датасет из 7442 клипов от 91-го человека. Клипы были сняты 48 актерами мужского пола и 43 актрисами женского пола, в возрасте от 20 до 73 лет. Для записи датасета люди говорили одно из двенадцати предложенных им предложений для шести различных эмоциональных состояний: гнев, радость, страх, отвращение, грусть и нейтральное, и 4 различных уровней эмоциональной интенсивности: низкий, средний, высокий и неопределенный.

Извлечение признаков в машинном обучении – это процесс преобразования исходных данных в набор признаков, который может быть использован в качестве входных данных модели машинного обучения. Процесс может включать в себя операции, такие как сжатие, нормализация или кодирование данных, а также использование различных алгоритмов для извлечения признаков, таких как преобразование Фурье или мел-кепстральное преобразование. Извлечение признаков является важной частью процесса машинного обучения, поскольку качество и количество признаков, используемых для обучения модели, может существенно влиять на точность и обобщающую способность модели.

Для этих целей нужно преобразовать их в нужный формат. MFCC (Мел-кепстральные коэффициенты) используются в области обработки звуковых сигналов, чтобы улучшить эффективность анализа и распознавания звуков. Они представляют собой признаки, которые отражают частотные характеристики звука, учитывая чувствительность человеческого уха к различным частотам. Они также позволяют уменьшить избыточность информации в звуковых сигналах и оптимизировать анализ звуковых признаков для задач распознавания эмоций. Человеческих слух воспринимает высоты звука не линейно, по отношению к частоте, и описывается следующим образом:

$$M(f) = 1125 \ln\left(1 + \frac{f}{700}\right). \quad (1)$$

График зависимости мел-шкалы от частоты представлен на рис. 4.

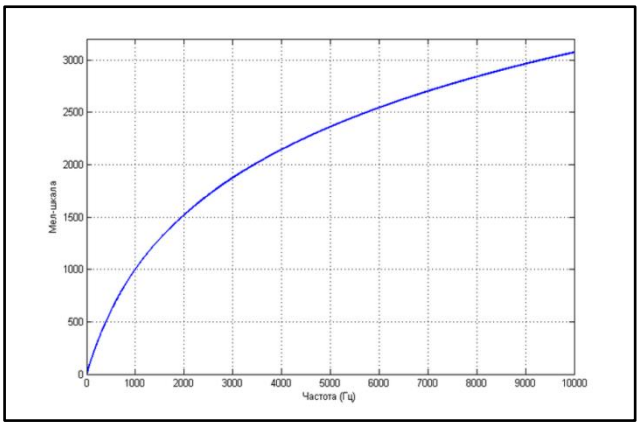


Рис. 4. График зависимости мел-шкалы от частоты

Подобные единицы измерения часто используют, так как они позволяют приблизиться к механизмам человеческого восприятия, которое пока что является лидирующей гипотезой в рамках известных систем распознавания речи [8].

Для извлечения звуковых признаков была использована библиотека openSMILE [1]. OpenSMILE (Open-Source Speech and Music Interpretation by Large-Space Extraction) является библиотекой, написанной на языке программирования C, предназначенная для извлечения признаков из аудио файлов. Используя данную библиотеку, можно извлечь такие признаки, как мел-кепстральные коэффициенты, коэффициенты энергии, аудиоспектральные коэффициенты и другие. Этот инструмент может быть использован в различных задачах аудио анализа, таких как распознавание эмоций или голосовой идентификации [9]. В рамках системы по распознаванию негативного и позитивного эмоционального состояния необходимо было разработать решение, способное классифицировать получаемые данные только на три класса, а существующие датасеты содержат от шести до восьми различных эмоций. Появилась необходимость в реорганизации найденных датасетов. Каждый из датасетов имеет свой идентификатор имени файла. В датасете SAVEE аудиофайлы названы таким образом, 'DC_d03.wav'. Первые 2 буквы префикса имени файла представляют инициалы говорящего. Буквы-префиксы описывают классы эмоций следующим образом:

- ‘a’ – anger(злость);
- ‘d’ – disgust(отвращение);
- ‘f’ – fear(страх);
- ‘h’ – happiness(радость);
- ‘n’ – neutral(нейтральное);
- ‘sa’ – sadness(грусть);
- ‘su’ – surprise(удивление).

Числа в конце названия файлов обозначают номер прочитанного предложения.

В датасете RAVDESS, каждый файл именовался следующим образом:

‘03-01-01-01-01-01-01.wav’, где каждый номер имеет следующие обозначения:

- Модальность (01 = аудио с видео, 02 = только видео, 03 = только аудио).
- Речевой канал (01 = речь, 02 = песня).
- Эмоция (01 = нейтральное, 02 = спокойный, 03 = радостный, 04 = грустный, 05 = злой, 06 = напуганный, 07 = раздражение, 08 = удивленный).
- Эмоциональная напряженность (01 = нормальное, 02 = сильное).
- Высказывание (01 = “Kids are talking by the door”, 02 = “Dogs are sitting by the door”).
- Повторение (01 = первое повторение, 02 = второе повторение).
- Актер (с 1-24, нечетные номера – мужчины, нечетные номера – женщины).

В датасете CREMA-D актеры и эмоции, как и все предыдущие датасеты, помечены в самом имени файла аудио. Пример файла из этого датасета: ‘1001_IEO_ANG_MD.wav’. Для данного датасета имеется файл, формата .csv, с идентификаторами распределения по эмоциям и актерам.

Датасеты были преобразованы в табличные структуры данных, которые были объединены для дальнейшего извлечения признаков.

Основная часть. Формула для вычисления мел-кепстральных коэффициентов состоит из нескольких шагов:

1. Преобразование Фурье для получения спектра частот.
2. Применение мел-фильтров для получения мел-спектра.
3. Вычисление логарифма мел-спектра.
4. Применение обратного преобразования Фурье для получения мел-кепстральных коэффициентов.

Мел-кепстральные коэффициенты часто используются для распознавания эмоций в аудио-сигнале, так как они могут дать информацию о частотной характеристике звука, которая может быть связана с эмоциональным состоянием человека. Например, низкие частоты могут быть связаны с тревожностью или гневом, в то время как высокие частоты

могут быть связаны с радостью или удовольствием. Использование мел-кепстральных коэффициентов позволяет извлечь эту информацию из аудио-сигнала и использовать ее для обучения модели распознавания эмоций.

CNN (Convolutional Neural Network – сверточные нейронные сети) модель для аудиосигнала обычно состоит из нескольких слоев:

- Слой входных данных: здесь аудиосигнал преобразуется в массив данных, который может быть обработан моделью.
- Слои свертки: это основа CNN модели. Они используются для извлечения признаков из входных данных.
- Слои пуллинга: это слои, которые используются для уменьшения размерности данных и уменьшения шума. Они также помогают увеличить обобщающую способность модели.
- Полносвязные слои: это слои, которые используются для классификации данных. Они используются для анализа извлеченных признаков и делают предсказание на основе этих признаков. Для распознавания эмоций по голосу часто используются нейронные сети с одним или несколькими сверточными слоями (CNN), которые способны извлекать важные признаки из аудиосигнала и передавать их в полносвязные слои для классификации.

С помощью TensorFlow Keras на языке Python с применением PyQt6 была разработана итоговая модель определения эмоций. На вход алгоритма подаются аудиоданные длительностью в 2.5 секунда и с частотой дискретизации 44100 Гц. В качестве датасетов использовались представленные ранее датасеты. Структура модели представлена на рис. 5.

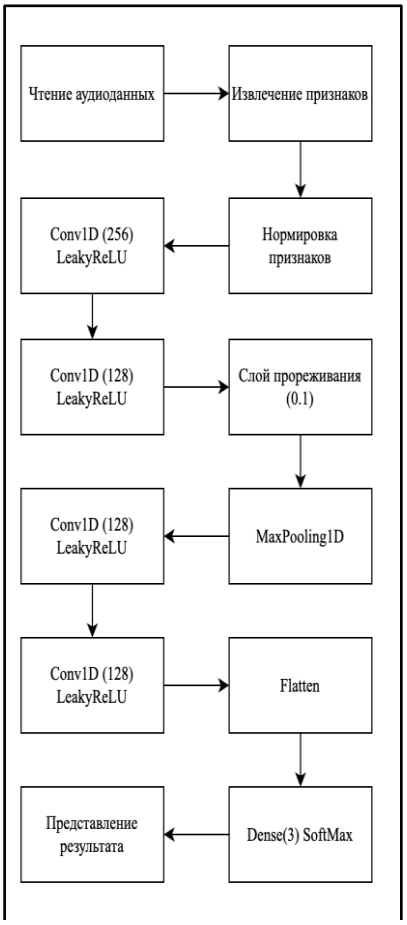


Рис. 5. Структура алгоритма определения эмоций

Результаты испытаний. Для определения качества разработанного решения проведены ряд испытаний. В качестве испытуемых приглашено 20 человек, которые являются студентами и сотрудниками Казанского Федерального Университета. Для испытаний использовалась клиент-серверная система мониторинга состояния здоровья человека. В качестве сервера выступал персональный компьютер Forsite(Intel® Core™ i9-9920X, NVIDIA Quadro RTX 6000; ОП 64 Гб; SSD 1Тб; клавиатура; мышь; монитор). Серверный персональный компьютер работал с СУБД PostgreSQL, в таблица которого выполнялись записи результатов измерений испытуемых посредством выполнения SQL запросов, получаемых от клиентского устройства. В качестве клиентских устройств использовались следующие модели смартфонов: Huawei nova 8i, Samsung A8 (2018), Xiaomi Mi 9 Lite, Redmi Note 9S, POCO X3 Pro, Samsung SM-A515F. Клиентские устройства, которыми пользовались испытуемые содержали установленное приложение. Каждый испытуемый проводил не менее 10 испытаний для проверки каждого типа классификации. Результаты испытаний представлены в табл. 1.

Таблица 1

Результаты испытания классификации эмоций.

Тип эмоции	Ошибки 1 рода	Ошибки 2 рода
Негативная	19.5 %	20.1 %
Положительная	22.4 %	31.5 %
Нейтральная	40.6 %	53.3 %

Обсуждение. В результате проведенных испытаний можно сделать вывод о том, что представленная модель позволяет практически с 20% ошибкой определять негативные эмоции. Это в принципе является более приоритетным классом для определения в рамках рассматриваемого применения решения. Однако является непригодной для определения других типов эмоций. Основными причинами таких результатов может являться небогатая обучающая выборка. Потенциально можно улучшить результаты, расширив количество рассматриваемых признаков, проведя балансировку обучающей выборки, а также увеличив размер ее размер.

Выводы. Рассмотрев тему классификации эмоций, разработан алгоритм, которая решает данную задачу используя алгоритмы машинного обучения, которую можно интегрировать в систему биомедицинского мониторинга. Испытания показали применимость полученной модели и алгоритма для определения негативных эмоций 19.5% ошибки первого рода и 20.1% ошибки второго рода. На тестовых записях положительных и нейтральных эмоций система показала большую ошибку, что делает решение непригодным для определения именно этих типов эмоций. Решение имеет потенциал для улучшения точности распознавания, путем увеличения размера обучающей выборки и расширения наборов признаков.

Благодарность. Данная работа выполнена в Казанском Федеральном Университете в рамках программы “Приоритет-2030”.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Opensmile 2.4.2 // PyPI - Указатель пакетов Python'a. – URL: https://pypi.org/project/opensmile/ (дата обращения: 12.12.2022).
2. Sentiment-Predictor-for-Stress-Detection-using-Voice // GitHub. – URL: https://github.com/sidmulajkar/sentiment-predictor-for-stress-detection (дата обращения: 10.10.2022).
3. Audio Emotion | Part 1 - Explore data // Kaggle. – URL: https://www.kaggle.com/code/ejlok1/audio-emotion-part-1-explore-data (дата обращения: 10.10.2022).
4. *Воронина В.В., Михеев А.В., Ярушкина Н.Г., Святов К.В.* Теория и практика машинного обучения: учеб. пособие. – Ульяновск: УлГТУ, 2017. – 290 с.
5. Surrey Audio-Visual Expressed Emotion (SAVEE) Database // The Centre for Vision, Speech and Signal Processing. – URL: http://kahlan.eps.surrey.ac.uk/savee/ (дата обращения: 29.11.2022).

6. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English // PLOS. – URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0196391 (дата обращения: 28.11.2022).
7. CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset // The National Center for Biotechnology Information. – URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4313618/ (дата обращения: 30.11.2022).
8. Лекция 1. Первичный анализ речевых сигналов // Alpha Cephei speech Recognition. – URL: https://alphacephei.com/ru/research (дата обращения: 12.12.2022).
9. *Панфилов И.А., Алексеев М.С., Сивцова Е.И.* Извлечение признаков голосового корсета // Цифровая трансформация экономических систем: проблемы и перспективы (экопром-2022). – СПб.: ПОЛИТЕХ-ПРЕСС, 2022. – С. 794-796.
10. *El Ayadi M., Kamel M.S., Karray F.* Survey on speech emotion recognition: Features, classification schemes, and databases // Pattern recognition. – 2011. – Vol. 44, No. 3. – P. 572-587.
11. *Abbaschian B.J., Sierra-Sosa D., Elmaghraby A.* Deep learning techniques for speech emotion recognition, from databases to models // Sensors. – 2021. – Vol. 21, No. 4. – P. 1249.
12. *Swain M., Routray A., Kabisatpathy P.* Databases, features and classifiers for speech emotion recognition: a review // International Journal of Speech Technology. – 2018. – Vol. 21. – P. 93-120.
13. *Chen L., Su W., Feng Y., Wu M., She J., Hirota K.* Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction // Information Sciences. – 2020. – Vol. 509. – P. 150-163.
14. *Gokilavani M., Katakam, H. Basheer.* Ravdness, crema-d, tess based algorithm for emotion recognition using speech // 2022 4th International conference on smart systems and inventive technology (ICSSIT). – IEEE, 2022. – P. 1625-1631.
15. *Zielonka M., Piastowski A., Czyżewski A., Nadachowski P., Operlejn M., Kaczor K.* Recognition of emotions in speech using convolutional neural networks on different datasets // Electronics. – 2022. – Vol. 11, No. 22. – P. 3831.
16. *Xie Y., Liang R., Liang Z., Huang C., Zou C., Schuller B.* Speech emotion classification using attention-based LSTM // IEEE/ACM Transactions on Audio, Speech, and Language Processing. – 2019. – Vol. 27, No. 11. – P. 1675-1685.
17. *Zhao J., Mao X., Chen L.* Speech emotion recognition using deep 1D & 2D CNN LSTM networks // Biomedical signal processing and control. – 2019. – Vol. 47. – P. 312-323.
18. *Wang J., Xue M., Culhane R., Diao E., Ding J., Tarokh V.* Speech emotion recognition with dual-sequence LSTM architecture // ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2020. – P. 6474-6478.
19. *Huang Z., Dong M., Mao Q., Zhan Y.* Speech emotion recognition using CNN // Proceedings of the 22nd ACM international conference on Multimedia. – 2014. – P. 801-804.
20. *Anvarjon T., Mustaqeem, Kwon S.* Deep-net: A lightweight CNN-based speech emotion recognition system using deep frequency features // Sensors. – 2020. – Vol. 20, No. 18. – P. 5212.

REFERENCES

1. Opensmile 2.4.2, *PyPI - Ukazatel' paketov Python'a* [PyPI - Python Package Index]. Available at: https://pypi.org/project/opensmile/ (accessed 12 December 2022).
2. Sentiment-Predictor-for-Stress-Detection-using-Voice // GitHub. – URL: https://github.com/sidmulajkar/sentiment-predictor-for-stress-detection (accessed 10 October 2022).
3. Audio Emotion | Part 1 - Explore data, *Kaggle*. Available at: https://www.kaggle.com/code/ejlok1/audio-emotion-part-1-explore-data (accessed 10 October 2022).
4. *Voronina V.V., Mikheev A.V., Yarushkina N.G., Svyatov K.V.* Teoriya i praktika mashinnogo obucheniya: ucheb. posobie [Theory and practice of machine learning: a textbook]. Ul'yanovsk: UlGTU, 2017, 290 p.
5. Surrey Audio-Visual Expressed Emotion (SAVEE) Database, *The Centre for Vision, Speech and Signal Processing. – URL: http://kahlan.eps.surrey.ac.uk/savee/* (accessed 29 November 2022).
6. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English, *PLOS*. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0196391 (accessed 28 November 2022).
7. CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset, *The National Center for Biotechnology Information*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4313618/ (accessed 30 November 2022).
8. Lektsiya 1. Pervichnyy analiz rechevykh signalov [Primary analysis of speech signals], *Alpha Cephei speech Recognition,* Available at: https://alphacephei.com/ru/research (accessed 12 December 2022).

9. *Panfilov I.A., Alekseev M.S., Sivtsova E.I.* Izvlechenie priznakov golosovogo korseta [Extraction of signs of a vocal corset], *Tsifrovaya transformatsiya ekonomicheskikh sistem: problemy i perspektivy (ekoprom-2022)* [Digital transformation of economic systems: problems and prospects (ecoprom-2022)]. Saint Petersburg: POLITEKH-PRESS, 2022, pp. 794-796.
10. *El Ayadi M., Kamel M.S., Karray F.* Survey on speech emotion recognition: Features, classification schemes, and databases, *Pattern recognition,* 2011, Vol. 44, No. 3, pp. 572-587.
11. *Abbaschian B.J., Sierra-Sosa D., Elmaghraby A.* Deep learning techniques for speech emotion recognition, from databases to models, *Sensors,* 2021, Vol. 21, No. 4, pp. 1249.
12. *Swain M., Routray A., Kabisatpathy P.* Databases, features and classifiers for speech emotion recognition: a review, *International Journal of Speech Technology,* 2018, Vol. 21, pp. 93-120.
13. *Chen L., Su W., Feng Y., Wu M., She J., Hirota K.* Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction, *Information Sciences,* 2020, Vol. 509, pp. 150-163.
14. *Gokilavani M., Katakam, H. Basheer.* Ravdness, crema-d, tess based algorithm for emotion recognition using speech, *2022 4th International conference on smart systems and inventive technology (ICSSIT).* IEEE, 2022, pp. 1625-1631.
15. *Zielonka M., Piastowski A., Czyżewski A., Nadachowski P., Operlejn M., Kaczor K.* Recognition of emotions in speech using convolutional neural networks on different datasets, *Electronics,* 2022, Vol. 11, No. 22, pp. 3831.
16. *Xie Y., Liang R., Liang Z., Huang C., Zou C., Schuller B.* Speech emotion classification using attention-based LSTM, *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* 2019, Vol. 27, No. 11, pp. 1675-1685.
17. *Zhao J., Mao X., Chen L.* Speech emotion recognition using deep 1D & 2D CNN LSTM networks, *Biomedical signal processing and control,* 2019, Vol. 47, pp. 312-323.
18. *Wang J., Xue M., Culhane R., Diao E., Ding J., Tarokh V.* Speech emotion recognition with dual-sequence LSTM architecture, *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 6474-6478.
19. *Huang Z., Dong M., Mao Q., Zhan Y.* Speech emotion recognition using CNN, *Proceedings of the 22nd ACM international conference on Multimedia,* 2014, pp. 801-804.
20. *Anvarjon T., Mustaqeem, Kwon S.* Deep-net: A lightweight CNN-based speech emotion recognition system using deep frequency features, *Sensors,* 2020, Vol. 20, No. 18, pp. 5212.

Статью рекомендовал к опубликованию д.т.н., профессор А.В. Боженюк.

Егорчев Антон Александрович – Казанский федеральный университет; e-mail: anton@egorchev.ru; Казань, Россия; директор института вычислительной математики и информационных технологий; к.т.н.

Пашин Дмитрий Михайлович – e-mail: dmitry.m.pashin@gmail.com; проректор по цифровой трансформации и инновационной деятельности; д.т.н.

Сарамбаев Никита Андреевич – e-mail: sarambaev@gmail.com; тел.: +79179029352; аспирант.

Фахрутдинов Адель Фердинандович – e-mail: timvaz@yandex.ru; тел.: +79872394153; аспирант.

Egorchev Anton Alexandrovich – Kazan Federal University; e-mail: anton@egorchev.ru; Kazan, Russia; director of Institute Computational Mathematics and IT; cand. of eng. sc.

Pashin Dmitry Mikhailovich – e-mail: dmitry.kfu@ya.ru; Vice-Rector for Digital Transformation and Innovation; dr. of eng. sc.

Sarambaev Nikita Andreevich – e-mail: sarambaev@gmail.com; phone: +79179029352; graduate student.

Fakhrutdinov Adel Ferdinandovich – e-mail: timvaz@yandex.ru; phone: +79872394153; graduate student.

УДК 004.94 DOI 10.18522/2311-3103-2024-4-101-110

А.В. Хлуденев, С.А. Сильвашко

СОПРЯЖЕННОЕ МОДЕЛИРОВАНИЕ БИС В САПР ПЕЧАТНЫХ УЗЛОВ

Виртуальное прототипирование выполняют в процессе разработки новых изделий с целью проверки проекта перед созданием физического прототипа, используя компьютерные модели. В САПР печатных узлов с этой целью используют SPICE симуляторы схем. Печатные узлы современных электронных устройств построены на одной или нескольких интегральных схем (ИС) высокой степени интеграции. Функционал больших интегральных схем (БИС) дополняют вспомогательные ИС и дискретные компоненты. В большинстве случаев требуемая эффективность обеспечивается при использовании БИС с процессорными ядрами. Поэтому симуляторы схем должны обеспечивать сопряженное моделирование аппаратных и программных средств. Приемлемыми по затратам вычислительных ресурсов являются модели БИС системного уровня. Основные достижения в области моделирования на системном уровне, включая сопряженное моделирование, связаны с разработкой самих БИС. В схемах печатных узлов они являются готовыми комплектующими изделиями. Эту специфику необходимо учитывать при реализации инструментов верификации схем печатных узлов. Модели БИС системного уровня должны встраиваться в модель всей схемы, быть экономичными и обеспечивать требуемую точность на границе внешних выводов. Модели цифровых БИС должны достоверно отображать задержки между изменениями уровней на выводах и диагностировать нарушения синхронизации. Модели БИС должны разрабатывать пользователи САПР печатных узлов, учитывая специфику проекта. Целью исследования является поиск решений для построения моделей БИС, включающих процессорные ядра, для прототипирования схем, используя OrCAD PCB Designer with PSpice. В статье рассматривается задача построения C/C++ модели для микроконтроллера семейства dsPIC, выполняющего обработку сигнала в реальном времени. Приведены пример построения модели, используя инструменты PSpice Model Editor, и результаты моделирования.

Печатный узел; виртуальное прототипирование; схемный симулятор; БИС; процессорное ядро; периферийный модуль; прямой доступ к памяти; модель системного уровня; сопряженное моделирование; симулятор системы команд; симуляция компилированного хостом программного обеспечения.

A.V. Khludenev, S.A. Silvashko

LSI COSIMULATION IN EDA FOR PCB DESIGN

Virtual prototyping is performed during product development to validate a design using a computer model before making a physical prototype. For this purpose, EDA for printed circuit board (PCB) design contain SPICE circuit simulator. Typically, modern PCB assemblies include one or more large-scale integrated circuits (LSI). The LSI functionality is complemented by auxiliary integrated circuits (IC) and discrete components. In most cases, the required efficiency is achieved by using LSIs that include processor cores. Therefore circuit simulators must provide a hardware/software co-simulation. System-level LSI models are acceptable in terms of computational resource costs. Major advances in system-level simulation, including co-simulation, come from the development of LSIs themselves. In PCB design, LSIs are fully fabricated components. This specificity must be taken into account when implementing tools for PCB design verifying. System-level LSI models must be integrated into the overall assembly circuit model. LSI models must provide the required accuracy only at the external pins. Models of digital LSIs must accurately reproduce delays between level changes at the pins and diagnose timing violations. EDA for PCB design users must develop LSI models tailored to the project specifics. The purpose of the research is to find solutions for building models of LSIs, containing processor cores, for prototyping circuits using OrCAD PCB Designer with PSpice. The article discusses the task of building a C/C++ model for the dsPIC33 microcontroller that performs signal processing in real time. An example of building a C/C++ model using the PSpice Model Editor tools and modeling results are given.

PCB assembly; virtual prototyping; circuit simulator; large-scale integrated circuit; processor core; peripheral module; direct memory access; system level model; co-simulation; instruction set simulator; host-compiled software simulation.

Введение. Стремление обеспечить расширенную функциональность и высокую надежность электронных устройств при снижении стоимости, потребляемой мощности, массы и габаритных размеров являются стимулом широкого применения в разработках БИС. В большинстве задач, некритичных по времени выполнения заданных функций, наилучшее соотношение цена – качество обеспечивается при реализации устройств на основе БИС микропроцессоров (МП) и микроконтроллеров (МК), далее МП БИС. Необходимая функциональность изделий обеспечивается выбором МП БИС с требуемыми характеристиками, а также в результате разработки и отладки встроенного программного кода. В конечных изделиях МП БИС вместе с другими компонентами устанавливаются в конструктивы на основе печатных плат. Разработку схем и конструкций таких изделий выполняют в среде САПР печатных узлов [1]. На этапе разработки электрической схемы печатного узла для анализа и верификации традиционно используют SPICE симуляторы и модели [2].

Симулятор PSpice A/D, входящий в состав САПР OrCAD PCB Designer, позволяет выполнять многоуровневый анализ схем, построенных на дискретных элементах, аналоговых и цифровых ИС. Макромодели цифровых ИС реализованы в виде подсхем, состоящих из базовых примитивов уровня логических вентилей [3]. В современных версиях PSpice A/D появилась возможность использовать алгоритмические C/C++ модели цифровых БИС системного уровня. В отличие от специализированных средств, например [4], в библиотеках OrCAD PCB Designer отсутствуют готовые модели БИС. Инструменты PSpice A/D ориентированы на ускоренный анализ схем печатных узлов, включающих от одной до нескольких БИС. Для этого следует использовать упрощенные проектно-ориентированные модели БИС, обеспечивающие требуемую точность на границе их внешнего интерфейса. Разработку таких моделей для своих проектов с учетом требований [5] должны выполнять пользователи. Исследование путей реализации PSpice C/C++ моделей цифровых БИС является актуальной задачей. В работе предлагается методика построения PSpice C/C++ моделей МП БИС, выполняющих целевую программу в режиме Bare Metal.

Постановка задачи. Для заданной МП БИС в составе схемы печатного узла построить математическую модель, отвечающую требованиям:

- должна встраиваться в PSpice модель схемы печатного узла;
- должна обеспечивать сопряженное моделирование аппаратных (HW) и программных (SW) средств;
- должна обеспечивать заданную точность моделирования сигналов на внешних выводах при минимальной сложности реализации и минимальных затратах вычислительных ресурсов.

Предлагаемая методика использует в качестве ресурса для выполнения последнего требования учет специфики выполняемых МП БИС задач и особенностей их технической реализации.

PSpice макромодели цифровых ИС. Основной задачей анализа схем цифровых устройств в PSpice A/D является обнаружение рассогласования сигналов во времени, рисков сбоя, нарушений условий синхронизации. Функционирование цифровых ИС в PSpice A/D описывается пятизначными моделями логических вентилей, триггеров и функциональных примитивов LOGICEXP, входящих в подсхему макромодели [3]. Свойства выводов описываются моделью входа - выхода, подсхемами интерфейса с аналоговой частью устройства и источниками питания. Динамические свойства описываются временной моделью, задающей задержки распространения сигналов с учетом разброса значений, и примитивом проверки временных ограничений CONSTRAINT.

Макромодель системного уровня в PSpice A/D также представляется подсхемой, в которую включен функциональный примитив LOGICEXP с квалификатором C_MODEL и ссылкой на dll файл исполняемого модуля модели, временной моделью типа ugate и моделью входов – выходов. Взаимодействие вычислительного ядра PSpice A/D с моделями устройств системного уровня обеспечивает Device Modeling Interface (DMI) [6]. Используя инструмент DMI Template code generator утилиты PSpice Model Editor, можно

автоматически сформировать файл PSpice макромодели и набор файлов проекта MS Visual Studio, включающий стандартные файлы DMI и файлы шаблона C/C++ модели БИС [7]. Исполняемый код модели БИС и DMI в виде dll файла создается в результате компиляции проекта в MS Visual Studio. Симулятор PSpice A/D загружает dll файл при выполнении Transient анализа.

Модели МП БИС системного уровня. Инструменты сопряженного HW/ SW моделирования появились и развивались для верификации проектов систем на кристалле, содержащих процессорные ядра [8–9]. Основные решения по разработке HW/SW симулятора связаны с реализацией модели процессорных ядер, выполняющих целевой программный код, симулятора аппаратных модулей МП БИС и синхронизации процессов в этих моделях. Модели более высокого уровня абстракции экономичнее по затратам вычислительных ресурсов на выполнение анализа, но менее точно учитывают время событий.

Модели процессорного ядра, способные имитировать его работу с точностью до машинного цикла, реализуют в виде симулятора набора команд (англ. instruction set simulator (ISS)) [10–11]. Код целевой программы формируют кросс компилятором. ISS интерпретирующего типа выполняет цикл, в котором симулируется выборка, декодирование и исполнение машинных команд целевым процессором. В ISS компилирующего типа инструкции целевого процессора преобразуют в эквивалентные макросы и затем компилируют в код хоста. В компилированной модели отсутствуют этапы выборки и декодирования инструкций, что приводит к ускорению моделирования [12]. Для сохранения точности учитывают затраченное время, используя механизм аннотации [13]. Скорость моделирования повышают, используя эффективные модели HW и механизмы синхронизации процессов [14]. Обзор коммерческих и академических инструментов сопряженного HW/SW моделирования приведен в [15].

Разработка моделей МП БИС на основе ISS оправдана при многократном применении в различных проектах. Для проектно ориентированных моделей эффективнее более простые решения. Точность модели на внешних выводах можно обеспечить при ее адекватности по времени при чтении и записи регистров портов и периферийных модулей. Время между другими событиями, протекающими в МП БИС, можно учитывать грубо. Часто в задачах реального времени процессорные ядра непосредственно не выполняют операции обмена с портами и регистрами данных периферийных модулей. В МП БИС с этой целью используют каналы прямого доступа к памяти (англ. Direct Memory Access (DMA)). В этом случае SW модель может быть "свободной по времени" и реализована путем компиляции адаптированного целевого C кода в исполняемый код хоста. Полагая, что целевая программа имеет модульную структуру и все изменения в регистрах и памяти происходят по фронту импульсов машинного цикла CLK, модель МП БИС можно построить, следуя правилам методики:

1) процессы в периферийных модулях протекают параллельно, поэтому после детектирования фронта CLK должна выполняться обработка моделей всех модулей (очередность обработки должна быть противоположной последовательности передачи сигналов между модулями, чтобы новые данные в регистрах действовали на входах приемников в следующем машинном цикле);

2) процессорные ядра выполняют операции обработки данных последовательно, поэтому после детектирования фронта CLK может выполняться моделирование только одного программного модуля (очередность обработки моделей программных модулей должна определяться алгоритмом программы);

3) вне очереди должно выполняться моделирование обработчиков прерываний при поступлении запросов на прерывания (при появлении нескольких запросов на прерывания должно выполняться моделирование обработчика для запроса с более высоким приоритетом).

Пример для построения модели. Экономичность реализации модели МП БИС зависит от учета особенностей реализации объекта моделирования. Рассмотрим в качестве примера фильтр с конечной импульсной характеристикой (КИХ) (рис. 1), реализованный

на МК dsPIC33FJ256GP506. МК семейства dsPIC33 содержат универсальное процессорное ядро MCU, ядро обработки сигналов DSP, контроллер DMA и набор периферийных модулей [16]. Процессорные ядра выполняют все инструкции (кроме команд передачи управления) за один машинный цикл, состоящий из двух машинных тактов. Процессорные ядра и периферийные модули тактируются от одного встроенного генератора с частотой импульсов F_{osc} = 80 МГц, частота машинного цикла F_{cy} = 40 МГц.

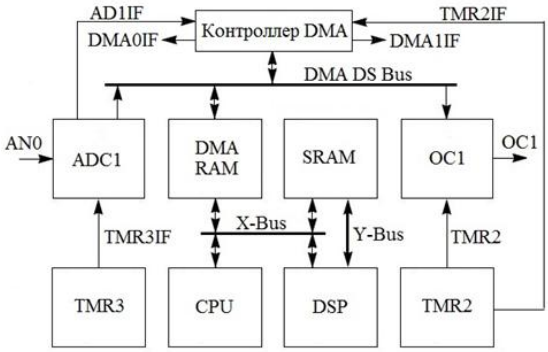


Рис. 1. Функциональная схема КИХ-фильтра

Для оцифровки входного сигнала с канала AN0 используется встроенный модуль аналого-цифрового преобразователя (АЦП) ADC1. АЦП последовательных приближений тактируется импульсами от делителя частоты с периодом $T_{AD} = 32/F_{cy}$. Запуск АЦП происходит при переполнении таймера TMR3 и установке флага TMR3IF, частота дискретизации F_s = 16 кГц. В течение интервала $3 \cdot T_{AD}$ происходит выборка входного напряжения, а затем в течение интервала $14 \cdot T_{AD}$ формируются 12 разрядов выходного кода [17]. При завершении преобразования устанавливается флаг AD1IF и инициирует передачу кода выборки по каналу DMA0, работающему в режиме ping-pong, в один их буферов обмена InBufferA или InBufferB, расположенных в области памяти данных DMA RAM [18]. При заполнении текущего буфера контроллер DMA устанавливает флаг прерывания DMA0IF и переходит к заполнению второго буфера.

Обработчик прерывания DMA0_IRQ инвертирует индикатор заполненного буфера InBufferIndicator и сбрасывает признак isReadBusy. Ядро MCU ожидает это событие и переходит к выполнению функции чтения заполненного буфера ADCChannelRead и копированию выборок в массив Samples в SRAM. При завершении копирования ADCChannelRead устанавливает признак isReadBusy. Схема алгоритма основного модуля целевой программы приведена на рис. 2.

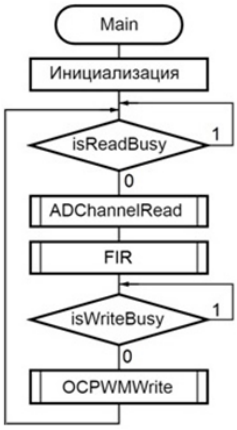


Рис. 2. Схема алгоритма целевой программы МК

Алгоритм КИХ-фильтрации выборок из массива Samples реализован в функции FIR. Основные операции выполняет ядро DSP при выборке данных из массивов линии задержки и коэффициентов. Коэффициенты были рассчитаны для формирования частотной характеристики полосового фильтра с полосой пропускания от 2 до 4 кГц. Подавление в диапазонах частот ниже 1 кГц и выше 5 кГц не менее 52 дБ. Обработанные выборки сохраняются в массиве Filterout в SRAM. Ядро MCU переходит к ожиданию сброса признака isWriteBusy.

Для восстановления аналогового сигнала используются широтно-импульсный модулятор (ШИМ), реализованный на встроенном модуле выходного компаратора OC1 [19], и внешний фильтр нижних частот (ФНЧ) [20]. Выходные выборки передаются из буфера OutBufferA или OutBufferB в памяти DMA RAM по каналу DMA1 в регистр выходного компаратора OS1RS по флагу TMR2IF при переполнении таймера TMR2. Модуль счета TMR2 определяет значение несущей частоты ШИМ сигнала F_{PWM} = 64 кГц. Канал DMA1 работает в одиночном блочном режиме без постинкремента адреса источника. Флаг прерывания DMA1IF устанавливается после четырехкратной передачи каждой выборки. Инкремент адреса источника и режим ping-pong реализованы программно в обработчике прерывания DMA1_IRQ. Когда все ячейки текущего буфера прочитаны, DMA1_IRQ инвертирует OutBufferIndicator, изменяет базовый адрес буфера и сбрасывает признак isWriteBusy. Ядро MCU переходит к выполнению функции OCPWMWrite для нормирования и записи элементов массива Filterout в свободный от чтения буфер обмена OutBufferA или OutBufferB. При завершении копирования OCPWMWrite устанавливает признак isWriteBusy. После выполнения функции OCPWMWrite ядро MCU переходит к ожиданию сброса признака isReadBusy.

Построение модели. Средства PSpice Model Editor формируют шаблон C/C++ модели только для БИС с полностью цифровым внешним интерфейсом. Поэтому для модуля АЦП была использована внешняя модель элемента ADC12break из библиотеки Breakout.OLB. Модель контроллера АЦП, формирующего сигнал ADC1IF, реализована в составе C/C++ модели. Для КИХ-фильтра были сформированы схемный символ для размещения на схеме (файл FIR_DMA.OLB), PSpice макромодель (файл FIR_DMA.LIB) и набор файлов проекта MS Visual Studio. В файле pspFIR_DMA.h проекта объявлены класс pspFIR_DMA и стандартные функции модели, идентификаторы переменных и констант (были дополнены функциями и идентификаторами пользователя). Файл проекта pspFIR_DMA.cpp содержит инсталлятор функций DMI installpspFIR_DMA. Шаблоны пользовательских функций модели приведены в файле FIR_DMA_user.cpp. Функция инициализации портов и переменных модели pspFIR_DMA::initialize вызывается PSpice A/D однократно при запуске Transient анализа. Функция pspFIR_DMA::evaluate вызывается PSpice A/D в каждом такте моделирования и выполняет операции:

- получает текущее время pTicks;
- получает PSpice уровни входных сигналов из pVectorStates;
- детектирует фронт синхросигнала CLK;
- преобразует PSpice уровни сигналов в C++ целочисленные значения;
- выполняет алгоритм C/C++ модели;
- преобразует C++ данные в PSpice уровни выходных сигналов;
- обновляет состояния выходов в pVectorStates.

Для данного примера наиболее простым решением будет построение модели, в которой в течение одного машинного цикла dsPIC33 хост полностью выполняет эквивалент одного модуля целевой программы. Схема алгоритма функции модели pspFIR_DMA::evaluate приведена на рис. 3. После обнаружения фронта CLK модель имитирует работу dsPIC33 в текущем машинном цикле. Получает уровни сигналов, проверяет флаги прерываний DMA0IF и DMA1IF. При поступлении запроса прерывания выполняет функцию обработчика DMA0_IRQ или DMA1_IRQ. При отсутствии запросов пре-

рывания анализирует значение переменной маркера (Marker). В каждом машинном цикле dsPIC33 хост выполняет только одну ветвь алгоритма модели с последующей передачей маркера. Операторы задержки при isReadBusy = 1 и isWriteBusy = 1 реализованы как пропуск передачи маркера.

После имитации работы процессорных ядер выполняется обработка моделей всех периферийных модулей в следующей последовательности:

1) канала DMA0 (обработка запроса AD1IF);
2) контроллера ADC1 (установка AD1IF через 17 периодов T_{AD} после запуска АЦП);
3) таймера TMR3 (установка TMR3IF при переполнении, запуск АЦП);
4) делителя частоты Clock Divider для тактирования ADC1;
5) канала DMA1 (обработка запроса TMR2IF);
6) выходного компаратора OC1 (сброс выхода OC1 при TMR2 = OC1RS);
7) таймера TMR2 (установка TMR2IF и выхода OC1 при переполнении).

В данном примере при использовании буферов DMA на 32 выборки один основной цикл целевой программы выполняется за 80000 машинных циклов. Из них при моделировании хост выполняет ветвь обработчика DMA1_IRQ восемь раз, остальные ветви (кроме ветвей пропуска) по одному разу.

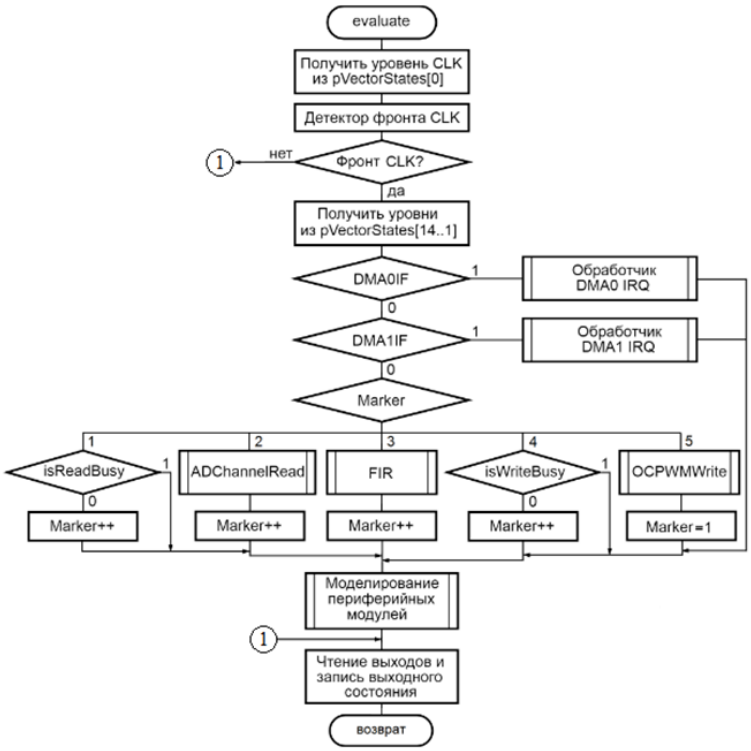


Рис. 3. Схема алгоритма функции модели pspFIR_DMA::evaluate

Результаты моделирования. На рис. 4 приведена схема для тестирования модели КИХ-фильтра в PSpice A/D. Параметры источников сигналов и аналоговых фильтров приведены на схеме. На рис. 5 приведен фрагмент результатов анализа реакции схемы на подачу двухтонального сигнала в течение 24 мс. Форма и параметры сигналов на выводах МП БИС соответствуют ожидаемым, подавляется гармоника с частотой 1 кГц. Грубые упрощения, допущенные при формировании модели SW компоненты, не оказывают влияние на полученные результаты. Параметры ШИМ сигнала на входе восстанавливающего фильтра (уровни высокого и низкого потенциалов и длительности фронтов) были учтены в PSpice модели входов - выходов КИХ-фильтра.

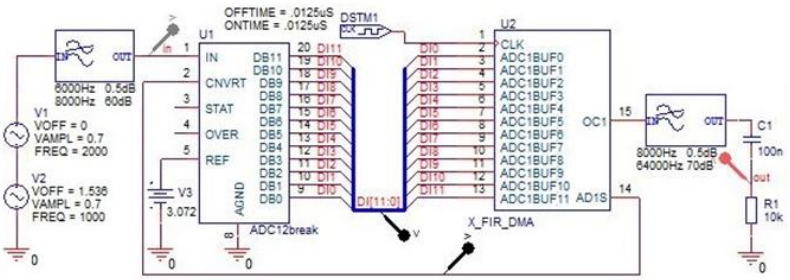


Рис. 4. Схема для тестирования модели КИХ-фильтра

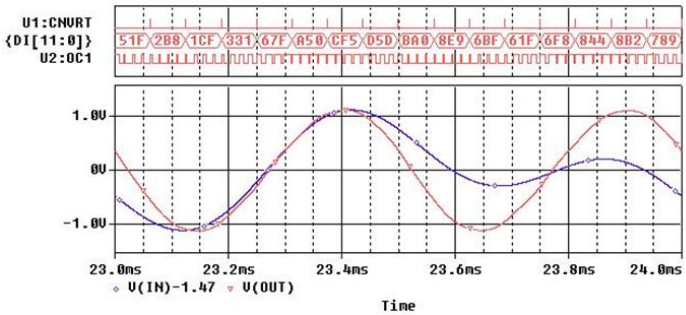


Рис. 5. Временные диаграммы

Модификации модели. Рассмотренный вариант модели будет давать идеализированные результаты работы МП БИС в случае, если процессорные ядра не будут успевать завершать обработку фреймов выборок до переключения буферов обмена контроллером DMA. Эти риски можно снизить, если выполнить доработку алгоритма модели МП БИС в одном из вариантов:

1) в течение одного машинного цикла хост полностью выполняет один модуль программы, при этом следующий программный модуль может быть выполнен после пропуска заданного количества машинных циклов;

2) если в модуле программы реализован циклический алгоритм, то в течение одного машинного цикла хост полностью выполняет один проход цикла программного модуля, следующий проход цикла программного модуля может быть выполнен после пропуска заданного количества машинных циклов.

Количество пропусков должно определяться временем выполнения программного модуля или одного прохода его цикла целевым процессором. В табл. 1 приведены значения времени выполнения программных модулей, определенные при выполнении кода, полученного кросс компилятором, с помощью инструментов симулятора – отладчика MPLAB IDE.

Таблица 1

Время выполнения модулей программы МК

Функция программы	Время, машинных циклов
DMA0_IRQ	20
DMA1_IRQ	22
ADCChannelRead	619
FIR	1048
OCPWMWrite	16388
Проверка isReadBusy	16
Проверка isWriteBusy	15

Заключение. Рассмотрена задача построения PSpice C/C++ моделей МП БИС для виртуального прототипирования схем печатных узлов. Учет особенностей технической реализации объекта моделирования позволяет снизить сложность реализации модели при заданной точности расчета сигналов на выводах МП БИС. Предложена методика построения модели МП БИС в форме алгоритма, управляющего вызовом эквивалентов модулей целевой программы и моделей периферийных модулей. При формировании эквивалентов на основе адаптированного С кода целевой программы МП БИС сокращаются трудоемкость и сроки разработки модели.

Приведен пример построения модели для МК семейства dsPIC33, выполняющего в реальном времени алгоритм КИХ-фильтрации при обмене кодами выборок входного и выходного сигналов с периферийными модулями через буферы DMA. При ограничениях на выполнение операций обмена с регистрами портов и периферийных модулей, точность модели на внешних выводах зависит от реализации моделей периферийных модулей (от машинного цикла до такта Transient анализа PSpice A/D). Такие модели целесообразно использовать при параллельном командном проектировании печатных узлов для верификации схем до завершения тестирования и отладки целевого программного кода. Предложены варианты модели МП БИС с учетом циклического характера алгоритмов и времени выполнения модулей целевой программы. В этих моделях необходимо учитывать время выполнения отлаженных модулей программы или одного прохода цикла.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. ГОСТ Р 70607–2022. Системы автоматизированного проектирования электроники. Состав и структура системы автоматизированного проектирования печатных узлов. – М.: Рос. ин-т стандартизации, 2023. – 8 с.
2. ГОСТ Р 70884–2023. Системы автоматизированного проектирования электроники. Информационное обеспечение. Модели SPICE. Общие положения. – М.: Рос. ин-т стандартизации, 2023. – 8 с.
3. PSpice user guide. Product version 17.2–2016. – Cadence, 2016. – 900 p. – URL: https:// resources.pcb.cadence.com/i/1180526-pspice-user-guide/25 (дата обращения: 20.05.2024).
4. *Васьков А.С., Смирнов Б.И.* Использование пакета Proteus для проектирования виртуальных моделей микропроцессорных устройств // Известия вузов России. Радиоэлектроника. – 2013. – Вып. 2. – С. 38-43.
5. ГОСТ Р 70806–2023. Системы автоматизированного проектирования электроники. Информационное обеспечение. Порядок разработки моделей SPICE. Постановка задачи. – М.: Рос. ин-т стандартизации, 2023. – 11 с.
6. PSpice device modeling interface API reference. Product version 17.2–2016. – Cadence, 2016. – 93 p.
7. Virtual prototyping in PSpice: Application note, V1. – FlowCAD, 2016. – 72 p. – URL: https://www.flowcad.de/AN/FlowCAD-AN_Device-Modeling-Interface.pdf (дата обращения: 20.05.2024).
8. *Teich J.* Hardware/software codesign: the past, the present, and predicting the future // Proc. IEEE. – 2012. – Vol. 100. – P. 1411-1430. – DOI: 10.1109/JPROC.2011.2182009.
9. *Rowson J.A.* Hardware/software co-simulation // Conf. on Design Automation. – 1994. – P. 439-440. – DOI: 10.1109/DAC.1994.204143.
10. *Liu J., Lajolo M., Sangiovanni-Vincentelli A.* Software timing analysis using HW/SW cosimulation and instruction set simulator // Proc. Int. Workshop on hardware/software codesign. – 1998. – P. 65-69. – DOI: 10.1145/278241.278299.
11. *Formaggio L, Fummi F., Pravadell G.* A timing-accurate HW/SW co-simulation of an ISS with SystemC // Proc. IEEE/ACM/IFIP Int. Conf. on Hardware/Software Codesign and System Synthesis. – 2004. – P. 152-157. – DOI: 10.1145/1016720.1016759.
12. *Zivojnovic V., Meyr H.* Compiled HW/SW co-simulation // Proc. Design Automation Conf. – 1996. – P. 690-695. – DOI: 10.1145/240518.240649.
13. *Lajolo M., Lazarescu M., Sangiovanni-Vincentelli A*. A compilation-based software estimation scheme for hardware/software co-simulation // Int. Conf. on Hardware/Software Codesign. – 1999. – P. 85-89. – DOI: 10.1145/301177.301493.
14. *Bringmann O., Ecker W., Gerstlauer A., [et al.]*. The next generation of virtual prototyping: Ultra-fast yet accurate simulation of HW/SW systems // Design, Automation & Test in Europe Conf. & Exhibition. – 2015. – P. 1698-1707. – DOI: 10.7873/DATE.2015.1105.

15. *Muttillo V., Pomante L., Santic M., Valente, G.* SystemC-based co-simulation / analysis for system-level hardware/software co-design // Computers and Electrical Engineering. – 2023. –Vol. 110. – 108803. – DOI: 10.1016/j.compeleceng.2023.108803.
16. dsPIC33FJXXXGPX06/X08/X10 high-performance, 16-bit digital signal controllers: Data sheet DS70286C. – Microchip Technology Inc., 2009. – 320 p. – URL: https://ww1.microchip.com/ downloads/en/DeviceDoc/70286C.pdf (дата обращения: 20.05.2024).
17. dsPIC33F/PIC24H family reference manual. Section 22. Direct memory access (DMA) DS70182A. – Microchip Technology Inc., 2006. – 60 p. – URL: http:// ww1.microchip.com/ downloads/en/ devicedoc/70183d.pdf (дата обращения: 20.05.2024).
18. dsPIC33F/PIC24H family reference manual. Section 16. Analog-to-digital converter (ADC) DS70183D. – Microchip Technology Inc., 2006-2012. – 88 p. – URL: https:// ww1.microchip.com/ downloads/en/DeviceDoc/70182A.pdf (дата обращения: 20.05.2024).
19. dsPIC33F family reference manual. Section 13. Output compare DS70209A. – Microchip Technology Inc., 2007. – 24 p. – URL: http://ww1.microchip.com/downloads/ en/DeviceDoc/70209A.pdf (дата обращения: 20.05.2024).
20. MPLAB starter kit for dsPIC® digital signal controllers: Users Guide DS51700A. – Microchip Technology Inc., 2008. – 42 p. – URL: https://ww1.microchip.com /downloads/en/DeviceDoc/51700B.pdf (дата обращения: 20.05.2024).

REFERENCES

1. GOST R 70607–2022. Sistemy avtomatizirovannogo proektirovaniya elektroniki. Sostav i struktura sistemy avtomatizirovannogo proektirovaniya pechatnykh uzlov [GOST R 70607–2022 Electronics automated design systems. Composition and structure of the computer-aided design of printed circuit assemblies]. Moscow: Ros. in-t standartizatsii, 2023, 8 p.
2. GOST R 70884–2023. Sistemy avtomatizirovannogo proektirovaniya elektroniki. Informatsionnoe obespechenie. Modeli SPICE. Obshchie polozheniya [GOST R 70884–2023 Electronics automated design systems. Information support. SPICE models. General provisions]. Moscow: Ros. in-t standartizatsii, 2023, 8 p.
3. PSpice user guide. Product version 17.2–2016. Cadence, 2016, 900 p. Available at: https:// resources.pcb.cadence.com/i/1180526-pspice-user-guide/25 (accessed 20 May 2024).
4. *Vas'kov A.S., Smirnov B.I.* Ispol'zovanie paketa Proteus dlya proektirovaniya virtual'nykh modeley mikroprotsessornykh ustroystv [Using the Proteus package for designing virtual models of microprocessor devices], *Izvestiya vuzov Rossii. Radioelektronika* [News of Russian universities. Radioelectronics], 2013, Issue 2, pp. 38-43.
5. GOST R 70806–2023. Sistemy avtomatizirovannogo proektirovaniya elektroniki. Informatsionnoe obespechenie. Poryadok razrabotki modeley SPICE. Postanovka zadachi [GOST R 70806–2023 Electronics automated design systems. Information support. Procedure for development of SPICE models. Formulation of the problem]. Moscow: Ros. in-t standartizatsii, 2023, 11 p.
6. PSpice device modeling interface API reference. Product version 17.2–2016. Cadence, 2016, 93 p.
7. Virtual prototyping in PSpice: Application note, V1. FlowCAD, 2016, 72 p. Available at: https://www.flowcad.de/AN/FlowCAD-AN_Device-Modeling-Interface.pdf (accessed 20 May 2024).
8. *Teich J.* Hardware/software codesign: the past, the present, and predicting the future, *Proc. IEEE,* 2012, Vol. 100, pp. 1411-1430. DOI: 10.1109/JPROC.2011.2182009.
9. *Rowson J.A.* Hardware/software co-simulation, *Conf. on Design Automation,* 1994, pp. 439-440. DOI: 10.1109/DAC.1994.204143.
10. *Liu J., Lajolo M., Sangiovanni-Vincentelli A.* Software timing analysis using HW/SW cosimulation and instruction set simulator, *Proc. Int. Workshop on hardware/software codesign,* 1998, pp. 65-69. DOI: 10.1145/278241.278299.
11. *Formaggio L, Fummi F., Pravadell G.* A timing-accurate HW/SW co-simulation of an ISS with SystemC, *Proc. IEEE/ACM/IFIP Int. Conf. on Hardware/Software Codesign and System Synthesis,* 2004, pp. 152-157. DOI: 10.1145/1016720.1016759.
12. *Zivojnovic V., Meyr H.* Compiled HW/SW co-simulation, *Proc. Design Automation Conf.,* 1996, pp. 690-695. DOI: 10.1145/240518.240649.
13. *Lajolo M., Lazarescu M., Sangiovanni-Vincentelli A*. A compilation-based software estimation scheme for hardware/software co-simulation, *Int. Conf. on Hardware/Software Codesign,* 1999, pp. 85-89. DOI: 10.1145/301177.301493.
14. *Bringmann O., Ecker W., Gerstlauer A., [et al.].* The next generation of virtual prototyping: Ultra-fast yet accurate simulation of HW/SW systems, *Design, Automation & Test in Europe Conf. & Exhibition,* 2015, pp. 1698-1707. DOI: 10.7873/DATE.2015.1105.

15. *Muttillo V., Pomante L., Santic M., Valente, G.* SystemC-based co-simulation / analysis for system-level hardware/software co-design, *Computers and Electrical Engineering,* 2023, Vol. 110, 108803. DOI: 10.1016/j.compeleceng.2023.108803.
16. dsPIC33FJXXXGPX06/X08/X10 high-performance, 16-bit digital signal controllers: Data sheet DS70286C. Microchip Technology Inc., 2009, 320 p. Available at: https://ww1.microchip.com/downloads/en/DeviceDoc/70286C.pdf (accessed 20 May 2024).
17. dsPIC33F/PIC24H family reference manual. Section 22. Direct memory access (DMA) DS70182A. Microchip Technology Inc., 2006, 60 p. Available at: http:// ww1.microchip.com/ downloads/en/devicedoc/70183d.pdf (accessed 20 May 2024).
18. dsPIC33F/PIC24H family reference manual. Section 16. Analog-to-digital converter (ADC) DS70183D. Microchip Technology Inc., 2006-2012, 88 p. Available at: https:// ww1.microchip.com/ downloads/en/DeviceDoc/70182A.pdf (accessed 20 May 2024).
19. dsPIC33F family reference manual. Section 13. Output compare DS70209A. Microchip Technology Inc., 2007, 24 p. Available at: http://ww1.microchip.com/downloads/ en/DeviceDoc/70209A.pdf (accessed 20 May 2024).
20. MPLAB starter kit for dsPIC® digital signal controllers: Users Guide DS51700A. Microchip Technology Inc., 2008, 42 p. Available at: https://ww1.microchip.com /downloads/en/DeviceDoc/51700B.pdf (accessed 20 May 2024).

Статью рекомендовал к опубликованию д.т.н., профессор Ю.А. Кравченко.

Хлуденев Александр Владимирович – Оренбургский государственный университет; e-mail: avhludenev@yandex.ru; г. Оренбург, Россия, тел.: +73532372874; кафедра промышленной электроники и информационно-измерительной техники; к.т.н.; доцент.

Сильвашко Сергей Анатольевич – e-mail: sisean@yandex.ru; тел.: +73532372874; кафедра промышленной электроники и информационно-измерительной техники; к.т.н.; доцент.

Khludenev Alexander Vladimirovich – Orenburg State University; e-mail: avhludenev@yandex.ru; Orenburg, Russia; phone: +73532372874; the department of Industrial Electronics and Information Measuring Engineering; cand. of eng. sc.; associate professor.

Silvashko Sergey Anatolievich – e-mail: sisean @yandex.ru; phone: +73532372874; the department of Industrial Electronics and Information Measuring Engineering; cand. of eng. sc.; associate professor.

УДК 004.912 DOI 10.18522/2311-3103-2024-4-110-122

Ю.М. Вишняков, Р.Ю. Вишняков

ФОРМАЛИЗАЦИЯ РАСПОЗНАВАНИЯ И ИДЕНТИФИКАЦИИ СЕМАНТИЧЕСКИХ ОБЪЕКТОВ В ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ТЕКСТОВЫХ ПОТОКАХ

Участившиеся случаи совершаемых в киберпространстве преступлений, в особенности, в социальных сетях и различного рода мессенджерах требуют создания адекватных и эффективных мер противодействия. Рост киберпреступлений настолько большой, что они уже могут нанести невосполнимый урон государству и обществу. Однако выявление подобного рода преступлений и преступных деяний, наталкивается на большие трудности, так как преступники присутствуют в социальных сетях виртуально и лингвистически, используют всячески их возможности и особенности для сокрытия следов своих преступлений. И, тем не менее, такими инструментами противодействия могли бы быть различного рода распознаватели и идентификаторы, способные автоматически обрабатывать естественный язык, выделять в нем специфические смысловые черты преступных деяний, распознавать и идентифицировать их. Поскольку по многим параметрам и обстоятельствам применительно к данным ситуациям использования нейросетевого подхода представляется нецелесообразным, в предлагаемой работе разрабатывается собственный формальный метод проектирования распознавателя для идентификации в текстовых потоках семантических объектов по их лингвистическим следам. Вводятся такие формальные понятия, как формальная модель семантического объекта, функция поведения, сценарий, лингвистический след, функция распознавания. Рассуждения строятся на теоретико-множественных положениях вычислительной теории семантической интерпретации и используют вычислительное представ-

ление смысла текстовых фрагментов для их сравнения на семантическую близость. Предлагаемый подход носит общий и универсальный характер, он позволяет формальным образом синтезировать распознаватель семантических объектов по их лингвистическим описаниям и поведению. В работе все рассуждения и построения иллюстрируются конкретными примерами.

Семантический объект; семантический след; семантический распознаватель; семантическое сравнение; текстовый фрагмент; семантическая близость; семантическая идентификация.

Yu.M. Vishnyakov, R.Yu. Vishnyakov

FORMALIZATION OF RECOGNITION AND IDENTIFICATION OF SEMANTIC OBJECTS IN NATURAL LANGUAGE TEXT STREAMS

The increasing incidence of crimes committed in cyberspace, particularly on social networks and various messengers, necessitates the development of adequate and effective countermeasures. The rise in cybercrime is so significant that it poses a potential threat of inflicting irreparable harm to the state and society. However, detecting such crimes and criminal activities is challenging because offenders operate virtually and linguistically within social networks, exploiting their features to conceal their traces. Nonetheless, various detection and identification tools capable of automatically processing natural language, highlighting specific semantic features of criminal activities, and recognizing and identifying them could serve as effective countermeasures. Given the impracticality of applying neural network approaches to these situations for several reasons, this study proposes a formal method for designing a recognizer to identify semantic objects in text streams based on their linguistic traces. Formal concepts such as the formal model of a semantic object, behavior function, scenario, linguistic trace, and recognition function are introduced. The reasoning is based on set-theoretical principles of computational theory of semantic interpretation and utilizes computational representations of the meaning of text fragments for their comparison in terms of semantic similarity. The proposed approach is general and universal, allowing for the formal synthesis of a recognizer for semantic objects based on their linguistic descriptions and behavior. All discussions and constructions in the work are illustrated with specific examples.

Semantic object; linguistic trace; semantic recognizer; semantic comparison; text fragment; semantic proximity; semantic identification.

Введение. В настоящее время обработка естественно-языковой информации NLP (Natural Language Processing) представляет одну из востребованных и активно развивающихся областей информатики, что связано с интенсивными коммуникациями в интернет-пространстве. В то же время NLP является и одной из слабо формализованных областей информатики, что определяется сложностью и многозначностью самого естественного языка.

В этой связи с авторы посчитали возможным дать некоторые ссылки на работы, которые с разных сторон касаются проблем NLP, хотя список подобного рода работ более чем обширен. Так, в работе [1–3] обсуждаются общие вопросы и проблемы прикладной лингвистики, а в работе [4] – язык программирования Python в качестве основного инструмента разработки. В работах [5, 6] демонстрируется использование языка программирования Python в прикладном аспекте – текстовом анализе и машинном обучении. Работы [7–9] посвящены решению конкретных вопросов обработки естественно-языковых данных, в работе [10] рассматривается вероятностная модель естественных языков, другие модели естественных языков рассмотрены в [11–14], модели языков в информационном поиске представлены [15, 16]. Из приведенного можно сделать вывод, что обработка естественно-языковой информации в различных аспектах, является более чем актуальна.

Однако со стороны пользователей постоянно усиливаются требования к точности и эффективности естественно-языковой обработки, и данное обстоятельство активно обсуждается в научном сообществе. Именно по этой причине совсем не случайно нейросетевой подход переживает ренессанс, но, как показывает практика, и он не представляет собой панацею.

Проблема состоит в том, что сегодня традиционная частотная парадигма семантики не в полной мере отвечает возрастающим требованиям пользователей и назрела необходимость в поиске иных подходов к моделированию смыслов. В этой связи уход в естественно-языковой обработке от частотной парадигмы представления семантики представляется весьма актуальным. В контексте данного обстоятельства авторы, например в

[17], развивают вычислительную теорию семантической интерпретации. Одним из ее приложений является поиск и идентификация объектов в естественно-языковых текстовых потоках [18, 19]. Под потоками понимаются чаты мессенджеров, социальные сети и пр., а под семантическим объектом - лингвистически представленный в текстовом потоке некий объект (субъект). Например, в одном случае это может быть музыкальный коллектив, в другом – роман или фильм, в третьем – преступная группа, создающие угрозы обществу.

Объекту присущи цель и индивидуальный образ, отображенные в характерные выражения и поведенческие схемы. Индивидуальность выражена в общей логике общения, в синтаксической структуре и логике сообщений, их семантике и пр. Именно такой лингвистический образ объекта понимается под семантическим объектом, а основная задача предлагаемого исследования заключается в том, чтобы выявить и идентифицировать присутствие объекта на основе его лингвистического образа и оставляемым им в текстовом потоке лингвистическим следам.

Востребованность подобных исследований явствует из публикации обозревателя «Новой газеты» Г. Мурсалиевой (НГ № 51 от 16.05.2016г.), которая рассматривалась в нашей работе [3]. В публикации шла речь об обосновавшейся в соцсети ВКонтакте преступной организации «Синий кит», склоняющих подростков к суициду. Владение приемами психологического воздействия в купе с возможностями социальных сетей для сокрытия следов преступлений позволяло преступникам долгое время действовать безнаказанно, создавая угрозы обществу, о чем свидетельствует большое число жертв преступлений. Очевидно, что подобного рода преступные деяния требует создания адекватных мер противодействия.

В связи с вышеизложенным целью настоящей работы является представление исследований по разработке методов формализованного описания и идентификации семантических объекта в текстовых потоках с оценкой степени достоверности. Предлагаемое исследование является развитием результатов работ [18, 19] и их продолжает.

Метод решения. В начале определим ряд базовых формальных понятий, которые положены в основу исследования.

Распознаватель семантических объектов (*Semantic Object Recognizer –RecSemObj*) представляет собой некоторое гипотетическое устройство (модель), который на основе информации о некотором семантическом объекте выявляет и идентифицирует присутствие данного семантического объекта в текстовом потоке и оценивает степень достоверности результата распознавания. *RecSemObj* можно представить тройкой вида:

$$RecSemObj = (T, SemObj, F), \quad (1)$$

где T – текстовый поток, *SemObj* (*Semantic Objekct – SemObj*) – семантический объект, F – функция распознавания.

Текстовый поток – это множество упорядоченных (во времени) и различимых по смыслу естественно-языковых конструкций (цепочек) – токенов (слов, предложений и их фрагментов):

$$T = \{t_{1,} t_2, \dots, t_{m,}\}. \quad (2)$$

Отдельный токен t_i, представляет собой элементарную неделимую единицу текстового потока.

Семантический объект (*Semantic objekct – SemObj*) – это тройка вида:

$$SemObj = (Q, P, Z), \quad (3)$$

где Q –множество лингвистических характеристик; P – функция поведения, Z – цель.

Лингвистическая характеристика. Лингвистическая характеристика q_i представляет собой слова, целостные по смыслу текстовые фрагменты или предложения, которые использует *SemObj* для достижения своих целей. Совокупность лингвистических характеристик семантического объекта (возможно упорядоченных и сведенных в классы) представляется в виде:

$$Q = \{q_1, q_2, \dots q_n\}. \quad (4)$$

Можно было бы сказать, что множество Q это своеобразный лексикон *SemObj*.

По аналогии с лексическим анализом лингвистической характеристике можно придать лексемоподобную нотацию:

$$< ID, \text{Текстовое значение}, \text{Класс} >, \tag{5}$$

где ID-идентификатор, Класс – принадлежность определенному классу (классы определяются исходя из конкретных целей). Подобное представление представляется удобным с точки зрения в реализации.

Из лингвистических характеристик можно составить различные последовательности (цепочки), причем потенциальное множество всех таких цепочек описывает полная итерация Q^* множества Q следующим образом:

$$\begin{cases} Q^* = \underbrace{Q^0 \cup Q^1 \cup Q^2 \cup \ldots . \cup Q^m}_{m \to \infty}; \\ Q^i == Q^{i-1} \times Q = Q \times Q^{i-1} = \underbrace{Q \times Q \times \ldots \times Q}_{i}; \\ Q^0 = \{\lambda\} - \text{аксиома}. \end{cases} \tag{6}$$

Здесь: × – операция прямого произведения множеств, λ – пустая цепочка, она не содержит ни одной характеристики. Q^0 – одноэлементное множество, состоящее из пустой цепочки λ. Других цепочек, не включенных в Q^*, нет.

Конкретное поведение *SemObj* направлено на достижение цели Z и задается некоторым сценарием. Сам сценарий составляется из последовательности текстовых заготовок, которыми являются лингвистические характеристики *SemObj:*

$$\propto = q_1^{\propto} q_2^{\propto} q_3^{\propto} \ldots q_r^{\propto}. \tag{7}$$

<u>Сценарий поведения *Sc* (*Sc - Scenario*) *SemObj*</u>. В общем случае семантический объект *SemObj* может иметь множество сценариев, приводящих к достижению цели Z:

$$\begin{cases} Sc = \{\alpha_1, \alpha_2, \ldots \alpha_n, \}; \\ Sc \subset Q^*. \end{cases} \tag{8}$$

<u>Функция поведения *P SemObj*</u>. Определим ее как функциональное соответствие (функцию) на множестве цепочек Q^* и множестве {0,1} и представим следующим образом:

$$P: Q^* \to \{0,1\}, \tag{9}$$

где

$$\begin{cases} P(\beta) = 1, \text{если } \beta \in Sc; \\ P(\beta) = 0, \text{ если } \beta \notin Sc.' \end{cases} \tag{10}$$

Данное функциональное соответствие задает поведение семантического объекта *SemObj*.

Исходя из (9) и (10), семантический объект *SemObj* можно представить в виде некоторого гипотетического устройства, которое последовательно отрабатывает сценарий *Sc,* строит функцию P(β) и достигает цели *Z*, общая схема которого показана на рис. 1:

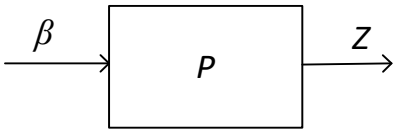


Рис. 1. Общая схема семантического объекта SemObj

Для построения формальной поведенческой модели *SemOb* можно воспользоваться различными подходами. Обсудим их кратко.

Нейросетевой подход. В этом случае множество сценариев *Sc* нужно рассматривать как обучающую выборку. Но опять же в случае бесконечного множества *Sc* придется создавать репрезентативную выборку, выполнять обучение сети и формировать функцию ошибки, что является далеко не тривиальной процедурой.

Другой – представляет собой формально грамматический подход. Его суть состоит в том, чтобы построить некоторую грамматику *G[Z]*. В данной грамматике *G[Z]* цель *Z* является ее начальным символом, множество лингвистических характеристик *Q* образуют словарь терминальных символов грамматики, отдельный сценарий α_i представляет собой предложение, а множество сценариев образует язык *L(G[Z])*. Тогда процедуру достижения цели можно представить выводом вида $Z => +\alpha_i$.

Еще один подход, который можно рассматривать как частный случай формально-грамматического подхода, основывается на понятиях регулярного множества, регулярного выражения и алгебры регулярных выражений. Здесь множество сценариев *Sc* соответствуют множеству значений некоторого регулярного выражения *E*, которое их порождает. И задача проектирования состоит в том, чтобы сконструировать регулярное выражение *E* и для него построить некоторое гипотетическое устройство (модель *SemObj*). Обсудим данный подход, поскольку, как нам представляется, большинство практических задач, укладываются в него. Кроме того, в языке программирования Pyhton, который стал основным средством разработки программистов, имеются средства описания и обработки подобного рода структур данных [2].

Обобщенное поведение семантического объекта можно представить диаграммой состояний. Как это показано на рис. 2, данная диаграмма имеет одно начальное состояние S, одно заключительное состояние Z и называется системой переходов.

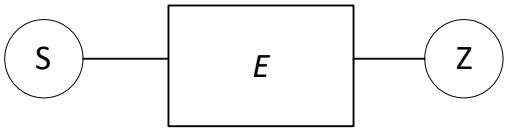


Рис. 2. Диаграмма состояний системы переходов семантического объекта

В системе переходов обобщенная дуга *E* (представлена прямоугольником) несет следующий смысл - любая цепочка $\alpha_i \in Sc$ переводит *SemObj* из состояния *S* в состояние *Z*.

Диаграмма состояний *SemObj*, в которой дуги переходов представляются только лингвистическими характеристиками и пустой цепочкой *λ*, называется приведенной. Будем считать, что модель *SemObj* построена, если она представлена своей приведенной системой переходов. Разбор сценария в приведенной системе переходов *SemObj* происходит последовательно по его лингвистическим характеристикам.

Для конструирования *SemObj* воспользуемся алгеброй регулярных выражений. В нашем случае алгебра определяется аксиоматически через операции "|" – ИЛИ, "*" – катенации, "{}" – итерации следующим образом:

1. ∅ (пустое множество), λ (пустая цепочка) и имена лингвистических характеристик $q_1,\ q_2, \dots q_n$ – регулярные выражения по определению (аксиома, основание);
2. Если e_1 и e_2 – регулярные выражения, то $e_1|e_2$ и $e_1 * e_2$ также регулярные выражения. Обычно в регулярных выражениях знак операции * опускают.

Например, если $e_1 = q_1|q_2$ и $e_2 = q_1q_2$, то $e_1e_2 = (q_1|q_2)q_1q_2 = q_1q_1q_2|q_2q_1q_2$ и для $e_1|e_2 = (q_1|q_2)|q_1q_2 = q_1|q_2|q_1q_2$;

1. Если e регулярное выражение, то и $\{e\}$ также регулярное выражение. В данном случае фигурные скобки {} – представляют операцию итерации;
2. Других регулярных выражений нет.

Диаграммы состояний систем переходов элементарных регулярных выражений представлены на рис. 3:

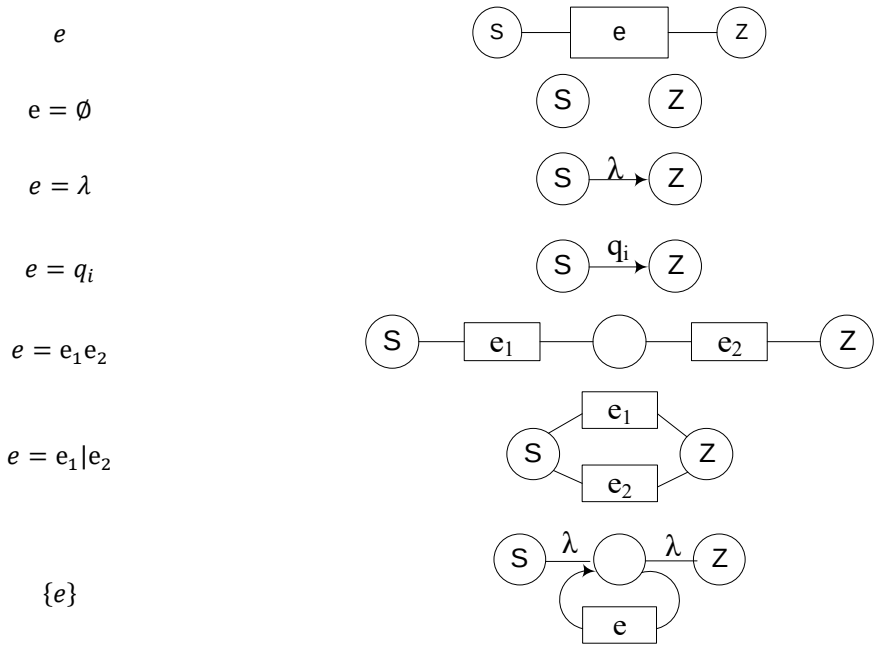


Рис. 3. Диаграммы состояний систем переходов элементарных регулярных выражений

Конструирование *SemObj* состоит в преобразовании системы переходов *SemObj*, показанной на рис. 2, к приведенной системе переходов. Преобразование выполняется декомпозицией получаемой на каждом шаге системы переходов по правилам рис. 3 до тех пор, пока не получится искомая приведенная система переходов.

Пример конструирования приведенной системы переходов для регулярного выражения $E = (q_1|q_2)q_3$ показан ниже на рис. 4:

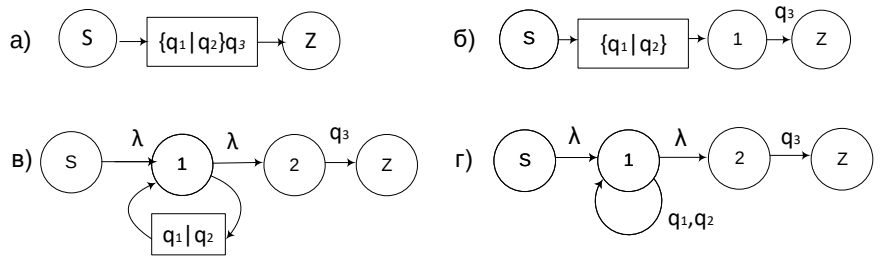


Рис. 4. Пример конструирования диаграммы состояний приведенной системы переходов

Заметим, что регулярное выражение *e* представляет собой формулу алгебры регулярных выражений. Значение *e* обозначается как $|e|$ и представляет множество цепочек, порождаемых регулярным выражением *e*. В случае лингвистической характеристики q_i значение имеет вид:

$$|q_i| = \{\text{"текст"}\}. \tag{11}$$

Например, пусть $|q_1| = \{$"расскажи о себе"$\}$, $|q_2| = \{$"ты сильный"$\}$. Значения для составных регулярных выражения имеют вид $|(q_1|q_2)| = \{$расскажи о себе, ты сильный$\}$ и $|q_1 q_2| = \{$расскажи о себе ты сильный$\}$.

Значение регулярного выражения $\{e\}$ представляется в виде $|\{e\}| = \{\lambda, e, ee, eee, \ldots \underbrace{ee \ldots e}_{n \to \infty}\}$. Например, если $e = q_2$, то $|\{q_2\}|$ представляет множество

$\{\lambda, \text{ты сильный}, \text{ты сильный ты сильный}, \text{ты сильный ты сильный ты сильный} \ldots\}$.

Пусть для *SemObj* рис. 4 заданы следующие конкретные сценарии, которые он отрабатывает:

$$\begin{cases} \alpha_1 = q_3 \\ \alpha_2 = q_1 q_2 q_3 \\ \alpha_3 = q_1 q_1 q_2 q_2 q_3 \end{cases}. \quad (12)$$

Отображение приведенной системы переходов на сценарий будем называть линейной разверткой *SemObj* и для (12) такие линейные развертки имеют вид:

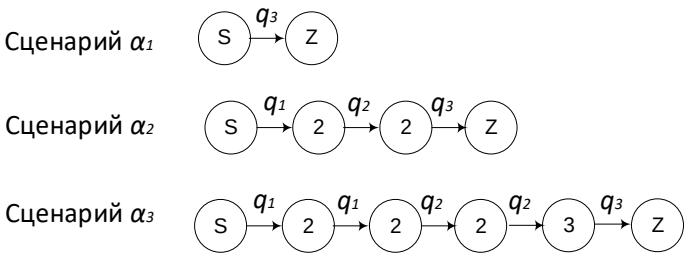


Рис. 5. Диаграммы состояний линейных разверток семантического объекта SemObj

Обратимся к вычислительной теории семантической интерпретации [8] и обсудим сравнение тестовых цепочек на семантическую близость. Пусть задана некоторая цепочка слов q:

$$q = a_1 a_2 \ldots a_n . \quad (13)$$

Она представляет целостный по смыслу текстовый фрагмент (предложение или его часть). Целостность означает, что каждое слово a_i входит в словосочетания, а сами словосочетания образуют отношение непосредственного подчинения на множестве слов цепочки.

Пусть $S(a_i)$ множество смысловых значений слова a_i, тогда множество смысловых значений цепочки $S(q)$ представляется функционалом вида:

$$\begin{cases} S(q) = \Phi\big(S(a_1), S(a2), \ldots, S(a_n)\big); \\ S(q) \subset S(a_i), \text{если } a_i \text{ главное слово } q. \end{cases} \quad (14)$$

Пусть заданы два текстовых фрагмента q и t и их смысловые значения S(q) и $S(t)$ соответственно. Введем функцию, которая определяет их смысловую близость и назовем ее критерием семантической близости *Cprox* [8]:

$$\begin{cases} Cprox\big(S(q), S(t)\big) \in D; \\ D = [0..1]. \end{cases} \quad (15)$$

Если $Cprox=0$, то семантическая близость между q и t отсутствует, при $Cprox=1$ имеет место полная семантическая близость.

В простейшем случае критерий семантической близости можно представить долей совпадающих элементов смысла p в семантической схеме цепочек q и t к общему числу элементов смысла m цепочки q:

$$Cprox\big(S(q), S(t)\big) = \frac{p}{m}. \quad (16)$$

Отметим, что традиционный частотный критерий релевантности является частным случаем (16), когда m представляет число слов в цепочке q, а p – число слов из q, совпадающих со словами в цепочке t. Из (16)) следует справедливость соотношения:

$$Cprox\big(S(q), S(t)\big) \neq Cprox\big(S(t), S(q)\big). \quad (17)$$

Из выше обсужденного для нас важен факт возможности измерения семантической близости двух текстовых отрезков.

Как ранее отмечалось, в текстовом потоке имеются токены, семантика которых в той или иной степени близка лингвистическим характеристикам *SemObj.* Такие токены будем называть лингвистическими следами характеристик, а их совокупность лингвистическим следом семантического объекта *SemObj*. Формализуем данные понятия.

Лингвистический след характеристики для некоторого *SemObj* (3) представляется тройкой вида:

$$\begin{cases} (q_i, t_j, w_k): \\ q_i \in Q,\ t_j \in T,\ \ w_k = Cprox(q_i, t_j)\ ,\ \ 0 < w_k \leq 1, \end{cases} \tag{18}$$

здесь величина w_i является мерой семантической близости, которую можно рассматривать как меру присутствия характеристики q_i в текстовом потоке.

Пусть тройка (q,w,t) задает некоторый семантический след в текстовом потоке, тогда из (18) по определению справедливо соотношение:

$$(q, w) = t. \tag{19}$$

Функция распознавания F . Пусть *SemObj* реализует в тестовом потоке *T* один из сценариев $\alpha \in Sc$ вида $\alpha = q_1 q_2 \dots q_n$, тогда существует такая же последовательность семантических следов (q_i, t_j, w_k) и последовательность токенов $\propto' = t_{r1} t_{r2} \dots t_{rn}$, представляющая отображение (19) сценария α на текстовый поток *T*. Тогда в общем случае функция распознавания имеет вид:

$$\begin{cases} F(\alpha, \propto') \in D: \\ D = [0..1]. \end{cases} \tag{20}$$

Функции распознавания *F* можно придать смысл сравнения на семантическое подобие (*Csemsim* - comparison for semantic similarity) сценария α и последовательности токенов $\propto'$ следующим образом:

$$F = Csemsim(\alpha, \propto'). \tag{21}$$

$Csemsim(\alpha, \propto')$ можно представить в виде:

$$Csemsim(\alpha, \propto') = Csemsim\big(Cprox(q_1, t_{r1}), Cprox(q_2, t_{r2}), \dots, Cprox(q_n, t_{rn})\big) \tag{22}$$

или с учетом (18) получим:

$$Csemsim(\alpha, \propto') = Csemsim(w_{k1}, w_{k2}, \dots, w_{kn}\ ). \tag{23}$$

В зависимости от практических целей для *F* можно подбирать различные виды функциональной аппроксимации, обеспечивая тонкую настройку *RecSemObj*. В нашем случае для понимания вполне достаточно понимать *F* как средневзвешенную величину:

$$F = Csemsim(\alpha, \propto') = \frac{(w_{k1} + w_{k2} + \dots + w_{kn}\ )}{n} \tag{24}$$

Рассмотрим функцию распознавания для сценария α_3 (12) *SemObj,* показанного на рис. 4:

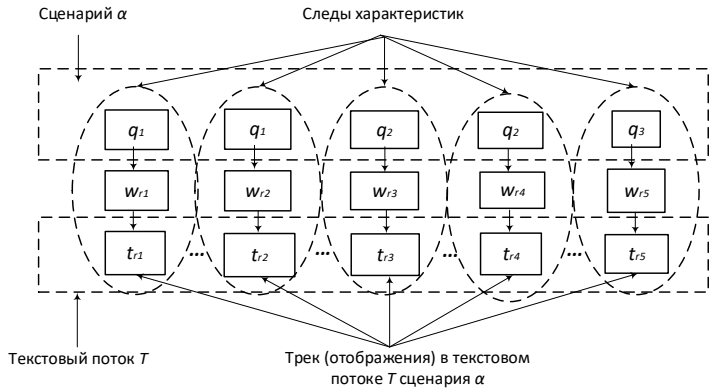


Рис. 6. Схема семантических следов SemObj в текстовом потоке

На рис. 6 представлен некоторый текстовый поток *T*, в который встроен сценарий α_3 *SemObj*. Сценарий представлен последовательностью характеристик $\alpha_3 = q_1 q_1 q_2 q_2 q_3$, которые будучи отображены в текстовый поток определяют в нем трек токенов $\alpha_3' = t_{r1} t_{r2} t_{r3} t_{r4} t_{r5}$. Характеристики сценария, токены трека и степени их семантической близости задают семантические следы *SemObj* в текстовом потоке *T*. На рис. 6 семантические следы выделены пунктирными овалами.

Выполнив подстановку в $\alpha_3' = t_{r1} t_{r2} t_{r3} t_{r4} t_{r5}$ в соответствии с (19), получим:

$$\alpha_3' = (q_1, w_{p1})(q_1, w_{p2})(q_2, w_{p3})(q_2, w_{p4})(q_3, w_{p5}), \tag{25}$$

тогда для трека токенов α_3' функция распознавания будет иметь вид:

$$F = Csemsim(\alpha, \propto') = \frac{(w_1 + w_2 + w_3 + w_4 + w_5)}{5}. \tag{26}$$

Проиллюстрируем вышеизложенное на примере реально подобного текстового потока. В нем некто Куратор (К) склоняет Жертву (Ж) совершить «ЭТО».

Набор лингвистических характеристик $Q = \{q_1, q_2, \dots q_{30}\}$ *SemObj* “Куратор”, сведанных в классы “Контакт”, “Вовлечение”, “Игра”, представлен в табл. 1. Поведение *SemObj* “Куратор” описывается нижеуказанными регулярным выражением (27) и приведенной системой переходов рис. 7:

$$E = q_1\{q_2|q_3| \dots |q_7\}q_8\{q_9|q_{10}| \dots |q_{16}\}q_{17}\{q_{16}|q_{17} \dots |q_{29}\}q_{24} \tag{27}$$

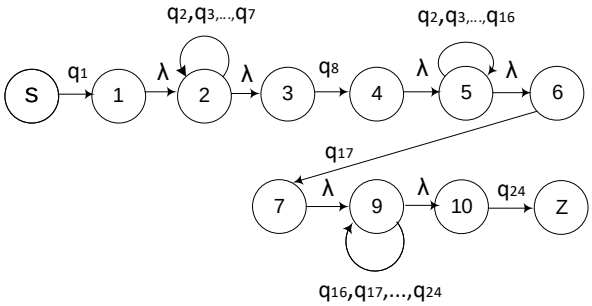


Рис. 7. Диаграмма состояний приведенной системы переходов SemObj “Куратор”

Таблица 1

Лингвистические характеристики *SemObj* “Куратор”

ID	Значение характеристики	Класс	ID	Значение характеристики	Класс
q_1	Привет, как зовут тебя?	КОНТАКТ	q_{16}	проблемы уйдут	ИГРА
q_2	расскажи о себе		q_{17}	твое задание	
q_3	Я живу в городе Н.		q_{18}	сядь на стул и замри	
q_4	попробую угадать твой адрес		q_{19}	молодец, ты справился	
q_5	пять лет назад переехал с родителями сюда		q_{20}	ножом надави на руку	
q_6	Мой папа врач, а мама учительница		q_{21}	терпи, даже если будет кровь	
q_7	раньше часто сорился с родителями		q_{22}	нужно продержаться пять минут	
q_8	когда-то я освоил одну игру	ВОВЛЕЧЕНИЕ	q_{23}	ты сильный	
q_9	игра помогло мне		q_{24}	сделай ЭТО	
q_{10}	стал уверенным и смелым		q_{25}	ничего не бойся	
q_{11}	веду игру в соцсетях и помогаю людям		q_{26}	избавишься от проблем и страданий	
q_{12}	даю задания, ты выполняешь		q_{27}	ты слабак и неудачник	
q_{13}	у тебя всё получится		q_{28}	ты никому не нужен	
q_{14}	Нужно делать все, что я говорю		q_{29}	наказание будет страшным	
q_{15}	закалим твой характер				

В табл. 2 представлен текстовый поток и для *SemObj* “Куратор” один из возможных сценариев α, который будучи отображен в текстовый поток порождает трек токенов $\propto'$.

Таблица 2

Текстовый поток с треком токенов сценария $\propto'$

№ пп	(q_i,w)	Трек сценария $\propto$
1.	**(q_1,1,0)**	К. **Привет,** я Костя, а **как зовут тебя?**
2.		Ж. Я Миша.
3.	**(q_2,1,0)**	К. Миша**, расскажи о себе**.
4.	**(q_3,1,0)**	К. **Я живу в городе Н.** А ты?
5.		Ж. Ого! А я тоже живу в городе Н.
6.	**(q_4,1,0)**	К. **Попробую угадать твой адрес**. Улица Усова, дом 35?
7.		Ж. Угадал! Здорово!
8.	**(q_5,0,5)**	К. Миша, я **переехал сюда с родителями** из города N.
9.	**(q_6,1,0)**	К. **Мой папа врач, а мама учительница.**
10.	**(q_7,1,0)**	К. Я **раньше часто сорился с родителями**, а ты, Миша?
11.		Ж. Постоянные ссоры, меня не понимают, заклевали!
12.	**(q_8,1,0)**	К. Миша, **я когда-то освоил одну игру**.
13.	**(q_9,1,0)**	К. **Игра помогла мне.**
14.		К. Я решил свои проблемы, наладились отношения.
15.	**(q_{10},1,0)**	К. **Стал уверенным и смелым.**
16.	**(q_{11},1,0)**	К. Теперь сам **веду игру в соцсетях и помогаю людям.**
17.		Ж. Костя, интересно! А что это за игра?
18.	**(q_{12},1,0)**	К. Я **даю задания, а ты выполняешь**.
19.	**(q_{13},1,0)**	К. Уверен, **у тебя всё получится**.
20.	**(q_{14},1,0)**	К. **Нужно делать все, что я говорю**.
21.	**(q_{15},1,0)**	К. Мы **закалим твой характер.**
22.	**(q_{16},1,0)**	К. **Проблемы** сами **уйдут**.
23.		К. Начнем?
24.		Ж. Да, хорошо, попробуем
25.	**(q_{17},1,0)**	К. Это **твое** первое **задание**.
26.	**(q_{18},1,0)**	К. **Сядь на стул и замри** на 15 минут.
27.		Ж**.** У меня получилось! Ура!
28.	**(q_{19},1,0)**	К. **Молодец, ты справился**!
29.	**(q_{17},0,5)**	К. Завтра новое **задание**
30.		Ж. Буду ждать, интересно!
31.	**(q_{17},1,0)**	К. Привет, Миша. Теперь **твое** новое **задание**.
32.		К. Укрепи характер.
33.		Ж. Костя, я весь внимание
34.	**(q_{20},1,0)**	К. Острием **ножа** сильно **надави на руку**.
35.	**(q_{21},1,0)**	К. **Терпи, даже если будет кровь.**
36.	**(q_{22},0,5)**	К. **продержись** хотя бы 10 **минут**.
37.		Ж. Я не смогу это сделать.
38.	**(q_{23},1,0)**	К. Сможешь, **ты сильный**
39.		Ж. Хорошо, я попробую.

40.		Ж. Я сделал, получилось!
41.	**(q_{23},1,0)**	К. Молодец, **ты сильный!**
42.	**(q_{17},1,0)**	К. Теперь **твое** самое сложное **задание**
43.	**(q_{24},1,0)**	К. **Сделай ЭТО!**
44.	**(q_{25},1,0)**	К. **Ничего не бойся!**
45.		Ж. Я не могу, я боюсь. Если меня не станет? А как же родители?
46.	**(q_{26},0,7)**	К. Разом **избавишься от** всех **проблем!**
47.	**(q_{27},1,0)**	К. Для друзей **ты слабак и неудачник**
48.	**(q_{28},1,0)**	К. **Ты никому не нужен!**
49.	**(q_{24},1,0)**	К. Ну! давай же! **сделай ЭТО!**
50.		Ж. Я очень боюсь!
51.	**(q_{28},1,0)**	К. Давай же, **ты никому не нужен!**
52.	**(q_{24},1,0)**	К. Немедленно **сделай ЭТО!!**
53.		Ж. Родители любят меня!
54.	**(q_{28},1,0)**	К. **Ты никому не нужен**!
55.	**(q_{24},1,0)**	К. **Сделай ЭТО!**
56.	**(q_{26},0,7)**	К. **Избавься от страданий!**
57.	**(q_{24},1,0)**	К. **Сделай ЭТО**
58.		Ж. Я боюсь, пожалей меня!
59.	**(q_{29},1,0)**	К. Не зли меня, **наказание будет страшным**
60.	**(q_{24},1,0)**	К. **Сделай ЭТО!**

В каждом токене жирным шрифтом выделены слова порождающей его лингвистической характеристики и указана степень *w* семантической близости с ней. Вычисленное значение функции распознавания $\mathrm{Csemsim}(\alpha, \propto')$ для нашего примера равно 0,95, что с высокой долей уверенности указывает на присутствует в потоке семантического объекта *SemObj* “Куратор”, реализующего свой сценарий поведения для достижения цели.

Эксперименты. Проведение экспериментов с представленной моделью семантического объекта осуществлялось на специально разработанном программном комплексе [20]. Эксперименты проводились для разных поведенческих сценариев семантического объекта и в различных текстовых потоках, они показали хорошую работоспособность сконструированного распознавателя семантических объектов и соответствие процесса распознавания заявляемым теоретическим характеристикам.

Заключение. В работе предложен формальный метод проектирования распознавателя для идентификации в естественно-языковых текстовых потоках семантических объектов по их лингвистическим следам. Разработаны формальная модель семантического объекта, функция поведения, сценарий, лингвистический след, функция распознавания. Предложена формальная методика синтеза распознавателя, основанная на алгебре регулярных выражений и автоматной модели – системе переходов. Для сравнения текстовых фрагментов на семантическую близость использовано вычислительное представление смысла. По мнению авторов предложенные решения можно рекомендовать в первую очередь для обнаружения и предотвращения преступлений в социальных сетях.

Исследование выполнено при финансовой поддержке Кубанского научного фонда в рамках научно-инновационного проекта «НИП-20.1.4».

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Николаев И.С.* Компьютерная и прикладная лингвистика / Николаев И.С., Митренина О.В., Ландо Т.М. (ред.). – М.: Ленанд, 2016. – 316 с.
2. *Шемякин Ю.И.* Начала компьютерной лингвистики: учеб. пособие. – М.: Изд-во МГОУ, А/О "Росвузнаука", 1992. – 120 с.

3. *Тестелец Я.Г.* Введение в общий синтаксис: учеб. пособие. – М.: Изд-во РГГУ, 2001. – 830 с.
4. *Прохоренок Н.А., Дронов В.А.* Python 3. Самое необходимое. – СПб.: БХВ-Петербург, 2019. – 608 с.
5. *Бенгфорт Бенджамин, Билбро Ребекка, Охеда Тони.* Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. – СПб.: Питер, 2019. – 368 с.
6. *Луис Педро Коэльо, Вилли Ричарт.* Построение Системы машинного обучения на языке Python: пер. с англ. Слинкин А.А. – М.: ДМК Пресс, 2016. – 302 с.
7. *Koncel-Kedziorski R., Hajishirzi H. and Sabharwal A. et al.* Parsing algebraic word problemsinto equations // Transactions of the Association for Computational Linguistics. – 2015. – 3. – P. 585-597.
8. *Devlin J., Chang M.W., Lee K. and Toutanova K.* Bert: Pre-training of deep bidirectional transformers for language understanding // arXiv preprint arXiv:1810.04805, 2018.
9. *Hu K., Wu H. and Qi K. et al.* A domain keyword analysis approach extending Term Frequency-Keyword Active Index with Google Word2Vec model // Scientometrics. – Springer, 2017. – P. 1-38.
10. *Налимов В.В.* Вероятностная модель языка. О соотношении естественных и искусственных языков. – М.: Наука, 1979. – 303 с.
11. *Гладкий А.В.* Синтаксические структуры естественного языка в автоматизированных системах общения. – М.: Наука. Главная редакция физико-математической науки, 1985. –144 с. – (Серия «Проблемы искусственного интеллекта»).
12. *Найханова Л.В., Евдокимова И.С.* Методы и алгоритмы трансляции естественно-языковых запросов к базе данных в SQL-запросы: монография. – Улан-Удэ: Изд-во ВСГТУ, 2004. – 148 с.
13. *Севбо И.П.* Графическое представление синтаксических структур и стилистическая диагностика. – Киев: Наукова думка, 1983. – 192 с.
14. *Рубашкин В.Ш.* Представление и анализ смысла в интеллектуальных информационных системах. – М.: Наука. Главная редакция физико-математической науки, 1989. – 192 с. – (Серия «Проблемы искусственного интеллекта»).
15. *Aggarwal C.C., Al-Garawi F., Yu P.S.* Intelligent crawling on the world wide web with arbitrary predicates // In Proc. of the WWW10, May 2001. – P. 96-105.
16. *Agichtein E., Lawrence S., Gravano L.* Learning search engine specific query transformations for question answering // In Proc. of the WWW10. – 2001. – P. 169-178.
17. *Вишняков Ю.М., Вишняков Р.Ю.* Вычислительная семантическая интерпретация текстов научно-технического стиля // Современные наукоемкие технологии. – 2016. – № 12–2. – С.236-242.
18. *Vishnyakov R.Y., Vishnyakov Y.M.* Identification of semantic objects // Journal of Physics: Conference Series, Bristol May 2021. – Vol. 1902. – DOI: 10.1088/1742–6596/1902/1/012104.
19. *Вишняков Ю.М. Вишняков Р.Ю.* О применении вычислительной теории семантической интерпретации к выявлению киберпреступлений // Актуальные проблемы прикладной математики, информатики и механики: Сб. трудов международной научно-технической конференции, Воронеж, 13–15 декабря 2021 г. – Воронеж, 2022. – С. 1526-1530.
20. *Степанова Е.В., Вишняков Ю.М.* Программное обеспечение для вычисления семантической близости текстов // Прикладная математика: современные проблемы математики, информатики и моделирования: Сб. материалов III-й Всероссийской научно-практической конференции молодых ученых. – Краснодар, 2021. – С. 218-223.

REFERENCES

1. *Nikolaev I.S.* Komp'yuternaya i prikladnaya lingvistika [Computer and applied linguistics], Nikolaev I.S., Mitrenina O.V., Lando T.M. (ed.). Moscow: Lenand, 2016, 316 p.
2. *Shemyakin Yu.I.* Nachala komp'yuternoy lingvistiki: ucheb. posobie [The beginnings of computational linguistics: a textbook]. Moscow: Izd-vo MGOU, A/O "Rosvuznauka", 1992, 120 p.
3. *Testelets Ya.G.* Vvedenie v obshchiy sintaksis: ucheb. posobie [Introduction to the general syntax: a textbook]. Moscow: Izd-vo RGGU, 2001, 830 p.
4. *Prokhorenok N.A., Dronov V.A.* Python 3. Samoe neobkhodimoe [Python 3. The most necessary]. Saint Petersburg: BKhV-Peterburg, 2019, 608 p.
5. *Bengfort Bendzhamin, Bilbro Rebekka, Okheda Toni.* Prikladnoy analiz tekstovykh dannykh na Python. Mashinnoe obuchenie i sozdanie prilozheniy obrabotki estestvennogo yazyka [Applied analysis of text data in Python. Machine learning and the creation of natural language processing applications]. Saint Petersburg: Piter, 2019, 368 p.
6. *Luis Pedro Koel'o, Villi Richart.* Postroenie Sistemy mashinnogo obucheniya na yazyke Python [Building a Machine learning System in Python]: transl. from engl. Slinkin A.A. Moscow: DMK Press, 2016, 302 p.

7. *Koncel-Kedziorski R., Hajishirzi H. and Sabharwal A. et al.* Parsing algebraic word problemsinto equations, *Transactions of the Association for Computational Linguistics,* 2015, 3, pp. 585-597.
8. *Devlin J., Chang M.W., Lee K. and Toutanova K.* Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805*, 2018.
9. *Hu K., Wu H. and Qi K. et al.* A domain keyword analysis approach extending Term Frequency-Keyword Active Index with Google Word2Vec model, *Scientometrics*. Springer, 2017, pp. 1-38.
10. *Nalimov V.V.* Veroyatnostnaya model' yazyka. O sootnoshenii estestvennykh i iskusstvennykh yazykov [A probabilistic model of language. On the relationship between natural and artificial languages]. Moscow: Nauka, 1979, 303 p.
11. *Gladkiy A.V.* Sintaksicheskie struktury estestvennogo yazyka v avtomatizirovannykh sistemakh obshcheniya [Syntactic structures of natural language in automated communication systems]. Moscow: Nauka. Glavnaya redaktsiya fiziko-matematicheskoy nauki, 1985, 144 p. (Seriya «Problemy iskusstvennogo intellekta» [Series "Problems of artificial intelligence"]).
12. *Naykhanova L.V., Evdokimova I.S.* Metody i algoritmy translyatsii estestvenno-yazykovykh zaprosov k baze dannykh v SQL-zaprosy: monografiya [Methods and algorithms for translating natural language database queries into SQL queries: monograph]. Ulan-Ude: Izd-vo VSGTU, 2004, 148 p.
13. *Sevbo I.P.* Graficheskoe predstavlenie sintaksicheskikh struktur i stilisticheskaya diagnostika [Graphical representation of syntactic structures and stylistic diagnostics]. Kiev: Naukova dumka, 1983, 192 p.
14. *Rubashkin V.Sh.* Predstavlenie i analiz smysla v intellektual'nykh informatsionnykh sistemakh [Representation and analysis of meaning in intelligent information systems]. Moscow: Nauka. Glavnaya redaktsiya fiziko-matematicheskoy nauki, 1989, 192 p. (Seriya «Problemy iskusstvennogo intellekta» [Series "Problems of artificial intelligence"]).
15. *Aggarwal C.C., Al-Garawi F., Yu P.S.* Intelligent crawling on the world wide web with arbitrary predicates, *In Proc. of the WWW10, May 2001,* pp. 96-105.
16. *Agichtein E., Lawrence S., Gravano L.* Learning search engine specific query transformations for question answering, *In Proc. of the WWW10,* 2001, pp. 169-178.
17. *Vishnyakov Yu.M., Vishnyakov R.Yu.* Vychislitel'naya semanticheskaya interpretatsiya tekstov nauchno-tekhnicheskogo stilya [Computational semantic interpretation of scientific and technical style texts], *Sovremennye naukoemkie tekhnologii* [Modern high-tech technologies], 2016, No. 12–2, pp. 236-242.
18. *Vishnyakov R.Y., Vishnyakov Y.M.* Identification of semantic objects, *Journal of Physics: Conference Series, Bristol May 2021.* – Vol. 1902. DOI: 10.1088/1742–6596/1902/1/012104.
19. *Vishnyakov Yu.M. Vishnyakov R.Yu.* O primenenii vychislitel'noy teorii semanticheskoy interpretatsii k vyyavleniyu kiberprestupleni [On the application of the computational theory of semantic interpretation to the identification of cybercrimes], *Aktual'nye problemy prikladnoy matematiki, informatiki i mekhaniki: Sb. trudov mezhdunarodnoy nauchno-tekhnicheskoy konferentsii, Voronezh, 13–15 dekabrya 2021 g.* [Actual problems of applied mathematics, computer science and mechanics: Proceedings of the International Scientific and Technical Conference, Voronezh, December 13-15, 2021]. Voronezh, 2022, pp. 1526-1530.
20. *Stepanova E.V., Vishnyakov Yu.M.* Programmnoe obespechenie dlya vychisleniya semanticheskoy blizosti tekstov [Software for calculating the semantic proximity of texts], *Prikladnaya matematika: sovremennye problemy matematiki, informatiki i modelirovaniya: Sb. materialov III-y Vserossiyskoy nauchno-prakticheskoy konferentsii molodykh uchenykh* [Applied mathematics: modern problems of mathematics, computer science and modeling: A collection of materials of the III All-Russian Scientific and practical Conference of Young Scientists]. Krasnodar, 2021, pp. 218-223.

Статью рекомендовал к опубликованию д.т.н., профессор В.В. Курейчик.

Вишняков Юрий Муссович – Кубанский государственный университет; e-mail: jury.vishnyakov@gmail.com; г. Краснодар, Россия; тел.: +79996314840; кафедра вычислительных технологий; д.т.н.; профессор.

Вишняков Ренат Юрьевич – e-mail: renat.vishnyakov@mail.ru; тел.: +79996300250; кафедра вычислительной математики и информатики; к.т.н.; доцент.

Vishnyakov Yury Mussovich – Kuban State University; e-mail: jury.vishnyakov@gmail.com; Krasnodar, Russia; phone: +79996314840; the department of Computing Technologies; dr. of eng. sc.; professor

Vishnyakov Renat Yurevich – e-mail: renat.vishnyakov@mail.ru; phone: +79996300250; the department of Computational Mathematics and Computer Science; cand. of eng. sc.; associate professor.

УДК 621.31 DOI 10.18522/2311-3103-2024-4-123-134

И.В. Еманов, Н.Е. Сергеев

ПРОГНОЗИРОВАНИЕ ПОТРЕБЛЕНИЯ ЭЛЕКТРОЭНЕРГИИ ПРОМЫШЛЕННЫХ ПРЕДПРИЯТИЙ (ОБЗОР)

Крупные потребители электроэнергии в основном производят покупку электроэнергии на оптовом рынке электроэнергии и мощности, например промышленные предприятия черной металлургии. Для производства продукции крупные промышленные предприятия ежедневно заказывают почасовые объёмы потребления электроэнергии на двое суток вперёд, при необходимости за сутки, предшествующие дню потребления, предприятия имеют право направлять скорректированные значения. При этом за отклонения от плановых почасовых объёмов предприятия несут дополнительные затраты, которые входят в тариф на электроэнергию. Одними из важнейших факторов, которые влияют на прогнозирование почасового потребления электроэнергии, являются: многообразие видов основного и вспомогательного оборудования, мощностей потребителей электроэнергии, осуществляющих технологические процессы предприятия; интенсивность загрузки производства и режимы работы в зависимости от производства сортамента; частое использования часов максимальной электрической мощности в течение суток; энергоемкое производство продукции. Для построения прогнозных данных временных рядов строится модель прогнозирования почасового потребления электроэнергии промышленным предприятием и имеет большое количество входных данных, которые имеют вероятностную составляющую. Рассмотрение различных методов прогнозирования временных рядов потребления электроэнергии промышленных предприятий представляется актуальной научно-технической задачей. Это обусловлено, требованиями минимизации, во-первых, скачков и сбоев в работе генерирующих мощностей энергосистемы региона, в котором расположено предприятие (т.к. нагрузка, например, предприятий черной металлургии может достигает до 10% от общего потребления региона), а во-вторых, дополнительных затрат, связанные с куплей/продажей объемов электроэнергии потребленной сверх заявки/невыбранной при неточном планировании почасовых объемов потребляемой электроэнергии, которые входят в тариф на электроэнергию.

Методы прогнозирования; краткосрочное прогнозирование электроэнергии; металлургическое предприятие; прогнозирование на сутки вперед; временные ряды.

I.V. Emanov, N.E. Sergeev

FORECASTING ELECTRICITY CONSUMPTION BY INDUSTRIAL ENTERPRISES (REVIEW)

Large consumers of electricity mainly purchase electricity on the wholesale electricity and capacity market, for example, industrial enterprises of ferrous metallurgy. For the production of products, large industrial enterprises daily order hourly volumes of electricity consumption for two days in advance, if necessary, enterprises have the right to send adjusted values for the day preceding the day of consumption. At the same time, for deviations from the planned hourly volumes, enterprises incur additional costs, which are included in the electricity tariff. One of the most important factors that affect the forecasting of hourly electricity consumption are: the variety of types of main and auxiliary equipment, the capacities of electricity consumers carrying out the technological processes of the enterprise; the intensity of production load and operating modes depending on the production of the product range; the frequent use of hours of maximum electric power during the Days; energy-intensive production. To build forecast data for time series, a model is built to predict hourly electricity consumption by an industrial enterprise and has a large number of input data that have a probabilistic component. Consideration of various methods for forecasting time series of electricity consumption of industrial enterprises seems to be an urgent scientific and technical task. This is due to the requirements of minimization, firstly, of jumps and failures in the operation of generating capacities of the energy system of the region in which the enterprise is located (since the load, for example, of ferrous metallurgy enterprises can reach up to 10% of the total consumption of the region), and secondly, additional costs associated with the purchase/sale of volumes of electricity consumed in excess of the application/unused in case of inaccurate planning of hourly volumes of electricity consumed, which are included in the electricity tariff.

Forecasting methods; short-term forecasting of electricity; metallurgical enterprise; day-ahead forecasting; time series.

Введение. На многих предприятиях черной металлургии существует свои методы определения прогноза почасового потребления электроэнергии на сутки вперёд, в основном завязанный на простом определении прогноза потребления электроэнергии в зависимости от загрузки производственных мощностей и проводимых планово-предупредительных ремонтов. При этом отсутствует в целом система помогающая спрогнозировать плановое потребление электроэнергии на сутки вперёд, основывающаяся на различных входных данных.

На металлургических предприятиях потребители электроэнергии, расположены в разных цехах - одни потребители электроэнергии могут влиять значительно на отклонения от плановых почасовых величин потребления, другие не существенно (создают условно постоянную нагрузку).

Наиболее значимыми крупными потребителями электроэнергии металлургического предприятия является электрооборудование:

1. Электросталеплавильного цеха (ЭСПЦ): дуговая сталеплавильная печь, печь-ковш, установка вакуумирования стали, машина непрерывного литья заготовок;

2. Трубопрокатного цеха: прошивной стан литой заготовки для производства бесшовных труб, стан извлекатель, калибровочный стан, редукционно-растяжной стан, термические участки индукционного нагрева труб;

3. Энергетического цеха: двигатели насосных станций оборотного цикла водоснабжения, двигатели водоподготовки для подачи воды для охлаждения оборудования основных производственных цехов, градирни, компрессорные и турбокомпрессорные станции воздухоснабжения потребителей завода.

Рассмотрим основные проблемы с обеспечением прогнозных данных на примере электросталеплавильного цеха металлургического завода.

1. Основные проблемы, связанные с обеспечением точных прогнозных данных потребления электроэнергии в ЭСПЦ. На точность построения прогнозных данных потребления электроэнергии в ЭСПЦ на сутки вперед влияют следующие факторы:

1. Информация по необходимому количеству сортаментов непрерывной литой заготовки и их объёму, которую необходимо выплавить за сутки.
2. Серийность выпуска производства продукции.
3. Начало и окончание проведения планово-предупредительных, капитальных ремонтов.
4. Среднее время плавки и обработки стали.
5. Наличие газокислородных горелок для расплава металла в дуговой сталеплавильной печи.
6. Используемого лома металла для производства стали [2].
7. Количества оборудования: дуговых сталеплавильных печей, печь-ковшей, установок вакуумирования стали, машин непрерывного литья заготовок.

Дополнительными факторами, влияющими на потребление электроэнергии, является качество используемых графитовых электродов, которые влияют на расход электроэнергии на тонну продукции. Чаще всего выбираются электроды в порядке следующего приоритета:

1. Механическая прочность.
2. Удельный расход электродов на тонну продукции.
3. Время плавки металла.
4. Удельный расход электроэнергии на тонну продукции.

Приобретаются и выбираются графитовые электроды, качество которых влияет на время плавки стали в печи по факту: из наличия на рынке продаж (в случае отсутствия выбора), либо какие приобретает коммерческий отдел, исходя из ценовой категории на данный материал. Отсюда можно сделать вывод, что удельный расход электроэнергии на тонну продукции находится не на первом месте.

Погодные факторы не учитываются для определения влияния расхода электроэнергии при производстве стали в ЭСПЦ, так как отсутствует устойчивая значимая зависимость удельного расхода электроэнергии на технологические нужды по основным агре-

гатам ЭСПЦ металлургического предприятия от среднемесячной температуры [1]. Однако на отклонения потребления электроэнергии от заявленных плановых величин может влиять высокая температура окружающей среды в летний период времени. В случае отсутствия достаточного количества охлаждающих мощностей оборудования вспомогательных цехов может приводить к внеплановым аварийным остановкам основных производственных цехов, связанных с перегревом оборудования.

При наличии оборудования в единичном экземпляре дуговой сталеплавильной печи, печь-ковша, потребление данного оборудования в течение суток является очень неравномерным и существует возможность регулирования мощности потребления электроэнергии, путём повышения или понижения напряжения на печных трансформаторах, чтобы сбалансировать величину потребления электроэнергии для минимального отклонения фактического значения потребления электроэнергии от плановой величины. Однако данное мероприятие способно привести к увеличению времени обработки стали и перегреву футеровки ковшей, в которых происходит расплав металла, и как следствие, прожогу ковшей с потерей металла, созданием аварийной остановки всего электросталеплавильного цеха, экономическим убыткам, дополнительным потерям на оплатах отклонений от плановых почасовых потреблений электроэнергии.

Большинство предприятий принимают единственное решение при планировании почасового потребления электроэнергии на сутки вперёд, это фактическая усреднённая мощность часового потребления дуговой сталеплавильной печью и печь-ковшом, а также вспомогательным оборудованием, для обеспечения усреднённых отклонений от часового потребления. При этом за эти отклонения потребитель несет дополнительных затраты, связанные с куплей/продажей объемов электроэнергии, потребленной сверх заявки/невыбранной при неточном планировании почасовых объемов потребляемой электроэнергии, которые входят в тариф на электроэнергию.

Пример распределения отклонений от суточных почасовых заявок металлургическим заводом за месяц по потребителям завода представлен на рис. 1. (в процентном соотношении). Как мы видим, на фоне общего потребления электроэнергии металлургическим заводом, на 1-м месте по отклонениям стоит дуговая сталеплавильная печь ЭСПЦ, на 2-м месте установка «Печь-ковш», отделение непрерывного литья заготовок (ОНЛЗ), вакууматор, на 3-м месте энергоёмкие участки трубопрокатного цеха, где производится бесшовная трубная продукция, её термообработка и отделка.

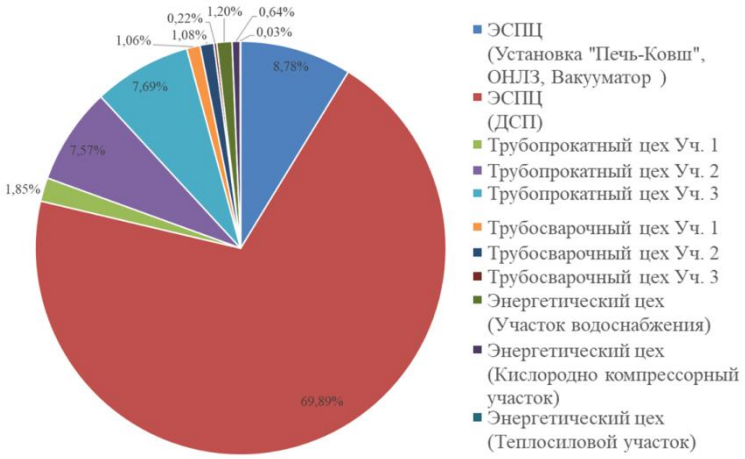


Рис. 1. Распределение отклонений от суточных почасовых заявок по потребителям металлургического завода за месяц

Пример неравномерного потребления электроэнергии ЭСПЦ в течение суток приведен на рис. 2.

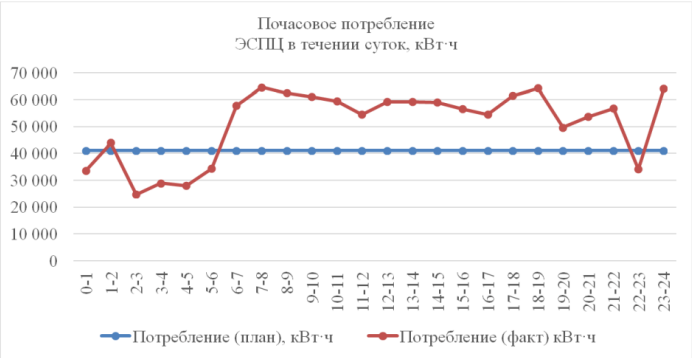


Рис. 2. Почасовое потребление электроэнергии электросталеплавильным цехом в течении суток

2. Методы прогнозирования временных рядов потребления электроэнергии. На сегодняшний день существует много различных методов прогнозирования потребления электроэнергии.

В зависимости от необходимости периода прогнозирования электроэнергии их можно разделить на следующие виды:

1. Долгосрочное – на несколько лет вперед.
2. Среднесрочное – на несколько месяцев вперед.
3. Краткосрочное – на один день или на несколько дней вперед.
4. Сверхкраткосрочное – на несколько минут или часов вперед.

Рассмотрим несколько методов, используемых для прогнозирования временных рядов, в частности потребления электроэнергии, определим их достоинства и недостатки.

Анализ временных рядов. При прогнозировании числовых значений временного ряда важно анализировать прошлые значения временных рядов и находить зависимости для определение прогнозного будущего значения.

Во временном ряде содержится информация об изменениях выбранного параметра в течение времени и могут наблюдаться определённые закономерности и тренды.

Временные ряды можно использовать, например, для определения информации о потреблении электроэнергии в определенный месяц, день, час суток, о производстве объемов энергетических ресурсов на технологические нужды предприятия, для получения информации о температуре окружающей среды в разные сезоны и временные промежутки.

Сам анализ временных рядов могут применять в разных направлениях, например, при производстве продукции, потреблении электроэнергии, изменении температуры и т.д. При анализе временных рядов необходимо правильно выбирать применяемые методы, модели, инструменты для получения корректных исследуемых результатов.

Рассмотрим некоторые методы анализа временных рядов.

Одними из методов анализа временных рядов являются стационарность и автокорреляция.

Стационарность – это свойство временного ряда, при котором, его усредненные и типовые отклонения не изменяются со временем.

Стационарный временной ряд не трудно проанализировать, а затем спрогнозировать будущее значение, но при этом, также есть нестационарные временные ряды, которые могут иметь линию тренда (рост или снижение), цикличность, сезонность (повторение событий). Пример нестационарного временного ряда приведён на рис. 3., где показаны затраты электроэнергии при производстве сжатого воздуха на технологические нужды цехов металлургического завода в течение года по месяцам.

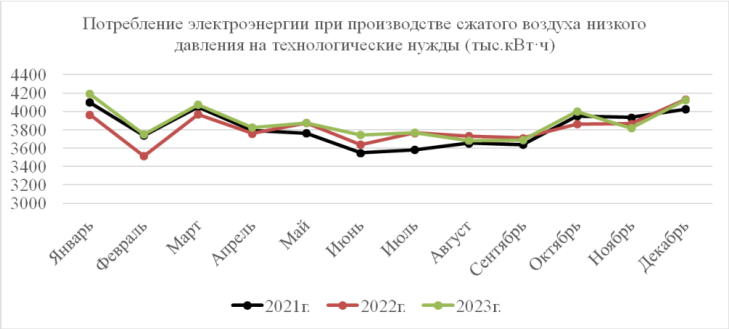


Рис. 3. Потребление электроэнергии при производстве сжатого воздуха низкого давления на технологические нужды по месяцам в течение года

Для того, чтобы проверить временной ряд на стационарность, необходимо его визуализировать с помощью графика автокорреляции (позволяет показать корреляцию между данными ряда в различные периоды времени), частной автокорреляции (позволяет учитывать корреляцию между двумя отдельными значениями, не беря в учёт все промежуточные значения).

Автокорреляция – это статистическая взаимосвязь двух или нескольких случайных величин между значениями ряда в различные периоды времени.

Значимость автокорреляции во временном ряде показывают параметры: тренд, сезонность, цикличность.

Любой метод имеет свои плюсы и минусы, применение конкретного метода зависит от выбранной области исследования.

Рассмотрим моделирование временных рядов на примере моделей ARIMA и SARIMA [3].

Моделирование временных рядов. Построение моделей временных рядов необходимо для прогнозирования будущих значений, это позволяет принимать взвешенные решения в исследуемых областях.

Дж. Боксом и Г. Дженкинсоном были предложены модели авторегрессионной интегрированной скользящей средней моделей (autoregressive integrated moving average – ARIMA).

ARIMA основывается на информации, содержащейся в исторических данных, исследуемых прогнозируемых рядов, что позволяет моделировать не стационарные временные ряды.

Модель ARIMA содержит три параметра: параметр авторегрессии; параметр скользящего среднего; параметр интегрирования.

В качестве примера, использования модели ARIMA может служить прогнозирование помесячного потребления электроэнергии на производство сжатого воздуха на технологические нужды на основе прошлых ежемесячных данных за предыдущие годы как показано на рис. 3. Если прошлые данные имеют цикличность или тренд, сезонность, то можно использовать модель ARIMA для учета обозначенных факторов в прогнозировании потребления на следующий год. Как видно из рис. 3 потребление электроэнергии меняется циклично из года в год, с небольшим приростом по потреблению электроэнергии в 2023 году по отношению к 2021 и 2022 годам.

Рассмотрим потребление электроэнергии трубопрокатным цехом (ТПЦ) участком №3 в течение недели на рис. 4. Как видно из диаграммы на рис. 4 почасовое потребление электроэнергии имеет неоднородную структуру по участку и зависит от разных факторов, которые необходимо учитывать.

Колебания величины потребления электроэнергии, показывают, что временной ряд неоднороден, имеет сложную практическую реализацию прогнозной модели, это свидетельствует о недостатке метода ARIMA. Также ARIMA свойственны недостатки корреляционных методов.

Рис. 4. Почасовое потребление электроэнергии в течении недели ТПЦ участком №3

Недостатками данного класса моделей являются: большое число параметров модели, идентификация которых неоднозначна и ресурсоемка, требует большого опыта со стороны исследователя; низкая адаптивность моделей (при получении новых вводных данных, модель нужно периодически переоценивать, а иногда и заново идентифицировать), а также линейность и, как следствие, отсутствие способности моделирования нелинейных процессов, часто встречающихся на практике [3].

Также к основному недостатку модели ARIMA относится требование к рядам данных для выполнения построения адекватной модели, где требуется не менее 40 наблюдений, а для следующей, далее рассматриваемой модели SARIMA – не менее 6–10 сезонов [3, 4], что не всегда возможно.

Однако, модель ARIMA очень популярна и часто используется для прогнозирования значений временных рядов потребления электроэнергии [5, 6, 11, 26], а также используется в сравнении с другими методами [7, 8, 10, 21].

Следующей моделью развития включающую в себя сезонность, является SARIMA, которая используется для моделирования поведения временных рядов в разные сезоны.

Модель SARIMA, это расширенная версия ARIMA с тремя дополнительными параметрами, которые определяют сезонность: период сезонности; параметр авторегрессии с сезонностью; параметр скользящего среднего с сезонностью.

SARIMA объединяет в себе методы авторегрессии, интегрированный скользящий среднего, сезонности. Объединение нескольких методов позволяет спрогнозировать будущие значения временных рядов с учетом прошлых сезонных данных (паттернов). Так, например, для прогнозирования временных рядов потребления электроэнергии компрессорным участком по производству сжатого воздуха на технологические нужды, модель SARIMA, позволит спрогнозировать месячное потребление как указано на рис.3, а также при анализе почасового потребления сжатого воздуха в течение суток, позволит определить сколько электроэнергии тратится почасово в течение рабочего дня.

Опыт других исследователей показывает, что метод SARIMA имеет периодическое усовершенствование [7, 9], применяется, например, для определения прогнозных данных потребления электроэнергии на долгосрочный период. Данный метод может применяться в комбинации с другими методами [8], которые уже создают гибридные модели.

Прогнозирование временных рядов. Для прогнозирования временных рядов потребления электроэнергии используются различные методы как по отдельности, так и в комбинации, учитываются различные входные данные.

Рассмотрим некоторые методы прогнозирования временных радов потребления электроэнергии, например методы экстраполяции и методы машинного обучения.

Методы экстраполяции – это методы, которые используют для прогнозирования будущих значений временного ряда, исходя из исторических значений временных рядов предыдущих периодов.

Методы экстраполяции включают в себя:

1. Метод скользящего среднего.
2. Экспоненциальное сглаживание.
3. Метод Хольта-Винтерса.

Метод скользящего среднего – это метод вычисления среднего значения ряда за определенный период для прогнозирования будущих значений временного ряда.

Экспоненциальное сглаживание – это метод, в котором используется экспоненциальная функция для прогнозирования будущих значений временных рядов.

Метод Хольта-Винтерса – это метод, в котором к модели экспоненциального сглаживания добавляется тренд и сезонность [22, 23].

Методы машинного обучения – это методы, в которых используют модели обучения для прогнозирования временных рядов, где вводными данными являются исторические данные прошлых значений временного ряда.

При использовании различных методов, алгоритмов, важным качеством является уменьшение погрешности возникающих ошибок, и установлении границ возникающей ошибки, для понимания корректности работы разработанных методов и алгоритмов [21].

Многие исследователи используют для прогнозов электроэнергии метод машинного обучения [12–14, 20]. В результате применения данного метода исследователи получили: модель машинного обучения на алгоритме градиентного бустинга библиотеки CatBoost; модель регрессии на алгоритме случайного леса Random Forest; модель принятия управленческих решений при формировании заявок на производство и потребление электроэнергии.

Рассмотрим некоторые методы машинного обучения, такие как регрессия и нейронные сети.

Регрессия – метод машинного обучения, в котором моделируется взаимосвязь между различными независимыми (входными) переменными данными и зависимыми (выходными) переменными данными [12].

При прогнозировании временных рядов независимые переменные значения ряда будут находиться в прошлых периодах, а зависимые переменные данные в будущих значениях ряда.

В метод машинного обучения входят в том числе построенные модели на нейронных сетях.

Нейронные сети – являются более сложными методами машинного обучения, в которых используются алгоритмы для обработки информации, похожих, как используются в человеческом мозге. По сравнению с методами экстраполяции, построенные алгоритмы на нейронных сетях, обычно имеют более высокую точность при прогнозировании временных рядов.

С учётом стремительного развития нейронных сетей в последние десятилетия наибольшее развитие получают методы машинного обучения, основанные на применении нейронных сетей для создания моделей прогнозирования потребления электроэнергии [15–20, 24, 25, 27].

Для методов, использующих в комбинации нейронные сети, возможно применение ансамблей нейронных сетей для улучшения точности прогнозирования, а также важна скорость переобучения нейронных сетей [27].

Заключение. Одним из перспективных направлений, является рассмотрение существующих методов и алгоритмов прогнозирования временных рядов потребления электроэнергии для улучшения точности краткосрочного почасового прогнозирования на металлургических предприятиях.

Рассматриваемые методы прогнозировании временных рядов имеют преимущества и недостатки.

Преимущества методов прогнозирования и анализа временных рядов.

Методы прогнозирования позволяют:

1. Получать будущее прогнозное значение временного ряда.

2. Анализировать корреляцию между выбранными данными.

3. Определять закономерность влияния разных входных данных на изменение будущих значений временного ряда.

Методы прогнозирования и анализа временных рядов, имеют недостатки:

1. Точность прогнозных значений временных рядов, может ограничить малое количество входных данных, что показывает необходимость учёта множества входных данных для выявления различных зависимостей при рассмотрении методов прогнозирования временных рядов.

2. Наличие случайных факторов и незапланированных событий могут привести к неверному прогнозу несмотря на то, что выбранные исторические данные позволили бы относительно точно составить прогноз, что требует разработки системы, метода учитывающий данные факторы.

3. Отсутствие данных, связанных с сезонностью, может повлиять на точность прогнозных значений временных рядов, что требует учета в методах прогнозирования введения сезонных коэффициентов для улучшения точности прогнозирования.

При использовании существующих методов и моделей для прогнозирования электроэнергии, необходимо обращать внимание на возможность снижения погрешности при их использовании как по отдельности, так и в комбинированных вариантах.

Для снижения стоимости оплат отклонений от заявленного почасового потребления электроэнергии, необходимо также учитывать динамику изменений почасовой стоимости электроэнергии в течение каждых суток, для выбора оптимального заказа почасового потребления на сутки вперёд.

Целью моего дальнейшего исследования, является проведение системного анализа существующих методов и алгоритмов прогнозирования потребления электроэнергии и их применимость для металлургических предприятий на примере одного из металлургических заводов. Данные исследования будут применены для решения следующей задачи - повышение точности краткосрочного прогнозирования почасового потребления электроэнергии на сутки вперед металлургических предприятий для уменьшения процента отклонения прогнозных данных от существующих фактических.

Данное исследование состоит в том, что проблема со значительными отклонениями почасовых прогнозных от фактических данных в потреблении электроэнергии металлургических предприятий до сих пор не решена и является актуальной по сей день. Разработка концепции реализации прогнозирования потребления электроэнергии промышленными предприятиями, а также применение существующих или новых методов прогнозирования потребления электроэнергии, будет интересна каждому крупному потребителю электроэнергии и других отраслей, имеющих схожие проблемы.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Ефременкова Т.И.* Исследование расхода электроэнергии на технологические цели в электросталеплавильном цехе металлургического предприятия методами корреляционно-регрессионного анализа // Современные технологии управления. – 2017. – № 6 (78). – ISSN2223-9339. – Режим доступа: https://sovman.ru/article/7801/.
2. *Мысик В.Ф., Жданов А.В.* Ресурсы и подготовка лома к плавке стали. – 2-е изд. испр. и допол. – Екатеринбург, 2017. – 337 с.
3. *Таскина А.О.* Прогнозирование цены на электроэнергию на рынке «на сутки вперёд» во второй ценовой зоне. – 2020. – Режим доступа: https://vital.lib.tsu.ru/vital/access/services/Download/vital:11747/SOURCE01.
4. *Ханк Д.Э., Уичери Д.У., Райте А.Дж.* Бизнес-прогнозирование: пер. с англ. – 7-е изд. – М.: Вильямс, 2003. – 506 с.
5. *Alberg D., Last M.* Short-term load forecasting in smart meters with sliding window-based ARIMA algorithms // Vietnam J. Comput. Sci. – 2018. – 5. – P. 241-249. – Режим доступа: https://doi.org/10.1007/s40595-018-0119-7.
6. *Rehan Jamil*. Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply-demand analysis for the year 2030 // Renewable Energy. – 2020. – Vol. 154. – P. 1-10. – ISSN 0960-1481. – https://doi.org/10.1016/j.renene.2020.02.117.

7. *Yuanyuan Wang, Jianzhou Wang, Ge Zhao, Yao Dong.* Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China // Energy Policy. – 2012. – Vol. 48. – P. 284-294. – ISSN 0301-4215. – https://doi.org/10.1016/j.enpol.2012.05.026.
8. *Xinbin Liang, Siliang Chen, Xu Zhu, Xinqiao Jin, Zhimin Du.* Domain knowledge decomposition of building energy consumption and a hybrid data-driven model for 24-h ahead predictions // Applied Energy. – 2023. – Vol. 344, 121244. – ISSN 0306-2619. – https://doi.org/10.1016/j.apenergy.2023.121244.
9. *Mehmet Bilgili, Engin Pinar.* Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye // Energy. – 2023. – Vol. 284, 128575. – ISSN 0360-5442. – https://doi.org/10.1016/j.energy.2023.128575.
10. *Русcков О.В.* Методы прогнозирования неравномерных нестационарных временных рядов на примере задачи планирования электропотребления промышленного предприятия. – 2023. – Режим доступа: https://elib.spbstu.ru/dl/2/r23-69.pdf/download/r23-69.pdf.
11. *Nti I.K., Teimeh M., Nyarko-Boateng O. et al.* Electricity load forecasting: a systematic review // Journal of Electrical Systems and Inf. Technol. – 2020. – 7, 13. – Режим доступа: https://doi.org/10.1186/s43067-020-00021-8.
12. *Моргоева А.Д.* Прогнозирование потребления электрической энергии промышленным предприятием с помощью методов машинного обучения // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2022. – Т. 333, № 7. – С. 115-125. – Режим доступа: https://izvestiya.tpu.ru/archive/article/view/3527/2751.
13. *Моргоева А.Д., Моргоев И.Д., Клюев Р.В., Хетагуров В.Н., Гаврина О.А.* Краткосрочное прогнозирование электропотребления обогатительной фабрики // Горный информационно-аналитический бюллетень. – 2023. – 5-1. – С. 157-169. – Режим доступа: https://www.giab-online.ru/files/Data/2023/5/05_2023_157-169.pdf.
14. *Клюев Р.В., Моргоева А.Д., Гаврина О.А., Босиков И.И., Моргоев И.Д.* Прогнозирование планового потребления электроэнергии для объединенной энергосистемы с помощью машинного обучения // Записки Горного института. – 2023. – № 261. – Режим доступа: https://cyberleninka.ru/article/n/prognozirovanie-planovogo-potrebleniya-elektroenergii-dlya-obedinennoy-energosistemy-s-pomoschyu-mashinnogo-obucheniya.
15. *Вялкова С.А.* Краткосрочное прогнозирование электропотребления Мегаполиса на основе ортогональных разложений и нейронных сетей. – 2022. – Режим доступа: https://www.npi-tu.ru/upload/medialibrary/f39/hxsy86lrce075msy4mt96l1nvbqz4ux8/avtoreferat_vialkova.pdf?ysclid=lue3pm3md4950760747.
16. *Староверов Б.А.* Теоретическое обоснование и практическая реализация систем прогнозирования электропотребления на основе ансамблей нейронных сетей в масштабе региональной сетевой компании. – 2023. – Режим доступа: https://www.dissercat.com/content/teoreticheskoe-obosnovanie-i-prakticheskaya-realizatsiya-sistem-prognozirovaniya-elektropotr.
17. *Алексеева И.Ю.* Краткосрочное прогнозирование электропотребления в электроэнергетических системах с использованием искусственных нейронных сетей. – https://www.dissercat.com/content/kratkosrochnoe-prognozirovanie-elektropotrebleniya-v-elektroenergeticheskikh-sistemakh-s-isp.
18. *Zhen Shao, Qingru Zheng, Chen Liu, Shuangyan Gao, Gang Wang, Yan Chu.* A feature extraction- and ranking-based framework for electricity spot price forecasting using a hybrid deep neural network // Electric Power Systems Research. – 2021. – Vol. 200, 107453. – ISSN 0378-7796. – https://doi.org/10.1016/j.epsr.2021.107453.
19. *Бажинов А.Н.* Метод прогнозирования объемов потребления электроэнергии предприятием черной металлургии на основе нейро-нечетких алгоритмов. – https://www.dissercat.com/ content/metod-prognozirovaniya-obemov-potrebleniya-elektroenergii-predpriyatiem-chernoi-metallurgii-.
20. *Хомутов С.О., Сташко В.И., Серебряков Н.А.* Повышение точности краткосрочного прогнозирования электропотребления групп точек поставки электроэнергии гарантирующих поставщиков второго уровня // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2020. – Т. 331, № 6. – С. 128-140. – Режим доступа: https://doi.org/10.18799/24131830/2020/6/2682.
21. *Chuang Li, Guojie Li, Keyou Wang, Bei Han.* Short-term electrical load forecasting based on error correction using dynamic mode decomposition // Applied Energy. – March 2020. – Vol. 261, 114368. – Режим доступа: https://www.sciencedirect.com/science/article/abs/pii/S0306261919320550?via%3Dihub.
22. *Taylor J.W., McSharry P.E.* Short-term load forecasting methods: An evaluation based on European data // IEEE Transactions on Power Systems. – November 2007. – Vol. 22, Issue 4. – P. 2213-2219. – ISSN 08858950. – https://www.scopus.com/record/display.uri?eid=2-s2.0-36348992080&origin=inward&txGid=21e6fb2c3300889b3ff56d705b52239d.

23. *Phillip G. Gould, Anne B. Koehler, J. Keith Ord, Ralph D. Snyder, Rob J. Hyndman, Farshid Vahid-Araghi.* Forecasting time series with multiple seasonal patterns // European Journal of Operational Research. – 2008. – Vol. 191, Issue 1. – P. 207-222. – ISSN 0377-2217, https://doi.org/10.1016/j.ejor.2007.08.024.
24. *Короткевич М.А., Курачинский В.В.* Прогнозирование электрической нагрузки энергосистемы на следующие сутки с использованием метода искусственных нейронных сетей // Энергетика. Известия высших учебных заведений и энергетических объединений СНГ. – 2010. – № 2. – URL: https://cyberleninka.ru/article/n/prognozirovanie-elektricheskoy-nagruzki-energosistemy-na-sleduyuschie-sutki-s-ispolzovaniem-metoda-iskusstvennyh-neyronnyh-setey.
25. *Чернецов В.И., Казаковский Е.Н.* Прогнозирование потребления электрической энергии с использованием нейронных сетей // НиКа. – 2006. – URL: https://cyberleninka.ru/article/n/prognozirovanie-potrebleniya-elektricheskoy-energii-s-ispolzovaniem-neyronnyh-setey.
26. *Jinliang Zhang, Yi-Ming Wei, Dezhi Li, Zhongfu Tan, Jianhua Zhou.* Short term electricity load forecasting using a hybrid model // Energy. – 2018. – Vol. 158. – P. 774-781. – ISSN 0360-5442, https://doi.org/10.1016/j.energy.2018.06.012.
27. *Староверов Б.А.* Теоретическое обоснование и практическая реализация систем прогнозирования электропотребления на основе ансамблей нейронных сетей в масштабе региональной сетевой компании. – 2023. – Режим доступа: https://www.dissercat.com/content/teoreticheskoe-obosnovanie-i-prakticheskaya-realizatsiya-sistem-prognozirovaniya-elektropotr.

REFERENCES

1. *Efremenkova T.I.* Issledovanie raskhoda elektroenergii na tekhnologicheskie tseli v elektrostaleplavil'nom tsekhe metallurgicheskogo predpriyatiya metodami korrelyatsionno-regressionnogo analiza [Study of electricity consumption for technological purposes in the electric steelmaking shop of a metallurgical enterprise using correlation and regression analysis methods], *Sovremennye tekhnologii upravleniya* [Modern control technologies], 2017, No 6 (78). ISSN2223-9339. Available at: https://sovman.ru/article/7801/.
2. *Mysik V.F., Zhdanov A.V.* Resursy i podgotovka loma k plavke stali [Resources and preparation of scrap for steel smelting]. 2nd ed. Ekaterinburg, 2017, 337 p.
3. *Taskina A.O.* Prognozirovanie tseny na elektroenergiyu na rynke «na sutki vpered» vo vtoroy tsenovoy zone [Forecasting the price of electricity on the day-ahead market in the second price zone], 2020. Available at: https://vital.lib.tsu.ru/vital/access/services/Download/vital:11747/SOURCE01.
4. *Khank D.E., Uicheri D.U., Rayte A.Dzh.* Biznes-prognozirovanie [Business forecasting]: transl. from engl. 7th ed. Moscow: Vil'yams, 2003, 506 p.
5. *Alberg D., Last M.* Short-term load forecasting in smart meters with sliding window-based ARIMA algorithms, *Vietnam J. Comput. Sci.*, 2018, 5, pp. 241-249. Available at: https://doi.org/10.1007/s40595-018-0119-7.
6. *Rehan Jamil*. Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply-demand analysis for the year 2030, *Renewable Energy,* 2020, Vol. 154, pp. 1-10. ISSN 0960-1481. Available at: https://doi.org/10.1016/j.renene.2020.02.117.
7. *Yuanyuan Wang, Jianzhou Wang, Ge Zhao, Yao Dong.* Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China, *Energy Policy,* 2012, Vol. 48, pp. 284-294. ISSN 0301-4215. Available at: https://doi.org/10.1016/j.enpol.2012.05.026.
8. *Xinbin Liang, Siliang Chen, Xu Zhu, Xinqiao Jin, Zhimin Du.* Domain knowledge decomposition of building energy consumption and a hybrid data-driven model for 24-h ahead predictions, *Applied Energy,* 2023, Vol. 344, 121244. ISSN 0306-2619. Available at: https://doi.org/10.1016/j.apenergy.2023.121244.
9. *Mehmet Bilgili, Engin Pinar.* Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye, *Energy,* 2023, Vol. 284, 128575. ISSN 0360-5442. Available at: https://doi.org/10.1016/j.energy.2023.128575.
10. *Russkov O.V.* Metody prognozirovaniya neravnomernykh nestatsionarnykh vremennykh ryadov na primere zadachi planirovaniya elektropotrebleniya promyshlennogo predpriyatiya [Methods for forecasting uneven non-stationary time series using the example of the problem of planning electricity consumption of an industrial enterprise], 2023. Available at: https://elib.spbstu.ru/dl/2/r23-69.pdf/download/r23-69.pdf.
11. *Nti I.K., Teimeh M., Nyarko-Boateng O. et al.* Electricity load forecasting: a systematic review, *Journal of Electrical Systems and Inf. Technol*., 2020, 7, 13. Available at: https://doi.org/10.1186/s43067-020-00021-8.
12. *Morgoeva A.D.* Prognozirovanie potrebleniya elektricheskoy energii promyshlennym predpriyatiem s pomoshch'yu metodov mashinnogo obucheniya [Forecasting the consumption of electrical energy by an industrial enterprise using machine learning methods], *Izvestiya Tomskogo politekhnicheskogo*

universiteta. Inzhiniring georesursov [Bulletin of Tomsk Polytechnic University. Georesources engineering], 2022, Vol. 333, No. 7, pp. 115-125. Available at: https://izvestiya.tpu.ru/archive/article/view/3527/2751.

13. *Morgoeva A.D., Morgoev I.D., Klyuev R.V., Khetagurov V.N., Gavrina O.A.* Kratkosrochnoe prognozirovanie elektropotrebleniya obogatitel'noy fabriki [Short-term forecasting of electricity consumption of the processing plant], *Gornyy informatsionno-analiticheskiy byulleten'* [Mining information and analytical bulletin], 2023, 5-1, pp. 157-169. Available at: https://www.giab-online.ru/files/Data/2023/5/05_2023_157-169.pdf.
14. *Klyuev R.V., Morgoeva A.D., Gavrina O.A., Bosikov I.I., Morgoev I.D.* Prognozirovanie planovogo potrebleniya elektroenergii dlya ob"edinennoy energosistemy s pomoshch'yu mashinnogo obucheniya [Forecasting the planned electricity consumption for the unified power system using machine learning], *Zapiski Gornogo institute* [Notes of the Mining Institute], 2023, No. 261. Available at: https://cyberleninka.ru/article/n/prognozirovanie-planovogo-potrebleniya-elektroenergii-dlya-obedinennoy-energosistemy-s-pomoschyu-mashinnogo-obucheniya.
15. *Vyalkova S.A.* Kratkosrochnoe prognozirovanie elektropotrebleniya Megapolisa na osnove ortogonal'nykh razlozheniy i neyronnykh setey [Short-term forecasting of electricity consumption of a megalopolis based on orthogonal decompositions and neural networks], 2022. Available at: https://www.npi-tu.ru/upload/medialibrary/f39/hxsy86lrce075msy4mt96l1nvbqz4ux8/avtoreferat_vialkova.pdf?ysclid=lue3pm3md4950760747.
16. *Staroverov B.A.* Teoreticheskoe obosnovanie i prakticheskaya realizatsiya sistem prognozirovaniya elektropotrebleniya na osnove ansambley neyronnykh setey v masshtabe regional'noy setevoy kompanii [Theoretical justification and practical implementation of power consumption forecasting systems based on neural network ensembles on the scale of a regional grid company], 2023. Available at: https://www.dissercat.com/content/teoreticheskoe-obosnovanie-i-prakticheskaya-realizatsiya-sistem-prognozirovaniya-elektropotr.
17. *Alekseeva I.Yu.* Kratkosrochnoe prognozirovanie elektropotrebleniya v elektroenergeticheskikh sistemakh s ispol'zovaniem iskusstvennykh neyronnykh setey [Short-term forecasting of electricity consumption in electric power systems using artificial neural networks]. Available at: https://www.dissercat.com/content/kratkosrochnoe-prognozirovanie-elektropotrebleniya-v-elektroenergeticheskikh-sistemakh-s-isp.
18. *Zhen Shao, Qingru Zheng, Chen Liu, Shuangyan Gao, Gang Wang, Yan Chu.* A feature extraction- and ranking-based framework for electricity spot price forecasting using a hybrid deep neural network, *Electric Power Systems Research,* 2021, Vol. 200, 107453. ISSN 0378-7796. Available at: https://doi.org/10.1016/j.epsr.2021.107453.
19. *Bazhinov A.N.* Metod prognozirovaniya ob"emov potrebleniya elektroenergii predpriyatiem chernoy metallurgii na osnove neyro-nechetkikh algoritmov [Method for forecasting electricity consumption volumes by a ferrous metallurgy enterprise based on neuro-fuzzy algorithms]. Available at: https://www.dissercat.com/content/metod-prognozirovaniya-obemov-potrebleniya-elektroenergii-predpriyatiem-chernoi-metallurgii-.
20. *Khomutov S.O., Stashko V.I., Serebryakov N.A.* Povyshenie tochnosti kratkosrochnogo prognozirovaniya elektropotrebleniya grupp tochek postavki elektroenergii garantiruyushchikh postavshchikov vtorogo urovnya [Improving the accuracy of short-term forecasting of electricity consumption of groups of electricity supply points of second-tier guaranteed suppliers], *Izvestiya Tomskogo politekhnicheskogo universiteta. Inzhiniring georesursov* [Bulletin of Tomsk Polytechnic University. Georesources engineering], 2020, Vol. 331, No. 6, pp. 128-140. Available at: https://doi.org/10.18799/24131830/2020/6/2682.
21. *Chuang Li, Guojie Li, Keyou Wang, Bei Han.* Short-term electrical load forecasting based on error correction using dynamic mode decomposition, *Applied Energy,* March 2020, Vol. 261, 114368. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0306261919320550? via%3Dihub.
22. *Taylor J.W., McSharry P.E.* Short-term load forecasting methods: An evaluation based on European data, *IEEE Transactions on Power Systems,* November 2007, Vol. 22, Issue 4, pp. 2213-2219. ISSN 08858950. Available at: https://www.scopus.com/record/display.uri?eid=2-s2.0-36348992080&origin=inward&txGid=21e6fb2c3300889b3ff56d705b52239d.
23. *Phillip G. Gould, Anne B. Koehler, J. Keith Ord, Ralph D. Snyder, Rob J. Hyndman, Farshid Vahid-Araghi.* Forecasting time series with multiple seasonal patterns, *European Journal of Operational Research,* 2008, Vol. 191, Issue 1, pp. 207-222. ISSN 0377-2217, Available at: https://doi.org/10.1016/j.ejor.2007.08.024.
24. *Korotkevich M.A., Kurachinskiy V.V.* Prognozirovanie elektricheskoy nagruzki energosistemy na sleduyushchie sutki s ispol'zovaniem metoda iskusstvennykh neyronnykh setey [Forecasting the electrical load of the power system for the next day using the method of artificial neural networks], *Energetika. Izvestiya vysshikh uchebnykh zavedeniy i energeticheskikh ob"edineniy SNG* [Power Engineering. News of higher educational institutions and energy associations of the CIS], 2010, No. 2.

Available at: https://cyberleninka.ru/article/n/prognozirovanie-elektricheskoy-nagruzki-energosistemy-na-sleduyuschie-sutki-s-ispolzovaniem-metoda-iskusstvennyh-neyronnyh-setey.
25. *Chernetsov V.I., Kazakovskiy E.N.* Prognozirovanie potrebleniya elektricheskoy energii s ispol'zovaniem neyronnykh setey [Forecasting the consumption of electrical energy using neural networks], *NiKa* [NiKa], 2006. Available at: https://cyberleninka.ru/article/n/prognozirovanie-potrebleniya-elektricheskoy-energii-s-ispolzovaniem-neyronnyh-setey.
26. *Jinliang Zhang, Yi-Ming Wei, Dezhi Li, Zhongfu Tan, Jianhua Zhou.* Short term electricity load forecasting using a hybrid model, *Energy,* 2018, Vol. 158, pp. 774-781. ISSN 0360-5442, https://doi.org/10.1016/j.energy.2018.06.012.
27. *Staroverov B.A.* Teoreticheskoe obosnovanie i prakticheskaya realizatsiya sistem prognozirovaniya elektropotrebleniya na osnove ansambley neyronnykh setey v masshtabe regional'noy setevoy kompanii [Theoretical justification and practical implementation of electricity consumption forecasting systems based on neural network ensembles on the scale of a regional grid company], 2023. Available at: https://www.dissercat.com/content/teoreticheskoe-obosnovanie-i-prakticheskaya-realizatsiya-sistem-prognozirovaniya-elektropotr.

Статью рекомендовал к опубликованию д.т.н., профессор В.В. Курейчик.

Еманов Иван Владимирович – Южный федеральный университет; e-mail: emanov@sfedu.ru; тел.: +79604685348; г. Таганрог; Россия; кафедра вычислительной техники; аспирант.

Сергеев Николай Евгеньевич – e-mail: nesergeev@sfedu.ru; тел.: +79001278025; кафедра вычислительной техники; д.т.н.; доцент; профессор.

Emanov Ivan Vladimirovich – South Federal University; e-mail: emanov@sfedu.ru; phone: +79604685348; Taganrog; Russia; the department of Computer Science; graduate student.

Sergeev Nikolay Evgenievich – e-mail: nesergeev@sfedu.ru; phone: +79001278025; the department of Computer Science; dr. of eng. sc.; assistant professor; professor.

УДК 004.942 DOI 10.18522/2311-3103-2024-4-134-144

Л.Э. Хайруллина, З.Н. Хакимов, Г.З. Хабибуллина

ПРЕДВАРИТЕЛЬНАЯ ВЕЙВЛЕТ-ОБРАБОТКА РЯДОВ ФИНАНСОВЫХ ДАННЫХ В СИСТЕМЕ WOLFRAM MATHEMATICA

Любой временной ряд представляет собой сочетание полезной информации и шумов. Поэтому в анализе финансовых временных рядов одним из ключевых моментов является предварительная обработка данных с целью уменьшения шумовой компоненты. Одним из перспективных способов очистки временного ряда является трешолдинг – разложение сигнала на вейвлет-спектр до заданного уровня, обнуление тех вейвлет-коэффициентов разложения, значения которых меньше определенного порогового значения, и последующая вейвлет-реконструкция сигнала по аппроксимирующим и очищенным на каждом уровне детализирующим коэффициентам. Трешолдинг проводится с применением современных программных средств, среди которых исследователи чаще всего отдают предпочтение среде Matlab. В данной работе представлена демонстрация возможностей системы компьютерной математики Wolfram Mathematica в предварительной обработке финансовых данных. Wolfram Mathematica обладает мощным функционалом, позволяющим проводить качественную обработку временных рядов. Система содержит большую коллекцию семейств вейвлетов, множественные варианты дискретных и непрерывных вейвлет-преобразований. В качестве объекта исследования была выбрана история ежедневных котировок акций Сбербанка за последние 3 года. Анализ полученных результатов показал, что на качество очистки сигнала влияет выбор базисного вейвлета – в нашем случае предпочтительным оказалось использование вейвлета Добеши 6-го порядка. Максимальное отношение сигнал/шум достигается при жесткой пороговой обработке с порогом «SURELevel». Проведенные исследования показали, что вейвлет-трешолдинг над детализирующими коэффициентами вейвлет-разложения является эффективным методом подавления выбросов и флуктуаций временного ряда. Очищенный сигнал повторяет форму исходного сигнала, все пики хорошо выражены. При этом в краткосрочном прогнозе получаются более точные прогнозные значения.

Финансовый временной ряд; шум; трешолдинг; пороговая обработка; вейвлет-анализ; вейвлет Добеши; Wolfram Mathematica.

L.E. Khairullina, Z.N. Khakimov, G.Z. Khabibullina

PRELIMINARY WAVELET PROCESSING OF FINANCIAL DATA SERIES IN THE WOLFRAM MATHEMATICA SYSTEM

Any time series is a combination of useful information and noise. Therefore, in the analysis of financial time series, one of the key points is the preprocessing of data in order to reduce the noise component. One of the promising ways to clean up the time series is threading – decomposing the signal into a wavelet spectrum to a given level, zeroing out those wavelet decomposition coefficients whose values are less than a certain threshold value, and subsequent wavelet reconstruction of the signal using approximating and refined detailing coefficients at each level. Tresholding is carried out using modern software tools, among which researchers most often prefer the Matlab environment. This paper presents a demonstration of the capabilities of the Wolfram Mathematica computer mathematics system in the preliminary processing of financial data. Wolfram Mathematica has powerful functionality that allows high-quality processing of time series. The system contains a large collection of wavelet families, multiple variants of discrete and continuous wavelet transformations. The history of Sberbank's daily stock quotes over the past 3 years was chosen as the object of the study. An analysis of the results showed that the quality of signal purification is influenced by the choice of a basic wavelet – in our case, the use of a 6th-order Daubechies wavelet turned out to be preferable. The maximum signal-to-noise ratio is achieved with rigid threshold processing with a "SURELevel" threshold. The conducted studies have shown that wavelet tresholding over the detailing coefficients of the wavelet decomposition is an effective method of suppressing outliers and fluctuations of the time series. The cleared signal repeats the shape of the original signal, all peaks are well expressed. At the same time, more accurate forecast values are obtained in the short-term forecast

Financial time series; noise; thresholding; threshold processing; wavelet analysis; Daubechies wavelet; Wolfram Mathematica.

Введение. Анализ временных рядов финансовых данных является сложной задачей из-за их нелинейности и нестационарности. Нестационарность выражается в изменчивости статистических характеристик ряда (математического ожидания, дисперсии) с течением времени и проявляется в виде тренда, цикличности, сезонности, случайном блуждании или их сочетании. Нестационарные сигналы чаще всего состоят из кратковременных высокочастотных и длительных низкочастотных компонентов, поэтому для анализа таких сигналов предпочтительнее использовать вейвлет-анализ [1]. He&Nguen [2] выделяют следующие особенности вейвлет-анализа применительно к временным рядам: 1) для вейвлет-анализа не требуется предположения о стационарности данных, 2) информация предоставляется как во временной, так и в частотной области, 3) вейвлет-анализ дает возможность находить разрывы и сдвиги в рядах данных. Подробный обзор возможностей дискретного вейвлет-анализа приведен в работе Chaovalit P., Gangopadhyay A. и др. [3]. В ней авторы систематизировали научные работы по применению вейвлет-анализа в разных областях, привели примеры прикладного применения этого математического аппарата и указали возможные будущие направления развития этой области.

Любой временной ряд представляет собой сочетание полезной информации и шумов. Значительный ряд работ посвящены описанию методов и технологий очистки временного ряда от шумов. Так, в коллективной работе Кропотова Ю.А. с соавторами [4] разработана математическая модель фильтрации шумов на основе пороговой обработки детализирующих коэффициентов вейвлет-разложения. Тем самым авторы добиваются увеличения отношения сигнал/шум и, соответственно, повышения точности представления данных в восстановленном ряду. В статье Белова А.А. [5] проведен анализ критериев при выборе пороговых уровней вейвлет-преобразования для сглаживания рядов данных. В работе Taranenko Y.K. [6] предложен алгоритм, позволяющий определить наиболее эффективные методы дискретной вейвлет-фильтрации в зависимости от характера измеряемого сигнала.

Отфильтровывая шум, аналитики могут улучшить прогнозирующую силу своих моделей. В коллективной работе Shaikh W. и др. [7] алгоритм вейвлет-декомпозиции объединен с различными традиционными моделями прогнозирования временных рядов, такими как метод опорных векторов, регрессионные модели, нейронные сети. Авторы де-

лают вывод, что такое комбинирование дает возможность получить лучшие прогнозные значения. Аналогичный вывод постулируется и в работе Lee H. [8] – вейвлет-анализ является эффективным инструментом снижения уровня шума при анализе финансовых временных рядов и последующем прогнозировании с помощью модели ARIMA. В статьях Benrhmach G. и др. [9], Boubaker H. и др. [10] разработаны гибридные методы моделирования финансового временного ряда, основанные на дискретном вейвлет-разложении ряда и дальнейшем применении искусственных нейронных сетей. Предобработка временных финансовых рядов состояния фондового рынка с помощью непрерывного вейвлет-преобразования и дальнейший анализ с помощью сверточных нейронных сетей изучается в работе Du B. и Barucca P. [11].

Чаще всего при проведении вейвлет-анализа исследователями используется система Matlab, языки программирования R или Python (см., напр., [12–14]). В настоящее время актуальной научной теоретической и практической задачей является поиск новых программно-аппаратных средств, позволяющих быстро и качественно проводить сложные вычисления. Одним из лидеров среди существующих программных продуктов является Wolfram Mathematica. К преимуществам применения данной системы относится и то, что благодаря возможности облачных вычислений (WolframCloud) пользователь становится слабозависимым от аппаратного обеспечения. Данная статья имеет своей целью демонстрацию предварительной обработки временного ряда финансовых данных с помощью вейвлет-анализа в системе компьютерной математики Wolfram Mathematica.

Вейвлет-анализ и пороговая обработка сигнала. Слово «wavelet» буквально означает «маленькая волна». Вейвлеты – это семейство локальных по времени и частоте функций. Все функции семейства получаются из одного вейвлета посредством его сдвига и растяжений. При анализе экономических данных в основном используются дискретные вейвлеты, обеспечивающие более удобную декомпозицию и реконструкцию сигналов.

Дискретное вейвлет-преобразование сигнала $f(t)$ состоит в его разложении в ряд [14]:

$$f(t) = c_0\psi(t) + \sum_{j=1}^{N}\sum_{k=0}^{2^j-1} c_{jk}\varphi_{jk}(t),$$

где $\psi(t)$ – называют отцовским вейвлетом, $\varphi_{jk}(t)$ – функции, полученные из материнского вейвлета $\varphi(t)$ путем сжатия и сдвига. Эти функции удовлетворяют условиям

$$\int_{-\infty}^{+\infty}\varphi(t)\,dt = 1,\ \int_{-\infty}^{+\infty}\psi(t)\,dt = 0.$$

Согласно теории многоуровневой аппроксимации сигналов, построенной Mallat [15], любой сигнал можно представить в виде суммы аппроксимирующего и детализирующих сигналов. В общем виде разложение сигнала до j-го уровня декомпозиции может быть представлено соотношением:

$$f(t) = A_j(t) + \sum_{m=1}^{j} D_m(t) = \sum_{k\in Z} a_{jk}\varphi_{jk}(t) + \sum_{k\in Z} d_{mk}\psi_{mk}(t),$$

где $a_{jk} = \left(f(t), \varphi_{jk}(t)\right)$ и $d_{mk} = \left(f(t), \psi_{mk}(t)\right)$ – аппроксимирующие и детализирующие коэффициенты, а базисные функции φ и ψ, соответственно, равны:

$$\varphi(t) = 2\sum_{l} L(l)\,\varphi(2t-l),\ L(l) = (\varphi(t), \varphi(2t-l)),$$

$$\psi(t) = 2\sum_{l} H(l)\,\varphi(2t-l),\ H(l) = (\psi(t), \varphi(2t-l)),\ l = 2n-1.$$

Коэффициенты аппроксимации содержат полезную информацию о поведении сигнала (грубое приближение сигнала). Коэффициенты детализации на каждом уровне разложения содержат уточняющую информацию о сигнале. В этих же коэффициентах могут содержаться шумы – случайные колебания данных, не подчиняющиеся определенным закономерностям. Они не несут полезной информации о главных тенденциях временного ряда данных и могут исказить интерпретацию сигнала. Поэтому анализ временных рядов

целесообразно предварять очисткой от шумов. Как правило, шумовые компоненты имеют меньшую амплитуду по сравнению с основным сигналом. В связи с этим одним из способов удаления шумов является трешолдинг - разложение сигнала на вейвлет-спектр до заданного уровня, обнуление тех вейвлет-коэффициентов разложения, значения которых меньше определенного порогового значения, и последующая вейвлет-реконструкция сигнала по аппроксимирующим и очищенным на каждом уровне детализирующим коэффициентам [4–5, 16–17].

Существуют разные виды пороговой обработки сигнала, из которых выделяют мягкий и жесткий. Жесткий трешолдинг предполагает обнуление всех коэффициентов, меньших заданного порога, остальные коэффициенты сохраняются неизменными:

$$x=\begin{cases} 0, |x| \le \delta, \\ x, |x| > \delta. \end{cases}$$

Мягкий трешолдинг дополнительно предполагает уменьшение по модулю всех остальных коэффициентов на величину порога:

$$x=\begin{cases} 0, |x| \le \delta, \\ \mathrm{sgn}(x)\,(|x|-\delta), |x| > \delta. \end{cases}$$

Порог может быть единым для всех коэффициентов и уровней разложения или различным в зависимости от частотных особенностей сигнала. От выбора порогового уровня зависит качество сглаживания сигнала – слишком малый порог не устранит шумы, слишком большой может привести к потере коэффициентов, содержащих существенную информацию о сигнале [5].

В качестве пороговых значений используются:

- порог на основе алгоритма Штейна несмещенной оценки риска:

$$\delta = arg\ min \sum_{i=1}^{N} \left(\overline{d_{ji}} - d_{ji} \right)^2,$$

где $\overline{d_{ji}}$ и d_{ij}, соответственно, эмпирические и истинные коэффициенты детализации сигнала,

- порог Донохо-Джонстона $\delta = \sigma\sqrt{2\ln N}$, где

$$\sigma = \frac{median\left\{ |x|, k = \overline{0, \frac{N}{2} - 1} \right\}}{0,6745}$$

- порог на основе минимаксной оценки

$$\delta = \begin{cases} \sigma\left(0,3936 + 0,1829\log_2 N\right), & N > 32 \\ 0, & N < 32 \end{cases}$$

$$\sigma = \frac{median\{|x|\}}{0,6745}$$

и некоторые другие.

Инструменты вейвлет-обработки в Wolfram Mathematica. В данной работе математическая обработка финансового временного ряда осуществляется с помощью системы компьютерной математики Wolfram Mathematica, обладающего мощной вычислительной и визуализирующей базой. Язык Wolfram Language обеспечивает полнофункциональную реализацию вейвлет-анализа [18].

Дискретное вейвлет-преобразование реализуется с помощью функции

```
DiscreteWaveletTransform[].
```

По умолчанию применяется вейвлет Хаара. В качестве материнского вейвлета в зависимости от характерных особенностей сигнала можно использовать вейвлеты:

```
DaubeshiesWavelet[], SymletWavelet[],
```

`BiorthogonalSplineWavelet[], CoifletWavelet[]` и др.

Восстановление сигнала осуществляется с помощью функции

`InverseWaveletTransform[].`

Для пороговой очистки используется функция WaveletThreshold:

`WaveletThreshold[dwd, tspec, wind],`

где *dwd* – коэффициенты дискретного вейвлет-преобразования, *tspec* – характеристика порога, *wind* – индексы вейвлет-коэффициентов, к которым будет применяться пороговая обработка.

В качестве характеристики порога можно выбрать встроенные:

- Hard (жесткая обработка с порогом δ),
- Soft (мягкая обработка с порогом δ)

и некоторые другие.

Пороговое значение δ также можно выбрать из имеющихся, например:

- "SURE" – несмещенная оценка риска Штейна;
- "SURELevel" – порог "SURE", применяемый для каждого уровня разложения сигнала;
- "Universal" – порог Донохо-Джонстона;
- "SUREHybrid" – комбинация порогового значения "SURE" и "Universal";
- "UniversalLevel" – порог "Universal", применяемый для каждого уровня разложения сигнала, и другие.

Эмпирическое исследование. В качестве объекта исследования выберем историю котировок акций Сбербанка за период с 1.01.2020 по 30.04.2024, в качестве единицы исследования примем ежедневные торги. Фурье-спектр этого ряда наглядно демонстрирует наличие высокочастотных шумов (рис. 1):

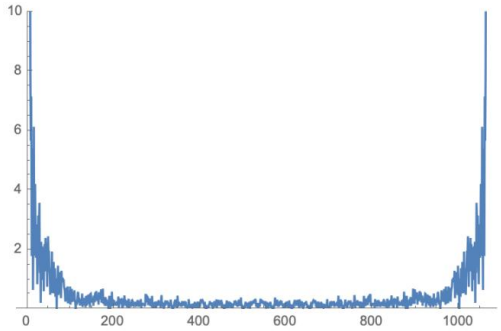


Рис. 1. Фурье-спектр сигнала

Проверка на стационарность данного временного ряда тестом Дики-Фулера подтверждает его нестационарность:

```
H = UnitRootTest[dat, Automatic, "HypothesisTestData"];
["TestDataTable", "DickeyFullerF"]
```

	Statistic	P-Value
Dickey-Fuller F	−0.0120161	0.679958

Для очистки сигнала от шума будем использовать ортогональные вейвлеты, обладающие возможностью быстрого разложения и реконструкции сигнала. Эффективность процедуры очистки сигнала в значительной степени зависит от вейвлет-базиса. Следуя [19–20], для выбора оптимального вейвлета будем использовать критерий минимума энтропий. Для этого рассчитаем энергию динамического ряда на каждом масштабе *j*:

$$E_j = \sum_k d_{j,k}^2$$

и найдем энтропию набора энергий:

$$WE = -\sum_{j=1}^{N} p_j \ln(p_j),$$

где $p_j = \dfrac{E_j}{\sum_{j=1}^{N} E_j}$ – относительная вейвлет-энергия, демонстрирующая распределение вейвлет-энергии по уровням разложения, N – максимальный уровень разложения. В качестве наилучшей выберем ту вейвлет-функцию, которой будет соответствовать наименьшая энтропия.

Для расчета вейвлет-энтропии использовалась опция EnergyFraction функции DiscreteWaveletTransform системы Wolfram Mathematica:

```
dwd = DiscreteWaveletTransform[dat, DaubechiesWavelet[4], 4]
efrac = dwd["EnergyFraction"];
ED4 = -Sum[efrac[[i, 2]] Log[efrac[[i, 2]]], {i, 1, Length[efrac}]
```

Энтропия рассчитывалась для вейвлетов Хаара, Мейера, Добеши и Симлета разных порядков. Разложение исходного сигнала возможно до 10 уровня. Однако исходя из того, что высшие уровни сигнала содержат лишь малую часть информации о сигнале, мы ограничились декомпозицией до 4 уровня. Анализ вейвлет-энтропий показал, что наиболее предпочтительным является использование вейвлета Добеши 6-го порядка (рис. 2), который можно рассматривать как компромисс между длиной области задания и регулярности базисной функции. Именно этот вейвлет был использован в дальнейших расчетах.

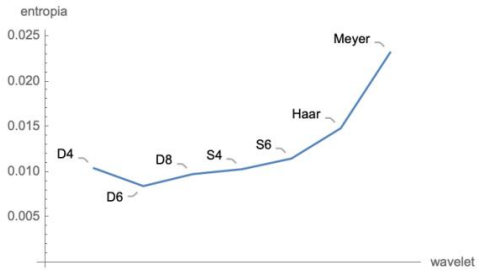


Рис. 2 Изменение вейвлет-энтропии в зависимости от вейвлет-базиса

На рис. 3 показано дерево последовательного разложения выбранного финансового ряда вейвлетом Добеши 6-го порядка на аппроксимирующие (a_i) и детализирующие (d_i) коэффициенты.

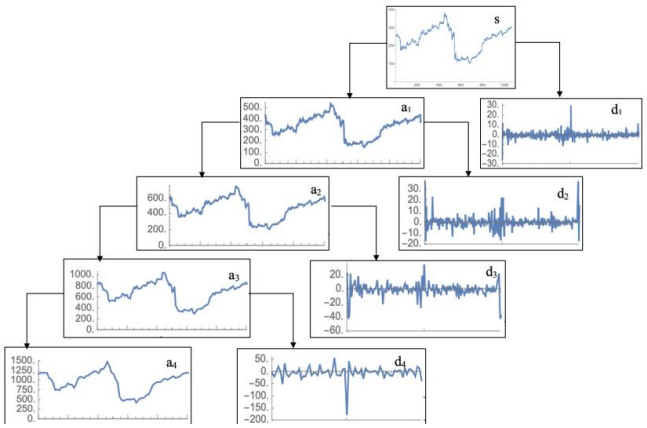


Рис. 3. Дерево разложения временного ряда

Очистка от шумов проводилась по детализирующим коэффициентам. В качестве порогового значения использовалась встроенная несмещенная оценка риска Штейна, порог Донохо-Джонстона и их модификации и комбинации. В Wolfram Mathematica очистка сигнала реализуется командами:

```
dwt = DiscreteWaveletTransform[dat, DaubechiesWavelet[6], 4]
WT = WaveletThreshold[dwt, {"Hard", "SURE"}];
WT["ThresholdTable"]
datthr = InverseWaveletTransform[WT];
```

С помощью опции ThresholdTable функции WaveletThreshold можно вывести пороговые значения очистки на каждом уровне разложения (табл.1).

Качество очистки оценивалось с помощью отношения сигнал/шум, определяемого формулой:

$$SNR = 10lg\ \frac{\sigma^2_{восст}}{\sigma^2_{шум}},$$

где $\sigma^2_{восст}$ – дисперсия восстановленного сигнала $\sigma^2_{шум}$ – дисперсия шума.

Таблица 1

Результаты пороговой очистки вейвлетом Добеши 6 порядка

Порог	Шумовой порог на каждом уровне разложения	SNR при пороговой обработке «Hard»	SNR при пороговой обработке «Soft»
"SURE"	1) 0,443732 2) 0,443732 3) 0,443732 4) 0,443732	56,822	44,286
"SURELevel"	1) 0,443732 2) 0,340379 3) 0,007522 4) 0,213248	57,492	45,849
"Universal"	1) 7,43896 2) 7,43896 3) 7,43896 4) 7,43896	28,274	25,509
"SUREHybrid"	1) 0,443732 2) 0,443732 3) 0,443732 4) 0,443732	56,822	44,286
"UniversalLevel"	1) 7,06539 2) 11, 9341 3) 19,9674 4) 53,8596	22,037	20,812

Как видим, наибольшее отношение сигнал/шум получено при жесткой пороговой обработке адаптивным порогом «SURELevel». На рис. 4 продемонстрирован сигнал после шумоподавления – трешолдинг позволил избавиться от шумов и при этом не исказил основную информацию по сигналу.

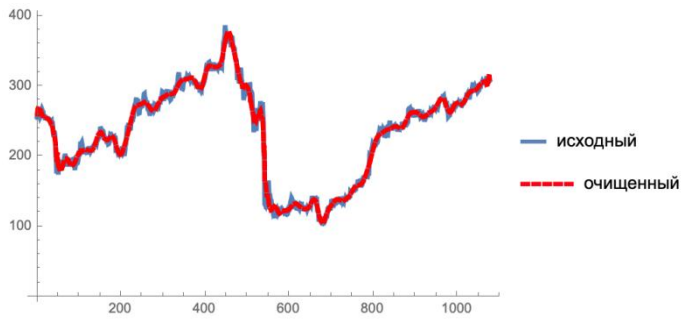


Рис. 4. Графики исходного и очищенного от шумов временного ряда

Очищенный временной ряд котировок применим для прогноза. Используя критерий Акаике, была подобрана оптимальная модель ARMA для исходных и очищенных данных и построен прогноз на 4 шага вперед (табл. 2):

```
TimeSeriesModelFit[dat, "ARMA"]["CandidateSelectionTable"];
tsm=TimeSeriesModelFit[dat, "ARMA"] // Normal;
forecast[tsm];
forecast["Path"]
```

Таблица 2

Прогнозные значения

Точные значения, май 2024	Прогноз по очищенным от шума данным	Прогноз по исходному временному ряду
307,37	307,767	308,072
307,53	307,472	307,835
306,01	307,273	307,633
308,22	307,075	307,432

Как видим, в краткосрочном прогнозе по очищенному ряду получаются более точные значения.

Выводы. Система Wolfram Mathematica обладает мощным функционалом, позволяющим проводить качественную обработку временных рядов. Анализ полученных результатов показал, что на качество очистки сигнала влияет выбор базисного вейвлета – в нашем случае предпочтительным оказалось использование вейвлета Добеши 6-го порядка. Максимальное отношение сигнал/шум достигается при жесткой пороговой обработке с порогом «SURELevel». Проведенные исследования показали, что вейвлет-трешолдинг над детализирующими коэффициентами вейвлет-разложения является эффективным методом подавления выбросов и флуктуаций временного ряда. Очищенный сигнал повторяет форму исходного сигнала, все пики хорошо выражены. При этом в краткосрочном прогнозе получаются более точные прогнозные значения.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Daubechies I.* The wavelet transform, time-frequency localization and signal analysis // IEEE transactions on information theory. – 1990. – 36 (5). – P. 961-1005.
2. *He T., Nguyen T.* Wavelet Analysis and Applications in Economics and Finance // Scholarship. – 2015. – 26. – URL: https://www.rroij.com/open-access/wavelet-analysis-and-applications-in-economics-and-finance.php?aid=62323&view=mobile.
3. *Chaovalit P., Gangopadhyay A., Karabatis G. and Chen Z.* Discrete wavelet transform-based time series analysis and mining // ACM Comput. Surv. – 2011. – 43, 2, Article 6. – 37 p. – URL: //doi.acm.org/10.1145/1883612.1883613.

4. *Кропотов Ю.А., Белов А.А., Проскуряков А.Ю.* Обработка временных рядов с применением вейвлет-преобразований для повышения точности представления информации // Транспортное машиностроение. – 2018. – № 8 (69). – URL: https://cyberleninka.ru/article/n/obrabotka-vremennyh-ryadov-s-primeneniem-veyvlet-preobrazovaniy-dlya-povysheniya-tochnosti-predstavleniya-informatsii (дата обращения: 22.04.2024).
5. *Белов А.А., Проскуряков А.Ю.* Сглаживание временных рядов на основе вейвлет-преобразования в системах автоматизированного экологического мониторинга // Методы и устройства передачи и обработки информации. – 2010. – № 1 (12). – С. 21-24.
6. *Taranenko Y.K.* Methods of Discrete Wavelet Filtering of Measurement Signals: an Algorithm for Choosing a Method // Meas Tech. – 2022. – 64. – P. 801-808. – https://doi.org/10.1007/s11018-022-02007-6.
7. *Shaikh W., Syed F., Pandhiani S. & Solangi M.* Wavelet Decomposition Impacts on Traditional Forecasting Time Series Models // Computer Modeling in Engineering & Sciences. – 2022. – Vol. 130. – P. 1517-1532. – DOI: 10.32604/cmes.2022.017822.
8. *Lee H.Y., Beh W.L. & Lem K.H.* Wavelet as a Viable Alternative for Time Series Forecasting // Austrian Journal of Statistics. – 2020. – 49. – P. 38-47.
9. *Benrhmach G., Namir K., Bouyaghroumni J. & Namir A.* Financial time series prediction using wavelet and artificial neural network. – 2021. – Vol 11, No. 5. – P. 5487-5500. – 10.28919/jmcs/5978.
10. *Boubaker H., Canarella G., Gupta R. et al.* A Hybrid ARFIMA Wavelet Artificial Neural Network Model for DJIA Index Forecasting // Comput. – 2023. – Econ 62. – P. 1801-1843. – https://doi.org/10.1007/s10614-022-10320-z.
11. *Du B., Barucca P.* Image Processing Tools for Financial Time Series Classification. – https://doi.org/10.48550/arXiv.2008.06042. – 2020.
12. *Lee G., Gommers R., Waselewski F., Wohlfahrt K. & Aaron* PyWavelets: A Python package for wavelet analysis // Journal of Open Source Software. – 2019. – 4. 1237. – 10.21105/joss.01237.
13. *Манонина И.В.* Обработка детализирующих вейвлет-коэффициентов для повышения точности рефлектометрических измерений // Научный вестник МГТУ ГА. – 2016. – № 5. – URL: https://cyberleninka.ru/article/n/obrabotka-detaliziruyuschih-veyvlet-koeffitsientov-dlya-povysheniya-tochnosti-reflektometricheskih-izmereniy (дата обращения: 22.04.2024).
14. *Дьяконов В.П.* Вейвлеты. От теории к практике. –2-е изд., перераб. и доп. – М.: СОЛОН-Пресс, 2010. – 400 с.
15. *Mallat S.G.* A Theory For Multiresolution Signal Decomposition: The Wavelet Representation // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1989. – No. 11. – P. 674-693.
16. *Московский С.Б., Сергеев А.Н., Лалина Н.А*. Очистка сигнала от шумов с использованием вейвлет-преобразования // Universum: Технические науки. – 2015. – № 2 (15). – URL: http://7universum.com/ru/tech/archive/item/1958.
17. *Московский С.Б., Сергеев А.Н., Сидорова Е.И., Марудов А.А.* Методы вейвлет-анализа в задачах обработки экспериментальных данных // EESJ. – 2019. – No. 4-4 (44). – URL: https://cyberleninka.ru/article/n/metody-veyvlet-analiza-v-zadachah-obrabotki-eksperimentalnyh-dannyh.
18. Wavelet Analysis. – URL: https://reference.wolfram.com/language/guide/Wavelets.html.
19. *Krotkikh S., Kirichenko L.* Analysis of event-related potentials of eeg signal using discrete wavelet transform // Radio Electronics, Computer Science, Control. – 2012. – 10.15588/1607-3274-2011-2-15.
20. *Мифтахова М.Э., Панасюк М.В.* Вейвлет-анализ динамики региональной социально-экономической системы // Ученые записки Казанского государственного университета. Серия «Естественные науки». – 2009. – Т. 151. – Вып. 1. – С. 247-262.

REFERENCES

1. *Daubechies I.* The wavelet transform, time-frequency localization and signal analysis, *IEEE transactions on information theory,* 1990, 36 (5), pp. 961-1005.
2. *He T., Nguyen T.* Wavelet Analysis and Applications in Economics and Finance, *Scholarship,* 2015, 26. Available at https://www.rroij.com/open-access/wavelet-analysis-and-applications-in-economics-and-finance.php?aid=62323&view=mobile.
3. *Chaovalit P., Gangopadhyay A., Karabatis G. and Chen Z.* Discrete wavelet transform-based time series analysis and mining, *ACM Comput. Surv*., 2011, 43, 2, Article 6, 37 p. Available at: doi.acm.org/10.1145/1883612.1883613.
4. *Kropotov Yu.A., Belov A.A., Proskuryakov A.Yu.* Obrabotka vremennykh ryadov s primeneniem veyvlet-preobrazovaniy dlya povysheniya tochnosti predstavleniya informatsii [Processing of time series using wavelet transforms to improve the accuracy of information presentation], *Transportnoe mashinostroenie* [ransport engineering], 2018, No. 8 (69). Available at: https://cyberleninka.ru/article/n/obrabotka-vremennyh-ryadov-s-primeneniem-veyvlet-preobrazovaniy-dlya-povysheniya-tochnosti-predstavleniya-informatsii (accessed 22 April 2024).

5. *Belov A.A., Proskuryakov A.Yu.* Sglazhivanie vremennykh ryadov na osnove veyvlet-preobrazovaniya v sistemakh avtomatizirovannogo ekologicheskogo monitoringa [Smoothing of time series based on wavelet transform in automated environmental monitoring systems], *Metody i ustroystva peredachi i obrabotki informatsii* [Methods and devices for information transmission and processing], 2010, No. 1 (12), pp. 21-24.
6. *Taranenko Y.K.* Methods of Discrete Wavelet Filtering of Measurement Signals: an Algorithm for Choosing a Method, *Meas Tech*., 2022, 64, pp. 801-808. Available at: https://doi.org/10.1007/s11018-022-02007-6.
7. *Shaikh W., Syed F., Pandhiani S. & Solangi M.* Wavelet Decomposition Impacts on Traditional Forecasting Time Series Models, *Computer Modeling in Engineering & Sciences,* 2022, Vol. 130, pp. 1517-1532. DOI: 10.32604/cmes.2022.017822.
8. *Lee H.Y., Beh W.L. & Lem K.H.* Wavelet as a Viable Alternative for Time Series Forecasting, *Austrian Journal of Statistics,* 2020, 49, pp. 38-47.
9. *Benrhmach G., Namir K., Bouyaghroumni J. & Namir A.* Financial time series prediction using wavelet and artificial neural network // Journal of Mathematical and Computational Science, 2021, Vol. 11, No. 5, pp. 5487-5500. 10.28919/jmcs/5978.
10. *Boubaker H., Canarella G., Gupta R. et al.* A Hybrid ARFIMA Wavelet Artificial Neural Network Model for DJIA Index Forecasting, *Comput*., 2023, Econ 62, pp. 1801-1843. Available at: https://doi.org/10.1007/s10614-022-10320-z.
11. *Du B., Barucca P.*_Image Processing Tools for Financial Time Series Classification. Available at: https://doi.org/10.48550/arXiv.2008.06042, 2020.
12. *Lee G., Gommers R., Waselewski F., Wohlfahrt K. & Aaron* PyWavelets: A Python package for wavelet analysis, *Journal of Open Source Software,* 2019, 4. 1237. 10.21105/joss.01237.
13. *Manonina I.V.* Obrabotka detaliziruyushchikh veyvlet-koeffitsientov dlya povysheniya tochnosti reflektometricheskikh izmereniy [Processing of detailing wavelet coefficients to improve the accuracy of reflectometric measurements], *Nauchnyy vestnik MGTU GA* [Scientific Bulletin of MSTU GA], 2016, No. 5. Available at: https://cyberleninka.ru/article/n/obrabotka-detaliziruyuschih-veyvlet-koeffitsientov-dlya-povysheniya-tochnosti-reflektometricheskih-izmereniy (accessed 22 April 2024).
14. *D'yakonov V.P.* Veyvlety. Ot teorii k praktike [Wavelets. From theory to practice]. 2nd ed. Moscow: SOLON-Press, 2010, 400 p.
15. *Mallat S.G.* A Theory For Multiresolution Signal Decomposition: The Wavelet Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1989, No. 11, pp. 674-693.
16. *Moskovskiy S.B., Sergeev A.N., Lalina N.A.* Ochistka signala ot shumov s ispol'zovaniem veyvlet-preobrazovaniya [Signal cleaning from noise using wavelet transform], *Universum: Tekhnicheskie nauki* [Universum: Technical sciences], 2015, No. 2 (15). Available at: http://7universum.com/ru/tech/archive/item/1958.
17. *Moskovskiy S.B., Sergeev A.N., Sidorova E.I., Marudov A.A.* Metody veyvlet-analiza v zadachakh obrabotki eksperimental'nykh dannykh [Wavelet Analysis Methods in Experimental Data Processing Problems], *EESJ,* 2019, No. 4-4 (44). Available at: https://cyberleninka.ru/article/n/metody-veyvlet-analiza-v-zadachah-obrabotki-eksperimentalnyh-dannyh.
18. Wavelet Analysis. Available at: https://reference.wolfram.com/language/guide/Wavelets.html.
19. *Krotkikh S., Kirichenko L.* Analysis of event-related potentials of eeg signal using discrete wavelet transform, *Radio Electronics, Computer Science, Control,* 2012. 10.15588/1607-3274-2011-2-15.
20. *Miftakhova M.E., Panasyuk M.V.* Veyvlet-analiz dinamiki regional'noy sotsial'no-ekonomicheskoy sistemy [Wavelet analysis of the dynamics of a regional socio-economic system], *Uchenye zapiski Kazanskogo gosudarstvennogo universiteta. Seriya «Estestvennye nauki»* [Scientific notes of Kazan State University. Series "Natural sciences"], 2009, Vol. 151, Issue 1, pp. 247-262.

Статью рекомендовал к опубликованию д.т.н., профессор А.В. Боженюк.

Хайруллина Лилия Эмитовна – Казанский федеральный университет; e-mail: lxayrullina@yandex.ru; Казань, Россия; тел.: +79272464003; кафедра информационных систем; к.ф.-м.н.; доцент.

Хакимов Зульфат Ниязович – e-mail: znkhakimov@stud.kpfu.ru; аспирант.

Хабибуллина Гузель Забировна – e-mail: hgz1980@rambler.ru; тел.: +79274057021; кафедра теории относительности и гравитации; к.пед.н.; доцент.

Khairullina Lilia Emitovna – Kazan Federal University; e-mail: lxayrullina@yandex.ru; Kazan, Russia; phone: +79272464003; the Department of Information Systems; cand. of phys. and math. sc.; associate professor.

Khakimov Zulfat Niyazovich – e-mail: znkhakimov@stud.kpfu.ru; graduate student.

Khabibullina Guzel Zabirovna – e-mail: hgz1980@rambler.ru; phone: +79274057021; the Department of Relativity and Gravity; cand. of ped. sc.; associate professor.

УДК 004.8 DOI 10.18522/2311-3103-2024-4-144-151

М.Д. Трегубенко

АНАЛИЗ МЕТОДОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ РАСПОЗНАВАНИЯ ДЕФЕКТОВ СОЛНЕЧНЫХ ПАНЕЛЕЙ (ОБЗОР)

В современном мире, где экологические проблемы становятся всё более актуальными, поиск альтернативных источников энергии становится приоритетной задачей. Одним из наиболее перспективных направлений является солнечная энергетика. Солнечная энергия является возобновляемым источником энергии, что делает её привлекательной для использования в различных областях, включая производство электроэнергии, отопление и охлаждение зданий, а также транспорт. Развитие солнечной энергетики может способствовать решению ряда экологических проблем, таких как загрязнение окружающей среды и изменение климата. Однако, оборудование для солнечных панелей подвержено различным типам дефектов и загрязнений. Неисправности могут негативно влиять на производительность и эффективность солнечных панелей, поэтому их обнаружение является критически важным для повышения надёжности и долговечности фотоэлектрических систем генерации энергии. Эффективный поиск дефектов позволяет минимизировать потери энергии, повысить надёжность системы и срок службы оборудования, а также снизить затраты на обслуживание. Кроме того, повышение производительности электрического оборудования способствует устойчивому развитию альтернативной энергетики, что позволяет уменьшить зависимость от традиционных источников энергии и снизить выбросы парниковых газов. Статья представляет собой обзор существующих методов обнаружения различных дефектов солнечных панелей с помощью методов компьютерного зрения и глубокого обучения. Для поиска неисправностей можно использовать изображения, полученные с помощью инфракрасной термографии (ИК), электролюминесцентной (ЭЛ) визуализации, либо в видимом спектре. Данная работа включает в себя анализ преимуществ и недостатков существующих методов поиска дефектов и загрязнений солнечных панелей, обсуждение факторов, влияющих на их эффективность, и представляет выводы по возможным дальнейшим исследованиям в данной сфере.

Компьютерное зрение; глубокое обучение; солнечные панели; нейронные сети детекция.

M.D. Tregubenko

ANALYSIS OF COMPUTER VISION METHODS FOR RECOGNISING SOLAR PANEL DEFECTS (REVIEW)

In today's world, where environmental problems are becoming more and more urgent, the search for alternative energy sources is becoming a priority. One of the most promising areas is solar energy. Solar energy is a renewable energy source, which makes it attractive for use in various areas, including power generation, heating and cooling of buildings, and transport. The development of solar energy can contribute to solving a number of environmental problems such as pollution and climate change. However, solar panel equipment is subject to various types of defects and contamination. Defects can adversely affect the performance and efficiency of solar panels, so their detection is critical to improve the reliability and durability of photovoltaic power generation systems. Effective fault finding can minimise energy losses, improve system reliability and equipment life, and reduce maintenance costs. In addition, improved performance of electrical equipment contributes to the sustainable development of alternative energy, thus reducing dependence on conventional energy sources and reducing greenhouse gas emissions. The paper presents an overview of existing methods for detecting various solar panel faults using computer vision

and deep learning techniques. Infrared thermography (IR), electroluminescence (EL) imaging, or visible spectrum imaging can be used to find the faults. This paper includes an analysis of the advantages and disadvantages of existing methods for finding defects and contamination in solar panels, discusses the factors affecting their performance, and presents conclusions for possible future research in this area.

Computer vision; deep learning; solar panels; neural networks; detection.

Введение. В современном мире, где экологические проблемы становятся всё более актуальными, поиск альтернативных источников энергии становится приоритетной задачей. Одним из наиболее перспективных направлений является солнечная энергетика, которая обладает рядом преимуществ перед традиционными источниками энергии. Солнечная энергия является возобновляемым источником, который не требует дополнительных затрат на топливо и не создаёт отходов. Кроме того, солнечные панели могут быть установлены практически в любом месте, что делает их особенно привлекательными для использования в отдалённых районах. Различные статистические результаты подтверждают значимость солнечной энергии и подключенных к сети солнечных батарей во всем мире. Несмотря на все преимущества, солнечная энергетика также сталкивается с рядом проблем, которые необходимо решить для её дальнейшего развития. Энергоэффективность солнечных элементов часто ограничивается возникающими дефектами, которые могут снизить их производительность и срок службы. Дефекты могут рассеивать энергию, создавая новые пути рекомбинации (потери), позволяя свету генерировать тепло, а не электричество, или даже потреблять энергию, хранящуюся в аккумуляторе, снижая эффективность фотоэлектрического модуля.

Одной из основных проблем, возникающих при использовании солнечных панелей, является их низкая стабильность, на которую могут влиять как внутренние, так и внешние факторы, включая намокание, тепловой стресс или влажные условия, приводящие к химическим и структурным изменениям [1]. В конечном итоге это может привести к увеличению стоимости и ограниченному сроку службы. Поэтому очень важно определить набор подходов к обнаружению дефектов для предиктивного обслуживания и мониторинга состояния фотоэлектрических модулей.

В статье [2] авторы подробно анализируют методы, которые выполняют поиск дефектов с помощью изображений, полученных с помощью инфракрасной термографии, электролюминесцентной визуализации. Однако к ним необходимо добавить методы, которые выполняют поиск дефектов солнечных панелей в изображениях, полученных в видимом спектре. По ним можно хорошо детектировать различные загрязнения и внешние факторы, которые могут понижать производительность панелей: пыль, ветки, вода, выделения птиц и т.д.

В данной статье рассматриваются и анализируются различные методы анализа данных для обнаружения дефектов солнечных панелей на основе компьютерного зрения.

Методы инфракрасного спектра. Анализ дефектов с помощью инфракрасного спектра считается одним из широко используемых неинвазивных методов, при котором излучение, испускаемое поверхностью любого тела, обрабатывается в инфракрасном диапазоне длин волн между 1,4 и 15 мкм. В этом методе используются тепловизоры, которые обычно способны измерять длины волн в среднем инфракрасном диапазоне 7-14 мкм. Для сбора данных требуются только инфракрасные (ИК) камеры без привлечения другого оборудования или прерывания работы фотоэлектрической системы [3].

В случае исправного фотоэлектрического модуля полученное распределение температуры на тепловом изображении является однородным по поверхности модуля. Дефекты солнечной панели, оказывающие значительное влияние на его тепловое поведение, можно обнаружить, наблюдая колебания температуры, приводящие к неоднородности тепловых распределений. Тепловые колебания возникают из-за тепла, выделяемого при облучении, или фотонов, которые не преобразуются в электроэнергию. Следовательно, накопление тепла повышает температуру, что приводит к появлению дефектов на полученных изображениях, которые могут характеризоваться повышенным последовательным сопротивлением, отключенным, зашунтированными ячейками или другими факто-

рами, влияющими на тепловое распределение. Например, дефекты, появляющиеся на тепловом изображении, могут быть отключенными модулями, распознаваемыми по более горячим участкам, затенением, распознаваемым по горячим точкам, и трещинами в ячейках, распознаваемыми по удлиненному нагреву ячеек [4]. Таким образом, инфракрасная термография является эффективным методом диагностики фотоэлектрических систем, позволяющим обнаруживать и локализовать дефекты без необходимости прерывания работы системы или использования дополнительного оборудования.

Классический алгоритм анализа тепловой энергии на основе обработки тепловых изображений предложен Курукуру и др. [5] на основе обнаружения краев Канни и преобразования Хофа. Сначала к изображению применялось извлечение признаков, которые проходили через алгоритм классификации для локализации и идентификации типа дефекта. Для получения карты краев сначала применялась фильтрация шума с помощью фильтра Гаусса, а затем вычислялся градиент интенсивности. Пиксели, не вносящие вклад в формирование края, удалялись с помощью не максимального подавления с последующим пороговым вычислением гистерезиса для определения градиента пикселя. Для разделения панелей линии на изображении определялись с помощью преобразования Хофа. Точность тестирования составила 93,1%.

Оливейра и др. [6] представили обзор воздушной инфракрасной томографии для автоматической инспекции фотоэлектрических станций, обобщив алгоритмы цифровой обработки изображений и алгоритм на основе глубокого обучения. В обзоре рассматривались три основных аспекта процедуры автоматического анализа изображений в ИК спектру: оптимизация траектории полета, ортофотомозаика солнечной станции и обнаружение загрязнений. Данный метод показал с точность до 90% при обнаружении и классификации неисправностей в 10 различных типах аномалий.

Также для анализа ИК изображений используются различные традиционные и более совершенные алгоритмы машинного обучения. В работе Али и др. [7] для обнаружения и классификации "горячих точек" использовалась модель машины опорных векторов (SVM). Для обучения SVM-модели был предложен гибридный вектор признаков, состоящий из RGB изображений, текстуры, гистограммы ориентированного градиента (HOG) и локального бинарного паттерна. Предварительная обработка состояла из преобразования изображения в одноканальное, выравнивания гистограммы и фильтрации шума (с помощью фильтра Габора). Гибридное извлечение признаков применялось перед обучением SVM. Модель SVM достигла точности тестирования 92 % и превзошла другие традиционные алгоритмы машинного обучения, которые анализировались в данной статье: квадратичным дискриминантным анализом, наивным Байесом, K-ближайших соседей и Бэггинг Ансамбль при более высокой эффективности вычислений и хранения данных.

Хайдари и др. [8] модифицировали архитектуру VGG16 для применения на солнечных станциях, в основном рассматривая три класса: исправные, горячие точки и подстроки. Набор данных был собран с помощью беспилотного летательного аппарата (БПЛА) и ручных портативных ИК-камер. Всего вручную было промаркировано 1116 тепловых изображений, из которых большинство относится к классу исправных. В данном исследовании гиперпараметры оригинальной сети VGG16, состоящей из 13 сверточных нейронных сетей (CNN) в 5 блоках [9], были изменены, что позволило достичь общей точности 98 %.

На температуру могут влиять несколько окружающих факторов, приводящих к неверным отклонениям в температурных измерениях, включая изменения излучательной способности (т.е. способности тел излучать инфракрасную энергию), отражения, угол обзора (поскольку ИК-камера может давать различные температуры в зависимости от угла обзора), неисправности камеры (по вине оператора или из-за ограничений ИК-камеры), а также помехи окружающей среды (например, ветер, солнечный свет и влажность). Кроме того, ложная идентификация неисправностей может происходить из-за помех от других теплоизлучающих тел [10].

Не смотря на высокие оценки точности некоторых моделей, они приспособлены только для одной задачи, являются недостаточно гибкими для поиска нескольких типов неисправностей и могут быть доработаны более современными алгоритмами глубокого обучения для улучшения обобщающей способности.

Методы электролюминесцентного спектра. В солнечных панелях падающая солнечная энергия преобразуется в электричество при определенном диапазоне коэффициента преобразования (КП). Энергетический КП солнечного элемента определяется как соотношение между максимальной электрической мощностью, которая может быть передана нагрузке, и мощностью падающего на устройство излучения. Электролюминесцентная (ЭЛ) визуализация - это метод, при котором дефекты поверхности модуля проверяются путем исследования областей фотоэлектрического модуля с изменяющимся КП, в которых область с низким КП может указывать на наличие дефектов. В системе формирования ЭЛ-изображений внешний источник питания постоянного тока подключается к фотоэлектрическому модулю в затемненном помещении для пропускания электрического тока и проявления фотоэмиссии возбужденного фотоэлектрического модуля с помощью инфракрасной чувствительной камеры. На полученном ЭЛ-изображении более яркая люминесценция наблюдается в областях с высоким уровнем КП, в то время как неактивные области или области с низким уровнем светового излучения указывают на дефекты, проявляющиеся в виде темных областей [11].

В работе [12] авторы представили систему обнаружения дефектов на основе глубокого обучения с использованием генеративно-состязательных сетей (GAN). Сначала система генерирует набор данных изображений высокого разрешения, используя небольшое количество существующих изображений. После этого на сгенерированном наборе данных обучается классификатор CNN для извлечения признаков. Утверждается, что такой подход позволяет повысить надежность классификации с точностью 83 %, при этом система тестировалась в сравнении с такими распространенными моделями, как MobileNet, Inception V3, VGG16 и ResNet50.

Дейч и др. [13] разработали комбинацию классического SVM и нейросетевой CNN в рамках контролируемой классификации для обнаружения дефектов солнечных элементов. Оба подхода различались по необходимой аппаратной эффективности: SVM обеспечил точность около 82% при низких вычислительных требованиях, в то время как CNN достиг более высокой точности около 88%. В ходе исследования также был получен набор данных, включающий около 2600 аннотированных изображений солнечных элементов ЭЛ. Для обучения CNN было использовано трансферное обучение путем адаптации архитектуры VGG19, первоначально обученной на наборе данных ImageNet, к набору данных ЭЛ.

Чен и др. [14] разработали быструю систему диагностики дефектов для обнаружения нескольких типов дефектов. Для обучения системы используется мультиалгоритмический классификатор электролюминесцентных изображений. Набор данных PVLP, включающий около 2,4 млн изображений фотоэлементов, был обучен на моделях Random Forest, ResNet и YOLO. Текущие результаты показывают, что точность классификации составляет 84 %, а время вывода не превышает 0,5 с.

Работа Фиорези и др. [15] представляет собой пример комбинированного использования обработки изображений и глубокого обучения для диагностики дефектов солнечных панелей. Была предложена модель, позволяющая различать четыре типа дефектов в мультикристаллических и монокристаллических ячейках. Используя комбинацию Deeplabv3 и ResNet-50, модель была обучена на более чем 17 000 изображений. По результатам тестирования точность составила около 95%. Стоит отметить, что некоторые изображения в процессе обучения были сгенерированы для уменьшения дисбаланса классов.

Таким образом стоит отметить, что имеются определенные методы решения задачи обнаружения дефектов в изображениях электролюминесцентного спектра, однако стоит отметить индивидуальный подход каждого предложенного метода к определенному набору данных.

Методы видимого спектра. В естественных условиях солнечные панели сталкиваются с проблемой осадков (воды, снега) и загрязнением поверхности пылью, выделениями птиц, листьями и т.п. Необходима регулярная очистка солнечных панелей от пыли и грязи, требующая затрат рабочей силы и ресурсов. Если пренебрегать очисткой, то налипание грязи может повредить поверхность панели, сократив срок ее службы, скопление пыли на инверторе или соединительных проводах может привести к перегреву, короткому замыканию или выходу из строя оборудования, что влияет на стабильность работы и безопасность всей системы. Для автоматизации процесса уборки используют камеры, закрепленные на беспилотных летательных аппаратах, которые снимают панели в видимом спектре и фиксируют загрязнение, которое потом передается операторы системы [16].

Фан и др. [17] предложили новый метод оценки степени накопления пыли на панелях на основе алгоритма улучшения изображения. Они использовали модель атмосферного рассеяния для анализа характеристик изображений чистых и покрытых пылью фотоэлектрических панелей и установили связь между коэффициентами модели и уровнем запыленности. Авторы проверили применимость и точность модели с помощью экспериментов при различных условиях облучения в помещении и на улице. Результаты показали, что модель может оценить уровень пыли на фотоэлектрических панелях с точностью 83,78%.

Чжоу и др. [18] разработали метод автоматической сегментации и классификации загрязненных солнечных панелей на основе изображений видимого спектра, полученных с помощью беспилотников. Метод использовал гибкую пороговую схему сегментации, которая сочетала в себе цветовое пространство серого и S-компоненту. Он также интегрировал текстуру и пиксельные характеристики электрических панелей для автоматического обнаружения загрязнения. На этапе автоматической классификации было оценено несколько нейронных сетей с использованием трансферного обучения. В процессе классификации точность составила 86,8%.

Оним и др. [19] представили новый набор данных изображений солнечных панелей в чистых и пыльных условиях. Авторы предлагают новую архитектуру CNN под названием SOLNET, которая может достичь точности 98,2% при определении уровня пыли на солнечных панелях. Авторы хотят в будущем расширить категории в наборе данных, чтобы включить различные степени и типы скопления пыли, а также собрать более разнообразный датастет с различными условиями окружающей среды. В будущем SolNet можно будет точно настроить, оптимизировав гиперпараметры и структуру сети. Также авторы хотят в будущем добавить возможность тонкой настройки архитектуры.

Йдрисси и др. [20] анализируют влияние пыли на солнечные системы с помощью исследовательской платформы GEP. Они предлагают систему обнаружения пыли, которая сочетает в себе метод классификации с помощью сверточной нейронной сети и алгоритма обработки изображений. Авторы используют передовой метод глубокого обучения для классификации различных степеней запыленности и достигают точности 96%, демонстрируя потенциал применения глубокого обучения в области солнечной энергетики. Они также используют GPS-координаты для записи местоположения каждого измерения, чтобы составить карту распределения пыли на солнечной станции.

Стоит отметить, что существующие методы анализа концентрируются в основном на пыли и исключают множество других факторов и типы загрязнений.

Сравнение результатов работы различных методов и подходов. Для возможности сравнения и унификации различных методов была взята классическая метрика точности (accuracy):

$$Accuracy = \frac{TP + FP}{(TP + FP + TN + FN)}. \quad (1)$$

Плюсы и минусы изображений различного спектра представлены в табл. 1.

Таблица 1

Сравнение изображений различного спектра

Вид спектра	Преимущества	Недостатки
Инфракрасный	Масштабируемость. Эффективное обнаружение горячих точек и перепадов температур.	Дорогой. Внешние факторы (например, колебания излучательной способности, отражения) могут влиять на точность. Надежность зависит от солнечного света.
Электролюминесентный	Локализация дефектов в солнечных элементах. Обнаружение дефектов внутри солнечных элементов.	Сложность масштабирования. Ограничено оптической деградацией (например, расслоение или разрушение стекла).
Видимый	Дешевизна. Масштабируемость. Обнаружение загрязнений.	Зависимость от внешних факторов (облачности, солнечного света) Плохо обнаруживает внутренние дефекты панели

Сводная таблица всех рассматриваемых в данной работе методов представлена в табл. 2:

Таблица 2

Сводная таблица рассматриваемых методов

№	Авторы работы	Спектр изображения	Архитектура алгоритма	Точность, %
1	Курукуру и др. [5]	ИК	Преобразование Хофа + фильтр Гаусса	93
2	Оливейра и др. [6]	ИК	CNN	90
3	Али и др. [7]	ИК	SVM + HOG	92
4	Хайдари и др. [8]	ИК	VGG16	98
5	Танг и др. [12]	ЭЛ	CNN + GAN	83
6	Дейч и др. [13]	ЭЛ	SVM + CNN	88
7	Чен и др. [14]	ЭЛ	Random Forest, ResNet, YOLO	84
8	Фиорези и др. [15]	ЭЛ	Deeplabv3 + ResNet-50	95
9	Фан и др. [17]	Видимый	Модель атмосферного рассеяния	83
10	Чжоу и др. [18]	Видимый	Цветовое пространство серого + S-компонента	86
11	Оним и др. [19]	Видимый	CNN (SOLNET)	98
12	Йдрисси и др. [20]	Видимый	CNN + GPS	96

Выводы. В статье представлен обзор исследовательской литературы по методам анализа данных и поиска дефектов и загрязнений в солнечных панелях. Анализ позволил сделать следующие выводы:

1. Различные виды спектров изображений хороши для поиска различных дефектов, каждый имеет свои плюсы и минусы.

2. Представлены различные методы анализа дефектов и загрязнений солнечных панелей. Результаты некоторых методов показывают большую точность, однако это наблюдается для узконаправленной задачи и небольшого набора данных, что в значительной степени понижает обобщающую способность алгоритма.

3. В исследованиях используются классические архитектуры, которые прошли проверку временем. Новаторские архитектуры не используются.

4. Представленные алгоритмы работают в одном спектре. Анализа изображений в различных спектрах в рамках одного алгоритма не проводятся.

5. Таким образом, проблема исследованных алгоритмов состоит в том, что существующие решения работают в одном спектре, заточены под конкретный набор данных, что ослабляет обобщающую способность, и не используют свежие архитектурные решения, которые могут увеличить значения качественных метрик.

6. Для решения данной проблемы необходимо проведение дальнейших исследований в направлении внедрения новых нейросетевых архитектур, увеличения универсальности и масштабируемости алгоритмов путем объединения и расширения существующих наборов данных и создания модели, которая работает с изображениями из нескольких каналов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Bouich A.* Study and characterization of hybrid perovskites and copper-indium-gallium selenide thin films for tandem solar cells: Diss. – Universitat Politècnica de València, 2021.
2. *Hijjawi U. et al.* A review of automated solar photovoltaic defect detection systems: Approaches, challenges, and future orientations, *Solar Energy,* 2023, Vol. 266, pp. 112186.
3. *Tsanakas J. A. et al.* Fault diagnosis and classification of large-scale photovoltaic plants through aerial orthophoto thermal mapping, *Proceedings of the 31st European Photovoltaic Solar Energy Conference and Exhibition,* 2015, Vol. 2015, pp. 1783-1788.
4. *Tsanakas J.A., Ha L., Buerhop C.* Faults and infrared thermographic diagnosis in operating c-Si photovoltaic modules: A review of research and future challenges, *Renewable and sustainable energy reviews,* 2016, Vol. 62, pp. 695-709.
5. *Kurukuru V.S.B., Haque A., Khan M.A.* Fault classification for photovoltaic modules using thermography and image processing, *2019 IEEE Industry Applications Society Annual Meeting*. IEEE, 2019, pp. 1-6.
6. *de Oliveira A.K.V., Aghaei M., Rüther R.* Automatic inspection of photovoltaic power plants using aerial infrared thermography: a review, *Energies,* 2022, Vol. 15, No. 6, pp. 2055.
7. *Ali M.U. et al.* A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography // Solar Energy, 2020, Vol. 208, pp. 643-651.
8. *Haidari P. et al.* Deep learning-based model for fault classification in solar modules using infrared images // Sustainable Energy Technologies and Assessments, 2022, Vol. 52, pp. 102110.
9. *Gopalakrishnan K. et al.* Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection, *Construction and building materials,* 2017, Vol. 157, pp. 322-330.
10. *Herraiz A.H., Marugán A.P., Márquez F.P.G.* A review on condition monitoring system for solar plants based on thermography, *Non-destructive testing and condition monitoring techniques for renewable energy industrial assets,* 2020, pp. 103-118.
11. *Tsai D.M., Wu S.C., Li W.C.* Defect detection of solar cells in electroluminescence images using Fourier image reconstruction, *Solar Energy Materials and Solar Cells,* 2012, Vol. 99, pp. 250-262.
12. *Tang W. et al.* Deep learning based automatic defect identification of photovoltaic module using electroluminescence images, *Solar Energy,* 2020, Vol. 201, pp. 453-460.
13. *Deitsch S. et al.* Automatic classification of defective photovoltaic module cells in electroluminescence images, *Solar Energy,* 2019, Vol. 185, pp. 455-468.
14. *Chen X., Karin T., Jain A.* Automated defect identification in electroluminescence images of solar modules, *Solar Energy,* 2022, Vol. 242, pp. 20-29.
15. *Fioresi J. et al.* Automated defect detection and localization in photovoltaic cells using semantic segmentation of electroluminescence images, *IEEE Journal of Photovoltaics,* 2021, Vol. 12, No. 1, pp. 53-61.
16. *Yang C. et al.* A Survey of Photovoltaic Panel Overlay and Fault Detection Methods, *Energies,* 2024, Vol. 17, No. 4, pp. 837.

17. *Fan S. et al.* A novel image enhancement algorithm to determine the dust level on photovoltaic (PV) panels, *Renewable energy,* 2022, Vol. 201, pp. 172-180.
18. *Zhou Y.J., Sun H.R.* Water photovoltaic plant contaminant identification using visible light images, *Sustainable Energy Technologies and Assessments,* 2022, Vol. 53, pp. 102476.
19. *Onim M.S.H. et al.* SolNet: a convolutional neural network for detecting dust on solar panels, *Energies,* 2022, Vol. 16, No. 1, pp. 155.
20. *El Ydrissi M. et al.* Dust InSMS: Intelligent soiling measurement system for dust detection on solar mirrors using computer vision methods, *Expert Systems with Applications,* 2023, Vol. 211, pp. 118646.

Статью рекомендовал к опубликованию к.т.н., доцент С.А. Кучеров.

Трегубенко Максим Дмитриевич – Южный федеральный университет; e-mail: mtregubenko@sfedu.ru; г. Таганрог, Россия; тел.: +79185244987; кафедра ИАСБ.

Tregubenko Maxim Dmitrievich – Southern Federal University; e-mail: mtregubenko@sfedu.ru; Taganrog, Russia; phone: +79185244987; the Department of IASB.

Раздел III. Электроника, нанотехнологии и приборостроение

УДК 666.655 DOI 10.18522/2311-3103-2024-4-152-157

М.А. Мараховский, Л.А. Дыкина, В.В. Филь, А.А. Панич

ИССЛЕДОВАНИЕ ВЛИЯНИЯ ПРИМЕСНЫХ ФАЗ ИСХОДНОГО СЫРЬЯ НА СВОЙСТВА СЕГНЕТОЭЛЕКТРИЧЕСКОЙ КЕРАМИКИ СИСТЕМЫ ЦТС

В процессе серийного изготовления сегнетоэлектрических материалов в исходном сырье периодически наблюдаются примеси различного типа и концентрации. ***Целью исследования*** *было определение влияния примесных фаз присутствующих в исходном сырье на диэлектрические и электрофизические свойства сегнетоэлектрической керамики. В работе методом спектрального анализа исследованы базовые сырьевые компоненты входящие в систему цирконата - титаната свинца на наличие примесных составляющих. В результаты была выявлена группа примесных фаз (Sb, Na, Bi, K, Fe) различной концентрации. Модельным объектом исследования был выбран промышленно выпускаемый сегнетоэлектрический материал со структурой перовскита и химической формулой* $Pb_{0,95}Sr_{0,05}(Zr_{0,53}Ti_{0,47})O_3$ *+ 1%* Nb_2O_5. ***Задача исследования*** *состояла в дозированном введении примесных легирующих добавок в состав исходного сегнетоэлектрического материала с целью возможного изменения конечных свойств. В ходе исследования установлена актуальность дозированного введения примесей K и Na в концентрации 1–2 % в систему ЦТС с целью снижения значений относительной диэлектрической проницаемости на 40–45 %. Методом растровой электронной микроскопии установлены зависимости формируемой керамической структуры от введённых примесных легирующих фаз. Установлены закономерности «тип примесной добавки – микроструктура – свойства». В результате исследования подтверждена эффективность дозированного введения примесных легирующих добавок K и Na в целях модификации свойств сегнетоэлектрической керамики системы ЦТС. Такое примесное легирование приводит к повышению значений удельной чувствительности по напряжению* (g_{33}) *до 34–37 мВ·м/Н. Сегнетоэлектрические материалы такого формата представляют высокий практический интерес для создания акустических преобразователей работающих в режиме приёма.*

Сегнетоэлектрический материал; исходное сырьё; примесная фаза; микроструктура; удельная чувствительность по напряжению.

M.A. Marakhovskiy, L.A. Dykina, V.V. Fil, A.A. Panich

INVESTIGATION OF THE EFFECT OF IMPURITY PHASES OF THE FEEDSTOCK ON THE PROPERTIES OF FERROELECTRIC CERAMICS OF THE PZT SYSTEM

In the process of mass production of ferroelectric materials, impurities of various types and concentrations are periodically observed in the feedstock. The aim of the study was to determine the effect of impurity phases present in the feedstock on the dielectric and electrophysical properties of ferroelectric ceramics. In this work, the basic raw materials components included in the lead zirconate - titanate system for the presence of impurity components were studied by spectral analysis. The results revealed a group of impurity phases (Sb, Na, Bi, K, Fe) of different concentrations. The model object of the study was an industrially produced ferroelectric material with a perovskite structure and the chemical formula $Pb_{0,95}Sr_{0,05}(Zr_{0,53}Ti_{0,47})O_3$ *+ 1%* Nb_2O_5. *The objective of the study was the dosed introduction of impurity alloying additives into the composition of the initial ferroelectric material in order to possibly change the final properties. The study established the relevance of the dosed introduction of K and Na impurities at a concentration of 1-2 % into the PZT system in order to reduce the values of relative permittivity by 40-45 %.*

The dependences of the formed ceramic structure on the introduced impurity alloying phases have been established by scanning electron microscopy. The regularities of the "type of impurity additive – microstructure – properties" have been established. As a result of the study, the effectiveness of dosed administration of impurity alloying additives K and Na in order to modify the properties of ferroelectric ceramics of the PZT system was confirmed. Such impurity alloying leads to an increase in the values of the specific voltage sensitivity (g_{33}) to 34-37 mV·m/N. Ferroelectric materials of this format are of high practical interest for the creation of acoustic transducers operating in reception mode.

Ferroelectric material; feedstock; impurity phase; microstructure; specific stress sensitivity.

Введение. Процесс серийного изготовления сегнетоэлектрических материалов системы ЦТС зачастую сопряжен с непостоянством качества и квалификации исходного сырья. Такое непостоянство приводит к периодическим трудностям и требует анализа исходного сырья, отработки пробных партий сегнетоэлектрического материала, корректировки технологического процесса из-за присутствия различных примесных фаз в исходном сырье [1–6]. Происхождение примесей и их концентрация в исходном сырье чаще всего варьируются для фиксированного компонента от партии к партии. Наличие таких примесей способно существенно повлиять на конечные свойства сегнетоэлектрических материалов и керамик на их основе [7–11]. Кроме того, наличие примесных фаз зачастую приводит к повышению числа отбракованных керамических изделий при производстве, повышает трудовые-, временные- и энергозатраты.

Целью данной работы было исследование влияния примесных фаз ионов металлов в исходном сырье на электрофизические параметры сегнетоэлектрических материалов. Исследование проводилось с применением передового научного оборудования в условиях реального производства пьезокерамических материалов в «Институте высоких технологий и пьезотехники Южного федерального университета».

Основная часть. Спектральный анализ базового исходного сырья (PbO, ZrO_2, TiO_2) сегнетоэлектрических материалов системы цирконата-титаната свинца (ЦТС), выполненного на рентгенофлуоресцентном волнодисперсионном спектрометре (ARL OPTIM"X 200W), выявил группу примесных фаз (Sb, Cd, Co, Na, Bi, K, Fe) различной концентрации.

С целью исследования влияния на электрофизические параметры сегнетоэлектрических материалов ионов металлов примесных фаз исходного сырья, было проведено дозированное введение примесных добавок в состав исходного сегнетоэлектрического материала. Концентрация вводимых добавок составляла 0,5–2 %.

В качестве модельного объекта исследования был выбран сегнетоэлектрический материал системы ЦТС со структурой перовскита и номинальной химической формулой ($Pb_{0,95}Sr_{0,05}(Zr_{0,53}Ti_{0,47})O_3$ + 1% Nb_2O_5. Указанный состав характеризуется высокими значениями относительной диэлектрической проницаемости ($\varepsilon^T_{33}/\varepsilon_0$ ~ 1620-1980), относительно высокими значениями пьезоэлектрического модуля (d_{33} ~ 310-460 пКл/Н), приемлемой температурной стабильностью ($T_к$ ~ 290 ºС), а также высокой воспроизводимостью свойств от партии к партии [12–17].

Сегнетоэлектрический материал указанного состава был изготовлен твердофазным синтезом с помолом в планетарной шаровой мельнице Fritsch Pulverisette 6. Синтез сегнетоэлектрического материала и спекание керамических элементов на его основе проводились в камерной печи Nabertherm L5/13/P330.

Качество спекания керамических элементов оценивалось по результатам рентгенофазового анализа, снимкам микроструктуры, полученным на растровом электронном микроскопе (JEOL JSM-6390LA) и по значениям плотности полученными методом гидростатического взвешивания в дистиллированной воде (аналитические весы AND GH-300).

Микроструктура полученных керамических элементов представлена на рис. 1.

Дозированное введение примесей приводило к формированию пористой структуры керамики по сравнению с микроструктурой исходного состава. Такие изменения сопровождались модификацией комплекса свойств керамических элементов, представленных в табл. 1.

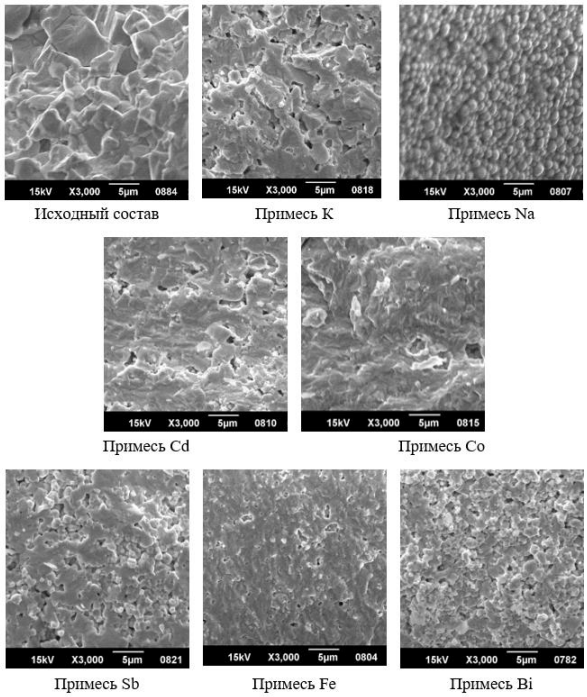


Рис. 1. Микроструктура керамических элементов полученных из исходного сегнетоэлектрического материала, а также с примесями

Таблица 1

Основные свойства исходного сегнетоэлектрического материала, а также с введением примесей

Сегнето-электрич. материал	$\varepsilon^T_{33}/\varepsilon_0$	d_{33}, пКл/Н	d_{31}, пКл/Н	g_{33}, мВ·м/Н	V_1^E, м/с
Исходный состав	1620-1980	310-460	150-200	21,6-26,3	2950-3150
Примесь K	726-970	240-320	81-120	37,3-37,4	2820-2910
Примесь Na	956-998	288-308	108-111	34,5-34,8	2860-2900
Примесь Cd	1180-1540	330-380	140-180	24,6-26,9	2970-2980
Примесь Co	912-941	188-196	78-82	20,6-20,8	3260-3280
Примесь Sb	1650-1885	300-420	155-185	18,1-22,2	2790-3000
Примесь Fe	1070-1110	50-70	20-25	4,6-6,3	3420-3430
Примесь Bi	1140-1170	240-255	100-110	21,0-21,9	3290-3350

Дозированное введение примесных фаз приводило к снижению основных свойств сегнетоэлектрического материала на 20–60 %. Наименьший вклад в изменение свойств оказало введение примесной фазы Sb (на 5 %). Из целого ряда полученных свойств, наибольший практический интерес представляли керамические элементы из исходного сегнетоэлектрического материала с примесями K и Na. Примесная фаза K привела к образо-

ванию большого числа пор, формируя менее плотную композиционную структуру керамики, способствующую увеличению анизотропии. Такого рода структура способствовала комплексному изменению свойств сегнетоэлектрической керамики и повышению значений удельной чувствительности по напряжению (g_{33}).

В случае дозированного введения примесной фазы Na, микроструктура керамики обладала зёрнами округлой формы с размерами 1-2 мкм, что в 5-8 раз меньше размеров зёрен керамики исходного состава. Однако, сокращение размеров зёрен приводит к снижению значений относительной диэлектрической проницаемости, поскольку увеличивается концентрация межзёренных границ, не обладающих сегнетоэлектрическими свойствами (рисунок 1). Такое изменение свойств отражается на повышении значений удельной чувствительности по напряжению (g_{33}) (табл. 1).

Данные полученные в процессе исследования соответствуют концентрации вводимых примесей 2 %. Концентрация дозировок вводимых примесных фаз определялась из результатов спектрального анализа исходного сырья. Вероятно, что с повышением концентраций примесей возможно формирование иной керамической структуры и новых конечных свойств.

Выводы. В результате исследования установлены зависимости влияния примесных фаз на диэлектрические и электрофизические свойства сегнетоэлектрической керамики системы ЦТС. Анализ полученных результатов показал актуальность дозированного введения примесей K или Na в концентрации 2 % с целью снижения значений относительной диэлектрической проницаемости на 40-45 % (до значений $\varepsilon/\varepsilon_0 = 720 – 970$). Это сопровождается повышением значений удельной чувствительности по напряжению (g_{33}) в 1,4 раза. Такой модифицированный состав сегнетоэлектрического материала может быть успешно использован в электронно-акустических преобразователях, работающих в режиме приёма [18–23].

Исследование выполнено при финансовой поддержке Министерства науки и высшего образования РФ: государственное задание в области научной деятельности, проект № FENW-2022-0033.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Haun M.J., Furman E., Jang S.J. et al.* Thermodynamic theory of $PbTiO_3$ // Journal of Applied Physics. – 1987. – 62 8. – P. 3331-3338.
2. *Малыхин А.Ю., Дыкина Л.А., Опарина А.Э. [и др.].* Патент № 2024620147 Российская Федерация. 11.01.2024.
3. *Bhatti H.S., Hussain S.T., Khan F.A., Hussain S.* Synthesis and induced multiferroicity of perovskite $PbTiO_3$; a review // Appl. Surf. Sci. – 2016. – 367, 291.
4. *Wight J.* Cellular Ceramics–Structure Manufacturing, Properties and Applications. – Wiley-VCH, Weinheim, 2005.
5. *Pabst W., et al.* Processing, microstructure, properties, applications and curvature-based classification schemes of porous ceramics / Alan Newton, ed. – NY, Nova Science Publishers Inc., 2017.
6. *Lee S.H. et al.* Fabrication of porous PZT-PZN piezoelectric ceramics with high hydrostatic figure of merits using camphene-based freeze casting // J. Am. Ceram. Soc. – 2007. – 90 (9). – 2807. – DOI: 10.1111/j.1551-2916.2007.01834.x.
7. *Zeng T. et al.* Effects of pore shape and porosity on the properties of porous PZT 95/5 ceramics // J. Eur. Ceram. Soc. – 2025 (2007). – 27 (4). – DOI: 10.1016/j.jeurceramsoc.2006.05.102.
8. IEEE standard on piezoelectricity, in ANSI/IEEE Std. – NY: The Institute of Electrical and Electronics Engineers Inc., 1988. – DOI: 10.1109/IEEESTD.1988.79638.
9. *Shvetsova N.A. et al*. Method of electromechanical characterization of ferroelectric materials, Ferroelectrics. – 2020. – 561 (1), 100. – DOI: 10.1080/00150193.2020.1736921.
10. *Aleshin V.A. et al.* Piezoelectric properties of layered bismuth-containing ferroelectric ceramics with a high degree of texture // J. Tech. Phys. – 1989. – 59, 152. – DOI: 10.1080/00150190211810.
11. *Третьяков Ю.Д.* Твердофазные реакции. – М.: Химия, 1978. – 360 с.
12. Материалы пьезокерамические. Технические условия (отраслевой стандарт) ОСТ 110444 – 87. – М., 1987.
13. *Фесенко Е.Г., Данцигер А.Я., Разумовская О.Н.* Новые пьезокерамические материалы. – Ростов-на-Дону: Изд-во Ростов. ун-та, 1983.

14. *Панич А.А., Мараховский М.А., Мотин Д.В.* Кристаллические и керамические пьезоэлектрики // Инженерный вестник Дона. – 2011. – № 1.
15. *Yamada H.* Pressureless sintering of PMN-PT ceramics // Journal of the european ceramic society. – 1999. – 19 (6-7). – P. 1053-1056.
16. *Lente M.H., Zanin A.L., Assis S.B., Santos I.A., Garcia D., Eiras J.A.* Ferroelectric domain dynamics in PMN-PT ceramics // Ferroelectrics. Gordon and Breach Science Publishers. – 2003. – 296. – P. 149-155.
17. *Liou Y.C.* Stoichiometric perovskite PMN-PT ceramics produced by reaction-sintering process // Materials science and engineering: B. – 2003. – 103 (3). – P. 281-284.
18. *Шарапов В.М., Минаев В.Г., Сотула Ж.В., Куницкая Л.Г.* Электроакустические преобразователи. – М.: Техносфера, 2013. – 296 с.
19. *Богуш М.В.* Пьезоэлектрическое приборостроение. Т. III. Пьезоэлектрические датчики для экстремальных условий эксплуатации. – Ростов-на-Дону: Изд-во СКНЦ ВШ, 2006. – 335 с.
20. *Белинкур Д., Керран Д., Жаффе Г.* Пьезоэлектрические и пьезомагнитные материалы и их применение в ультразвуке // Физическая акустика. Т. 1. Ч. А. / под ред. У. Мэзона.– М.: Мир, 1966. – С. 204-324.
21. *Ананьева А.А.* Керамические приемники звука. – М., 1963.
22. *Мараховский М.А., Мараховский В.А., Мирющенко Э.А., Панич Е.А.* Исследование возможности создания новых полифазных пьезоматериалов для гидроакустических преобразователей. ГА-2018. – СПб., 2018. – 616 с.
23. *Мараховский М.А., Нестеров А.А., Свирская С.Н., Панич А.Е.* Изменение механизма спекания порошков пьезоматериалов, как способ снижения температуры формирования плотных керамических каркасов // Инженерный вестник Дона. – 2013. – № 2.

REFERENCES

1. *Haun M.J., Furman E., Jang S.J. et al.* Thermodynamic theory of $PbTiO_3$, *Journal of Applied Physics,* 1987, 62 8, pp. 3331-3338.
2. *Malykhin A.Yu., Dykina L.A., Oparina A.E. [i dr.].* Patent № 2024620147 Rossiyskaya Federatsiya. 11.01.2024 [Patent No. 2024620147 Russian Federation. 11.01.2024].
3. *Bhatti H.S., Hussain S.T., Khan F.A., Hussain S.* Synthesis and induced multiferroicity of perovskite $PbTiO_3$; a review, *Appl. Surf. Sci*., 2016, 367, 291.
4. *Wight J.* Cellular Ceramics–Structure Manufacturing, Properties and Applications. Wiley-VCH, Weinheim, 2005.
5. *Pabst W., et al.* Processing, microstructure, properties, applications and curvature-based classification schemes of porous ceramics, Alan Newton, ed. NY, Nova Science Publishers Inc., 2017.
6. *Lee S.H. et al.* Fabrication of porous PZT-PZN piezoelectric ceramics with high hydrostatic figure of merits using camphene-based freeze casting, *J. Am. Ceram. Soc*., 2007, 90 (9), 2807. DOI: 10.1111/j.1551-2916.2007.01834.x.
7. *Zeng T. et al.* Effects of pore shape and porosity on the properties of porous PZT 95/5 ceramics, *J. Eur. Ceram. Soc.*, 2025 (2007), 27 (4). DOI: 10.1016/j.jeurceramsoc.2006.05.102.
8. IEEE standard on piezoelectricity, in ANSI/IEEE Std. NY: The Institute of Electrical and Electronics Engineers Inc., 1988. DOI: 10.1109/IEEESTD.1988.79638.
9. *Shvetsova N.A. et al*. Method of electromechanical characterization of ferroelectric materials, *Ferroelectrics,* 2020, 561 (1), 100. DOI: 10.1080/00150193.2020.1736921.
10. *Aleshin V.A. et al.* Piezoelectric properties of layered bismuth-containing ferroelectric ceramics with a high degree of texture, *J. Tech. Phys*., 1989, 59, 152. DOI: 10.1080/00150190211810.
11. *Tret'yakov Yu.D.* Tverdofaznye reaktsii [Solid-phase reactions]. Moscow: Khimiya, 1978, 360 p.
12. Materialy p'ezokeramicheskie. Tekhnicheskie usloviya (otraslevoy standart) OST 110444 – 87 [Piezoceramic materials. Technical conditions (industry standard) OST 110444 – 87]. Moscow, 1987.
13. *Fesenko E.G., Dantsiger A.Ya., Razumovskaya O.N.* Novye p'ezokeramicheskie materialy [New piezoceramic materials]. Rostov-on-Don: Izd-vo Rostov. un-ta, 1983.
14. *Panich A.A., Marakhovskiy M.A., Motin D.V.* Kristallicheskie i keramicheskie p'ezoelektriki [Crystalline and ceramic piezoelectrics], *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don], 2011, No. 1.
15. *Yamada H.* Pressureless sintering of PMN-PT ceramics, *Journal of the european ceramic society,* 1999, 19 (6-7), pp. 1053-1056.
16. *Lente M.H., Zanin A.L., Assis S.B., Santos I.A., Garcia D., Eiras J.A.* Ferroelectric domain dynamics in PMN-PT ceramics, *Ferroelectrics. Gordon and Breach Science Publishers,* 2003, 296, pp. 149-155.

17. *Liou Y.C.* Stoichiometric perovskite PMN-PT ceramics produced by reaction-sintering process, *Materials science and engineering: B,* 2003, 103 (3), pp. 281-284.
18. *Sharapov V.M., Minaev V.G., Sotula Zh.V., Kunitskaya L.G.* Elektroakusticheskie preobrazovateli [Electroacoustic transducers]. Moscow: Tekhnosfera, 2013, 296 p.
19. *Bogush M.V.* P'ezoelektricheskoe priborostroenie. T. III. P'ezoelektricheskie datchiki dlya ekstremal'nykh usloviy ekspluatatsii [Piezoelectric instrument making. Vol. III. Piezoelectric sensors for extreme operating conditions]. Rostov-on-Don: Izd-vo SKNTS VSh, 2006, 335 p.
20. *Belinkur D., Kerran D., Zhaffe G.* P'ezoelektricheskie i p'ezomagnitnye materialy i ikh primenenie v ul'trazvuke [Piezoelectric and piezomagnetic materials and their application in ultrasound], *Fizicheskaya akustika* [Physical acoustics], Vol. 1, Part A., ed. by U. Mezona. Moscow: Mir, 1966, pp. 204-324.
21. *Anan'eva A.A.* Keramicheskie priemniki zvuka [Ceramic sound receivers]. Moscow, 1963.
22. *Marakhovskiy M.A., Marakhovskiy V.A., Miryushchenko E.A., Panich E.A.* Issledovanie vozmozhnosti sozdaniya novykh polifaznykh p'ezomaterialov dlya gidroakusticheskikh preobrazovateley. GA-2018 [Study of the possibility of creating new polyphase piezoelectric materials for hydroacoustic transducers. GA-2018]. Saint Petersburg, 2018, 616 p.
23. *Marakhovskiy M.A., Nesterov A.A., Svirskaya S.N., Panich A.E.* Izmenenie mekhanizma spekaniya poroshkov p'ezomaterialov, kak sposob snizheniya temperatury formirovaniya plotnykh keramicheskikh karkasov [Changing the sintering mechanism of piezoelectric material powders as a way to reduce the formation temperature of dense ceramic frameworks], *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don], 2013, No. 2.

Статью рекомендовал к опубликованию д.т.н., профессор Б.Г. Коноплев.

Мараховский Михаил Алексеевич – Южный федеральный университет; e-mail: marmisha@mail.ru; г. Ростов-на-Дону; Россия; тел.: +79045015431; к.т.н.; начальник сектора.

Дыкина Любовь Александровна – e-mail: diykinala@mail.ru; тел.: +79043415495; начальник лаборатории.

Филь Вадим Владимирович – e-mail: vfil@sfedu.ru; тел.: +79889968173; м.н.с.

Панич Александр Анатольевич – e-mail: rctt.rctt@mail.ru; тел.: +79034888887; д.т.н.; директор Института высоких технологий и пьезотехники.

Marakhovskiy Mikhail Alekseevich – Southern Federal University; e-mail: marmisha@mail.ru; Rostov-on-Don, Russia; phone: +79045015431; cand. of eng. sc.; head of the sector.

Dykina Lyubov Alexandrovna – e-mail: diykinala@mail.ru; phone: +79043415495; head of the laboratory.

Fil Vadim Vladimirovich – e-mail: vfil@sfedu.ru; phone: +79889968173; junior researcher.

Panich Alexander Anatolievich – e-mail: rctt.rctt@mail.ru; phone: +79034888887; dr. of eng. sc.; director of Institute of High Technologies and Piezotechnics.

УДК 53.088.7 DOI 10.18522/2311-3103-2024-4-157-169

В.В. Сердюков, К.С. Коротков, А.В. Голан, А.Т. Маньшина, С.Е. Калюжная

ОПРЕДЕЛЕНИЕ РАЗНОСТИ ФАЗ В ОПТОЭЛЕКТРОННЫХ И СВЧ ИНТЕРФЕЙСНЫХ МОДУЛЯХ СИСТЕМ СВЯЗИ С МНОГОУРОВНЕВЫМИ ФОРМАТАМИ МОДУЛЯЦИИ

Целью исследования является расчет и проектирование устройства, измеряющего разности фаз сигналов, с возможностью приема команд управления и передачи результатов через высокоскоростной Ethernet интерфейс. Любой современный измерительный прибор оптического или сверхвысокочастотного (СВЧ) диапазона имеет в своей конструкции важный элемент, без которого невозможно ни одно измерение, а именно – векторный вольтметр, измеряющий сдвиг фаз и отношение амплитуд сигналов. Реализацией таких приборов практически не занимаются и такие разработки в основном являются интеллектуальной собственностью крупных компаний, поэтому проектирование и создание такого устройства в широкодоступном исполнении крайне необходимо и актуально. Рассмотрены современные форматы модуляции и реализация транспондеров по

передаче оптических сигналов с использованием многоуровневых форматов квадратурной манипуляции фазовым сдвигом с двойной поляризацией (DP QPSK) и 16-ти позиционной квадратурно-амплитудной модуляции с двойной поляризацией (DP 16QAM). Так же рассмотрены основные методы построения векторных вольтметров с использованием микроконтроллеров и field programmable gate arrays (FPGA), произведено моделирование оптических каналов связи и создано устройство измерения сдвига фаз. В результате исследования получен векторный вольтметр на FPGA, который, в свою очередь, может быть использован для создания установки по измерению сдвига фаз смесителей с подключением по Ethernet интерфейсу для передачи данных. Также на языке аппаратного программирования Verilog HDL для FPGA Altera Cyclone V составлен программный код для электронно-вычислительной машины (ЭВМ) по измерению разности фаз двух гармонических сигналов. Реализована программа на языке СИ для процессора ARM Cortex A9 в среде Quartus Prime Lite в составе сверхбольшой интегральной схемы (СБИС) Cyclone V, передающая результаты измерений в реальном времени по интерфейсу 1GB на ЭВМ с возможностью приема команд управления.

ПЛИС; FPGA; смесители; сдвиг фаз; СВЧ; оптоэлектроника; векторный вольтметр; СБИС.

V.V. Serdukov, K. S. Korotkov, A.V. Golan, A.T. Manshina, S.E. Kaliuzhnaya

RESEARCH OF THE PHASE DIFFERENCE IN OPTOELECTRONIC AND MICROWAVE INTERFACE MODULES OF COMMUNICATION SYSTEMS WITH MULTILEVEL MODULATION FORMATS

The purpose of the study is to calculate and design the device that measures the phase differences of signals, with the ability to receive control commands and transmit the results via a high-speed Ethernet interface. Any modern measuring device of the optical or ultrahigh frequency (microwave) range has an important element in its design, without which no measurement is possible, namely, a vector voltmeter that measures the phase shift and the ratio of signal amplitudes. Practically no one is engaged in the implementation of such devices and such developments are mainly the intellectual property of large companies, therefore, the design and creation of such a device in a widely available version is necessary. We have considered modern modulation formats and the implementation of transponders for the transmission of optical signals using multi-level formats of quadrature phase shift manipulation with double polarization (DP QPSK) and 16-position quadrature amplitude modulation with double polarization (DP 16QAM), as well as the basic methods for constructing vector voltmeters using microcontrollers and field programmable gate arrays (FPGA), optical communication channels were simulated and a phase shift measurement device was created. As a result of our research, we obtained a vector voltmeter on an FPGA, which, in turn, can be used to create an installation for measuring the phase shift of mixers connected via an Ethernet interface. Also, in the Verilog HDL hardware programming language for the Altera Cyclone V FPGA, a program code has been compiled for an electronic computing machine to measure the phase difference of two harmonic signals. A C program has been implemented for the ARM Cortex A9 processor in the Quartus Prime Lite environment as part of the Cyclone V ultra-large integrated circuit (VLSI), transmitting measurement results in real time over the 1GB interface to a computer with the ability to receive control commands.

FPGA; mixers; phase shift; microwave; optoelectronics; vector voltmeter; VLSI.

Введение. Экспоненциальный рост потребности современного общества в скорости и объеме получения информации повышает требования к помехоустойчивости, защищенности и высокой пропускной способности ВОЛС. Перед инфокоммуникационными компаниями стоит проблема увеличения пропускной способности в существующих магистральных и локальных сетях [1]. Особое внимание заслужили новые форматы модуляции, использование которых возможно при использовании современного стационарного оборудования. Однако внедрение новых форматов модуляции невозможно без разработки более совершенных интерфейсных модулей, реализующих эти форматы. Подобные модули способны измерять не только амплитуду, но и фазу сигналов. В оптических интерфейсах зачастую измерение разности фаз может производиться интерференционными методами на компонентах интерфейсов до использования АЦП. В радиосистемах, если они низкоскоростные, (обычно они реализуются с помощью FPGA), задача измерения фазового сдвига решается программным способом – формированием аппаратной части из

вентилей. Однако при скоростях выше 1 Гб/с достичь требуемых параметров измерения фазового сдвига возможно лишь с применением более сложных систем наподобие ASIC (интегральная схема целевого применения) и NPU (блок сетевой обработки).

В связи с этим, необходимо произвести проектирование структурной и функциональной схемы устройства, измеряющего разности фаз сигналов.

Современные форматы модуляции в оптических системах связи. Модуляция как способ преобразования информации используется очень давно, начиная с импульсно-кодовой модуляции, которая использовалась на витой паре, до различных фазовых и амплитудных манипуляций в оптоволокне [2].

На оптических линиях связи возможно использование следующих форматов модуляции:

- ASK – амплитудная;
- PSK – фазовая;
- FSK – частотная;
- DPSK – дифференциальная фазовая;
- BPSK – двоичная фазовая;
- QPSK – квадратурная фазовая;
- QAM – квадратурная амплитудная модуляция.

Чаще остальных используются ASK, PSK форматы ввиду малой стоимости и простоты реализации. При более высоких скоростях, выше 10 Гбит/с, в этих форматах модуляции появляются различные нелинейные эффекты, что неблагоприятно сказывается на коэффициенте битовых ошибок (BER) [3]. Поэтому увеличение скорости до 40 Гбит/с и выше требует внедрения новых, более эффективных и менее восприимчивых к искажениям форматов модуляции. В последнее время активно исследовались и внедрялись различные форматы модуляции. Особое внимание получили: BPSK, QPSK, QAM [4–9].

Моделирование оптического канала связи. Наибольшую функциональную нагрузку в подсистеме приемо-передачи сигналов ВОЛС выполняет транспондер, представляющей собой коммуникационное оборудование, выполняющее оптоэлектронное преобразование оптического сигнала [10]. Важной составляющей современных транспондеров является блок коррекции ошибок FEC [11, 12].

Для достижения до некоторой степени достоверных результатов при сравнении работы смоделированных оптоэлектронных схем необходимо приблизить параметры каждой систем к реальным насколько это возможно. Для достижения данной цели смоделирован оптический канал связи, который встроен в каждую смоделированную систему передачи.

Волоконная линия, состоящая из 4 регенерационных участков, представляет собой цепь из чередующихся волокон типа SMF и DCF. После каждого участка установлен соответствующий ему оптический усилитель для компенсации затухания. Общая длина канала связи составила 300 км.

В явном виде в моделях представлена только одна пара волокон, а требуемое количество участков достигнуто за счет использования зацикливания. Данная возможность программной среды позволяет не строить длинную цепь повторяющихся волокон (и другого оптоэлектронного оборудования), а представить только наименьшую единицу цепи и повторить ее произвольное количество. Характеристики используемых волокон представлены в табл. 1.

Таблица 1

Характеристики используемых волокон

Параметр	SMF-волокно	DCF-волокно
Затухание	$0{,}152\ \frac{\text{дБ}}{\text{км}}$	$0{,}235\ \frac{\text{дБ}}{\text{км}}$
Дисперсия	$20{,}2\ \frac{\text{пс}}{\text{нм}\cdot\text{км}}$	$-38\ \frac{\text{пс}}{\text{нм}\cdot\text{км}}$
Наклон дисперсии	$0{,}06\ \frac{\text{пс}}{\text{нм}^2\cdot\text{км}}$	$-0{,}12\ \frac{\text{пс}}{\text{нм}^2\cdot\text{км}}$
Коэффициент поляризационно-модовой дисперсии	$0{,}02\ \frac{\text{пс}}{\sqrt{\text{км}}}$	$0{,}05\ \frac{\text{пс}}{\sqrt{\text{км}}}$

Для фильтрации прошедшего волоконный тракт сигнала от шумов и помех перед приемником находится оптический фильтр с гауссовым профилем (Gaussian Optical Filter) с центральной длиной волны спектра 1550 нм, полосой пропускания 50 ГГц. После прохождения фильтра сигнал попадает на DP-QPSK или DP-16-QAM ресивер когерентного типа. Полная схема смоделированного канала связи изображена на рис. 1.

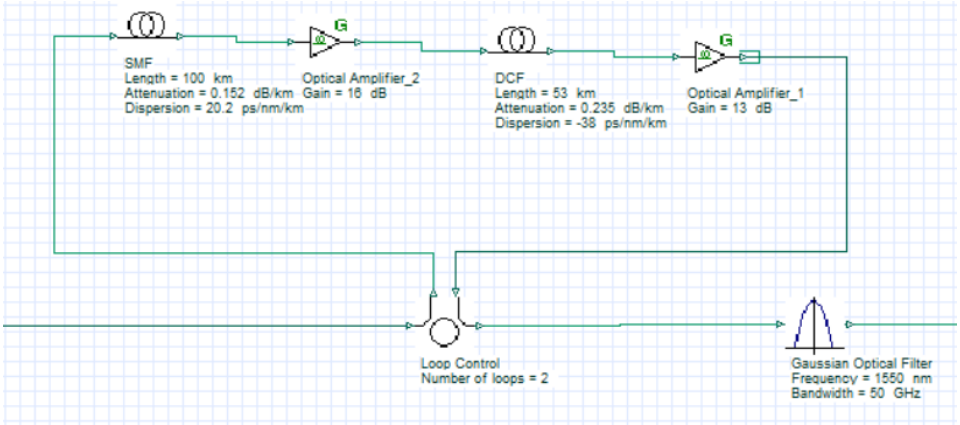
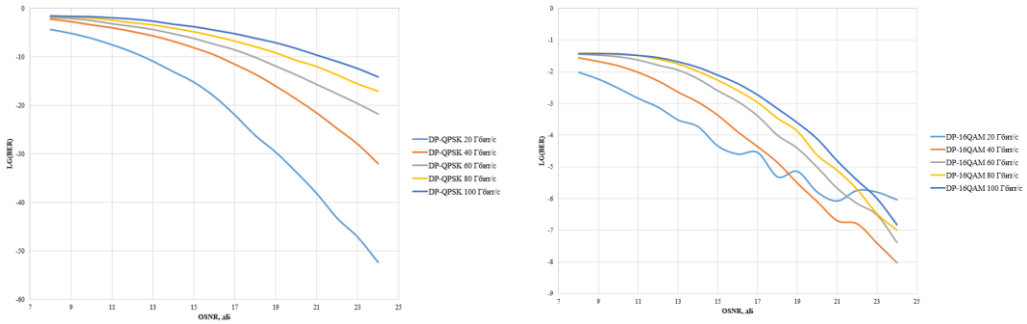


Рис. 1. Схема смоделированного канала связи

Для смоделированных оптоэлектронных схем проведена эмуляция работы в среде OptiSystem, результатом которой является совокупность численных результатов передачи, а именно Q-фактора и EVM.

Каждая схема подверглась передаче 65536 бит информации с различными скоростями передачи данных. Скорость передачи варьировалась в промежутке от 20 до 100 Гбит/с (с шагом в 20 Гбит/с). Параметр OSNR изменялся от 8 до 24 дБ с единичным шагом. В общей сложности произведено 85 вычислительных итераций. В результате каждой итерации сняты характеристики получившихся сигнальных созвездий, на основе полученных данных, был вычислен коэффициент битовых ошибок. После сбора всех полученных данных были построены соответствующие графики зависимости коэффициента битовых ошибок от соотношения «сигнал-шум» для обоих исследуемых форматов модуляции. Построенные графики представлены на рис. 2.



а б

Рис. 2. Зависимость BER от OSNR: а – при DP-QPSK-модуляции; б – при DP-16QAM-модуляции

На основе рис. 2 можно сделать следующие выводы:

1. При использовании формата DP-QPSK проявляется линейная зависимость BER от OSNR на участке от 14 до 20 дБ, при формате DP16QAM линейность проявляется меньше.

2. У DP-QPSK-формата скорость изменения коэффициента битовых ошибок значительно выше, чем у формата DP-16QAM. К примеру, на скорости передачи данных 20 Гбит/с при переходе OSNR от 21 до 23 дБ наблюдается изменение BER на 10 порядков при DP-QPSK передаче, а при DP-16QAM передаче всего лишь на порядок.

Цифровые методы измерения фазы сигнала. В 2008 году компания Agilent Technologies объявила об изобретении нелинейного векторного анализатора цепей и создании новой системы X-параметров [13–17]. Помимо нового подхода к измерению активных устройств, работающих в нелинейных режимах (например, усилителей), данная технология открыла и новые возможности для измерения параметров смесителей. В 2009 году в университете Вайкато, г. Гамильтон, Новая Зеландия [18], а затем и в компании Agilent Technologies, США [19, 20], была разработана новая методика измерения параметров устройств с преобразованием частоты, в том числе и их сдвигов фаз, которая основана на применении генератора комбинационных частот – генератора гармоник.

В настоящее время разработано большое количество методов измерения фазы с использованием различных математических алгоритмов, реализуемых как программно на микроконтроллерах, так и на FPGA [21,22].

Схема измерителя представлена на рис. 3. Сигналы от источников S1 и S2 поступают в два идентичных канала, преобразуются в цифровой вид с помощью 14-разрядного АЦП и передаются в FPGA.

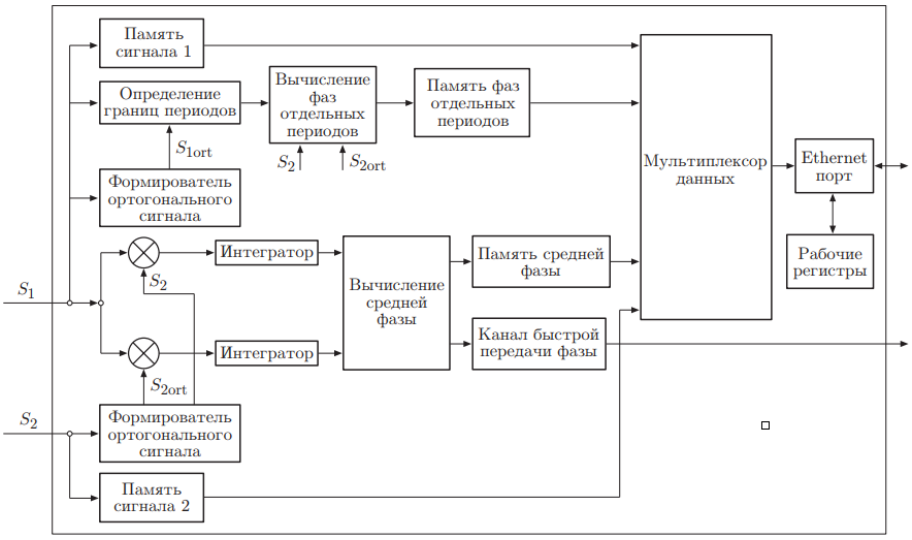


Рис. 3. Схема измерителя фаз на основе FPGA

В основе процесса измерения фазы лежит квадратурное синхронное детектирование. Для опорного сигнала S_2 задаётся ортогональная компонента S_{2ort}. Формируется эта компонента путём прохождения сигнала S_2 через цепочку регистров, запись в каждый последующий регистр осуществляется с задержкой, равной периоду тактовой частоты FPGA. Компонента S_{2ort} создаётся на основе сигналов S_{2i} и S_{2j}, взятых с выходов регистров i, j. При вычислении S_{2ort} используется линейная интерполяция.

Вычисление квадратурных компонент F_{sin} и F_{cos} (1), (2):

$$F_{sin} = \int_0^{nT} S_1 S_2 dt = \frac{nT}{2} \sin(\varphi_0), \quad (1)$$

$$F_{cos} = \int_0^{nT} S_1 S_{2ort} dt = \frac{nT}{2} \cos(\varphi_0), \quad (2)$$

где φ_0 – разность фаз между первыми гармониками S1 и S2; n – число периодов сигнала, в течение которого осуществляется интегрирование.

Далее разность фаз φ_0 находится как арктангенс отношения Fsin/Fcos. Арктангенс вычисляется с использованием итерационного алгоритма CORDIC [23, 24].

Подбор компонентов и выбор схемы реализации устройства, измеряющего фазу сигнала. На основе проведенного анализа методов модулирования несущей, используемых в современных системах связи, а также устройства интерфейсных модулей, непосредственно формирующих передаваемый сигнал, установлено, что:

- в интерфейсном модуле должны производиться измерение амплитуды и фазы сигнала;
- для увеличения скорости передачи данных интерфейсные модули, как правило, представляют собой либо специальные логические интегральные схемы (ASIC), либо программируемые пользователем вентильные матрицы (FPGA).

Задачами разрабатываемого устройства являются:

- оцифровка двух синхронизированных гармонических радиосигналов частотой 278 кГц;
- вычисление в реальном времени разности фаз и отношения амплитуд этих сигналов;
- передача обработанных данных на ПК по протоколу Ethernet.

Исходя из предъявляемых задач, можно сформировать требования к компонентам и схеме реализации разрабатываемого устройства, измеряющего фазу сигнала:

- наличие двух высокоскоростных АЦП большой разрядности;
- высокая скорость обработки данных;
- возможность проведения калибровок устройства, необходимых при изменении параметров измеряемых сигналов;
- наличие Ethernet модуля, поддерживающего работу со стандартами не ниже Fast Ethernet;
- устройство должно быть реализовано без использования навесного монтажа, для уменьшения влияния внешних помех.

Исходя из задач и требований программная обработка, реализованная на микроконтроллере, признана нецелесообразной из-за больших задержек при обработке и меньшей производительности микроконтроллеров по сравнению с аппаратными средствами. Так как для немассовых устройств применение ASIC является нецелесообразным по различным причинам, в первую очередь из-за высокой сложности разработки. Поэтому выбран СБИС с программируемой логикой Cyclone V от компании Intel [27].

Наиболее подходящей под требования является плата ADC-SoC от компании Terasic, показанная на рис. 4, где обозначены как элементы, подключенные к FPGA, так и к микропроцессору (обозначен как HPS) [28].

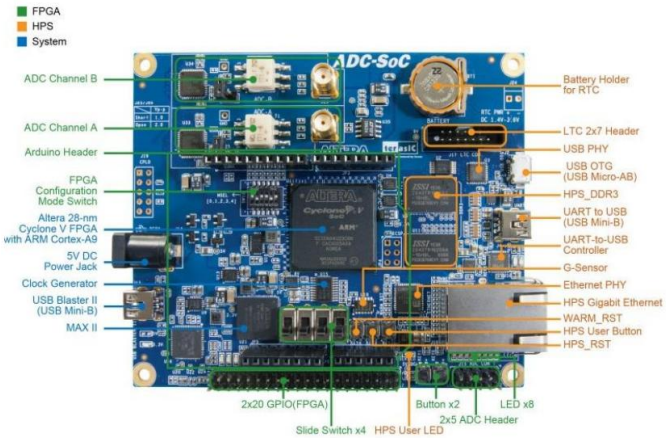


Рис. 4. Плата ADC-SoC

На основе проведенного анализа и с учетом выбранных компонентов обоснованы структурная и функциональная схемы, показанные на рис. 5.

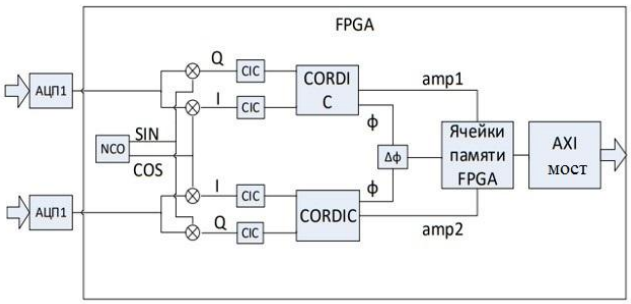


Рис. 5. Функциональная схема блока математической обработки сигналов

Особенностью данного чипа является то, что на одном кристалле расположены и FPGA, и микропроцессор ARM Cortex A9, имеющие возможность обмена данными с помощью шины AXI, что позволит разделить процессы обработки данных и их передачи [29].

Как видно из рисунка 5 оцифрованные сигналы подаются на цифровые смесители, где перемножаются с ортогональными колебаниями, сгенерированными NCO. На выходах смесителей для каждого входного сигнала фактически будут наблюдаться действительная и мнимая компоненты.

Далее подключаются CIC-фильтры, пропускающие только сигнал с частотой равной разности частот входного сигнала и гетеродина. Вычисление фазы и амплитуды каждого сигнала выполняется с помощью алгоритма CORDIC. Вычисление фазы может проводиться с точностью до десятых и сотых долей градуса. После чего рассчитывается разность фаз двух гармонических сигналов и отношение их амплитуд.

Проведя анализ сигналов, FPGA записывает полученные данные в ячейки памяти, к которым посредством AXI Bridge имеет доступ микропроцессор ARM Cortex A9 [28]. Микропроцессор, с установленной операционной системой на базе Linux, проводит расчет отношения амплитуд сигналов, переводит значение в градусы, формирует кадры протокола Ethernet и передает их на физический уровень.

Разработка на языке Verilog блока обработки измеряемых сигналов. Задача блока NCO, используя в качестве входного сигнала только тактирующие импульсы, генерировать гармонические колебания частотой 278 кГц. Чтобы получить нужную частоту сигнала на выходе, необходимо чтобы на входе конвейера CORDIC по каждому тактовому сигналу с шины «clk» по определенному алгоритму изменялась фаза. Программный код составлен в Quartus Prime Lite версия 18.0, являющейся частью дистрибутива Intel FPGA.

Для работы реализуемого алгоритма необходим мультипликативный преобразователь частоты. На рисунке 6 показана структурная схема цифрового преобразователя частоты. Отдельно произведён расчет и программная реализация CIC фильтра

Расчет и генерация программного кода CIC фильтра выполнялись в демонстрационной версии среды MATLAB 2021 с использованием расширения Filter Builder. Решено для расчета фазы использовать сигнал, являющийся разностью частот измеряемого сигнала и гетеродина. Таким образом, сигнал на выходе фильтра при неизменной амплитуде и разности фаз сигналов будет представлять собой некоторой постоянное значение, описываемое выражением $s(t) = \cos(\varphi)$, где φ – разность фаз измеряемого сигнала и гетеродина. Такая реализация значительно снижает требования к фильтру, т.к. позволяет подобрать минимум первого лепестка АЧХ равный сумме частот преобразуемых сигналов, не учитывая минимумы последующих лепестков.

С учетом описанных выше особенностей наиболее эффективен для применения в разрабатываемой программе децимирующий CIC фильтр, т.к. «прореживание» выходной последовательности данных при условии, что на выходе некоторое значение, изменяющее в небольших пределах и невысокой скоростью, позволяет дополнительно снизить нагрузку на аппаратную составляющую.

Также программно реализован алгоритм CORDIC, вычисляющий фазу и амплитуду сигналов. Наиболее значимым элементом модуля cordic_phase является таблица углов поворота. Таблица углов задается внутри функции с использованием поведенческого блока case.

Модуль rotator_phase является ядром алгоритма CORDIC – именно данный модуль непосредственно выполняет операции сдвига. Объединенные в конвейер модули rotator_phase последовательным приближением вычисляют значения амплитуды и фазы каждого из измеряемых сигналов. Программный код модулей cordic_phase и rotator_phase имеет объем 78 и 53 строки соответственно.

Задачей модуля PHASE_METER является объединение описанных выше блоков и модулей в одну структуру, показанную на рис. 7.

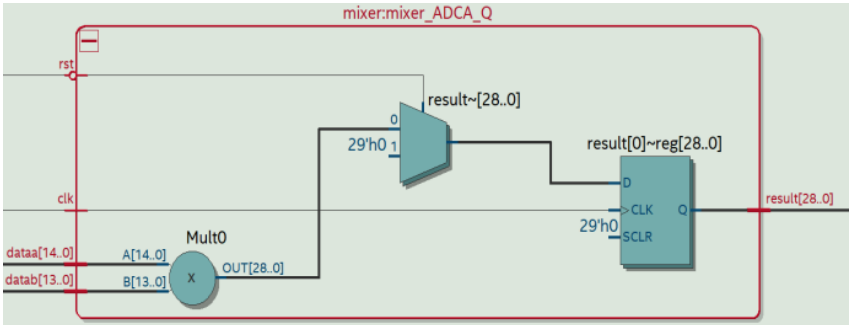


Рис. 6. Структурная схема цифрового преобразователя частоты

На структурной схеме кроме описанных выше модулей также присутствуют дополнительные регистры.

- input clk – тактовый сигнал;
- input rst – сигнал сброса;
- input[14:0] data – вход гетеродина, разрядность увеличена на 1 бит для передачи сигналов с учетом знака;
- input [13:0] datab – вход, соединяемый с АЦП.

Единственный выходной сигнал данного модуля – result [28:0]. Они необходимы для корректной передачи сигналов между модулями – отсутствие этих связующих регистров приводит к полной невозможности передачи данных следующему модулю.

В разрабатываемой системе ведущими блоками являются IP-ядра Terasic_AD9254 (с подключенным IP-ядром PHASE_METER) и Arria V/Cyclone V Hard Processor System, ведомыми, соответственно, On-Chip Memory (RAM or ROM) Intel FPGA IP и PHASE_METER. Задачей блоков Terasic_AD9254 (исходный код IP-ядра открыт) является управление АЦП и прием оцифрованных сигналов для дальнейшей обработки (блок подключен к физическим входам/выходам чипа).

С помощью Lightweight HPS-to-FPGA в FPGA из микропроцессора передаются команды на начало измерений и информация о количестве отсчетов АЦП, которые нужно принять и обработать. Все интерфейсы Avalon Mapped-Memory данного блока являются ведущими. Схема соединений IP-ядер, составлена с использованием среды системной интеграции Platform Designer.

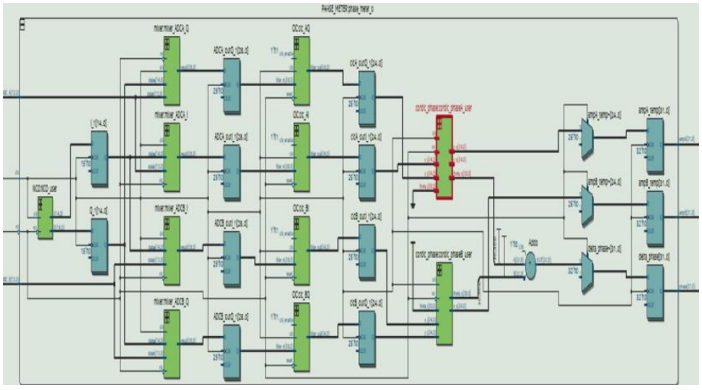


Рис. 7. Структурная схема модуля PHASE_METER

Для работы системы, необходимо чтобы в блоки PHASE_METER и Terasic_AD9254 поступили команды на начало измерений. Источником команд является микропроцессор, передающий команды с помощью моста Lightweight HPS-to-FPGA. Работа микропроцессора при взаимодействиях с FPGA определяется приложением, скомпилированным под ОС Linux и запускаемым из файлового пространства операционной системы. Смонтированный образ ОС Linux записан на флэш-карту, входящую в комплект поставки платы Terasic ADC-SoC. Работа через операционную систему избавляет от необходимости написания программного кода, реализующего передачу информации по Ethernet интерфейсу, т.к. уже содержит все необходимые драйверы. Погрешность измерена на собранном тестовом стенде, состоящем из генератора сигнала и созданное устройства. Для проведения измерений частота выходного сигнала на выходе генератора равнялась 278 кГц. Проведены две серии измерений: в первой – при постоянном отношении амплитуд изменялась разность фаз сигналов в диапазоне от -180^0 до 180^0, во второй – при постоянной разности фаз изменялась амплитуда сигналов (отношение амплитуд сигналов на выходе генератора изменялась от -12 до 12 дБ с шагом 1 дБ). Погрешность измерения для разностей фаз и отношений амплитуд соответственно разработанного устройства не превышают ±0,02 дБ и ±0,1 дБ.

Заключение. Рассчитано и спроектировано устройство, измеряющее разность фаз сигналов, с возможностью приема команд управления и передачи результатов через высокоскоростной Ethernet интерфейс.

Для достижения цели, в инженерном программном обеспечении OptiSystem проведено моделирование работы транспондеров DP-QPSK и DP-16QAM форматов модуляции. На основе моделирования проанализированы полученные характеристики работы линии длиной 300 км на одной из несущих DWDM систем передачи при скорости передачи данных от каналообразующего оборудования до 100 Гбит/с, а также отношение сигнал-шум канала передачи в пределах от 8 до 24 дБ. В результате анализа определено, что требуемая полоса пропускания канала DWDM относительно формата NRZ при DP-QPSK модуляции уменьшается в 4 раза относительно потока каналообразующего оборудования, а при DP-16QAM модуляции – в 8 раз. Таким образом, для передачи потока в 100 Гбит/с DP-QPSK требует полосы пропускания канала 25 Гбит/с, а DP-16QAM – 12.5 Гбит/с.

Для применения в стенде по измерению фазовой задержки гетеродинного преобразования составлены структурная и функциональная схемы устройства, измеряющего разность фаз двух гармонических сигналов заданной частоты 278 кГц [29].

На языке аппаратного программирования Verilog для процессов, обрабатываемых FPGA, составлена программа объемом 818 строк для реализации функциональной схемы устройства, измеряющего разность фаз двух гармонических сигналов [30]. Также на языке программирования СИ составлена программа объемом 180 строк для микропро-

цессора ARM Cortex A9, входящего в состав СБИС Cyclone V. С помощью как готовых, так и самостоятельно созданных IP-ядер собрана логическая схема, для программирования FPGA.

На основании эксперимента определены относительные погрешности разработанного устройства, не превышающие ±0,02 и ±0,1 дБ при измерении разности фаз сигналов и отношения их амплитуд соответственно.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Наний О.Е., Трещиков В.Н.* Анализ форматов модуляции для систем DWDM со скоростью 40 Гбит/с // Вестник связи. – 2012. – № 1.
2. *Филиппов А.В., Мальцева Н.С.* Особенности развития систем DWDM // Технические средства систем управления и связи: Матер. V Международной конференции «Информационные технологии и технические средства управления». – Астрахань: Астраханский государственный технический университет, 2021. – С. 57-60.
3. *Величко М.А., Наний О.Е., Сусьян А.А.* Новые форматы модуляции в оптических системах связи // LIGHTWAVE Russian Edition. – 2005. – № 4.
4. *Листвин В.Н., Трещиков В.Н.* DWDM системы. – М.: Техносфера, 2015. – 296 с.
5. *Голиков А.М.* Модуляция, кодирование и моделирование в телекоммуникационных системах. Теория и практика. – СПб.: Лань, 2018. – 419 с.
6. *Лучковский Д.П., Сухарев Е.Н., Куклина А.И.* Квадратурная амплитудная модуляция // Молодежь. Общество. Современная наука, техника и инновации. – 2016. – № 15. – С. 50-51.
7. *Li M., Hong W., Zhang X., Li S., Song W., Han C., Zhu M., Li W., Huang D.* Transparent wavelength conversion for a novel orthogonal FSK/IM modulation format at 40-Gb/s based on FWM effect of SOA // Proceedings of SPIE. – 2008. – No. 12.
8. *Гуркин Н.В., Наний О.Е., Трещиков В.Н.* Оптические когерентные DWDM системы связи с канальной скоростью 100 Гбит/с // T-Comm: Телекоммуникации и транспорт. – 2013. – № 4.
9. *Леонов А.В., Наний О.Е., Трещиков В.Н.* Совершенствование форматов модуляции в оптических системах связи DWDM // Первая миля. – 2019. – № 1. – С. 30-37.
10. *Проскурин А.А., Близнюк В.И., Ивкин В.Б.* Реализационные аспекты разработки отечественного транспондера когерентных волоконно-оптических систем передачи со скоростью 100 Гбит/с // Известия Тульского государственного университета. Технические науки. – 2020. – № 2. – С. 64-69.
11. *Гуркин Н.В., Трещиков В.Н., Новиков А.Г., Наний О.Е.* Российское DWDM оборудование с канальной скоростью 100 Гбит/с // T-COMM: Телекоммуникации и транспорт. – 2012. – Т. 6, № 4. – С. 65-67.
12. *Juliano S., Martins F., Prata A., Lopes S.* High Performance Microwave Point-to-Point Link for 5G Backhaul with Flexible Spectrum Aggregation // 2015 IEEE MTT-S International Microwave Symposium, 17–22 May 2015. – Phoenix, USA, 2015.
13. *Root D.E.* A New Paradigm for Measurement, Modeling, and Simulation of Nonlinear Microwave and RF Components // Berkeley Wireless Research Center. Presentation. Agilent Technologies. – Apr. 17, 2009. – – URL: http://www.learningace.com/doc/3086134/0867637b6f4bb9900348a7da473cf467/bwrc_root_april09.
14. *Root D.E., Horn J., Betts L., Gillease C., Verspecht J.* X–parameters: The new paradigm for measurement, modeling, and design of nonlinear RF and microwave components // Microwave Engineering Europe. – December 2008. – P. 16-21.
15. *Vye D.* Fundamentally Changing Nonlinear Microwave Design // Microwave Journal. – 2010. – Vol. 53, No. 3. – P. 22-38.
16. *Николаев Е.* X–параметры эффективный инструмент для анализа электрических цепей // Электроника. Наука. Технология. Бизнес. – 2012. – № 1 (00115). – С. 7679.
17. *Саяпин В.Ю.* Описание нелинейных цепей на основе X–параметров и методика их измерения // Доклады ТУСУРа. – 2012. – № 2 (26). – Ч. 1. – С. 83-86.
18. *Scott J., Hoy M.* Group–Delay Measurement of Frequency–Converting Devices using a Comb Generator // IEEE Transactions on Instrumentation and Measurement. – 2010. – Vol. 59, Issue 11. – P. 3012-3017.
19. *Павлов Л.А., Первова Н.В.* Структуры и алгоритмы обработки данных: учебник. – 2-е изд., испр. и доп. – С.Пб.: Лань, 2020. – 256 с.
20. *Баженов В.Г., Богдан Г.А., Кравченко М.В.* Цифровая система измерения фазовых сдвигов радиоимпульсных сигналов // Международный научно-исследовательский журнал. – 2016. – № 4. – С. 36-38.

21. *Naik R., Nadaf R.* Sine-Cosine Computation Using CORDIC Algorithm // International Journal of Advanced Research in Computer and Communication Engineering. – 2015. – Vol. 4, Issue 9.
22. *Jack Volder*. The CORDIC computing technique // In Papers presented at the the March 3-5, 1959, western joint computer conference (IRE-AIEE-ACM '59 (Western)). Association for Computing Machinery, New York, NY, USA, 1959. – P. 257-261. – https://doi.org/10.1145/1457838.1457886.
23. *Garrido M., Källström P., Kumm M. and Gustafsson O.* CORDIC II: A New Improved CORDIC Algorithm // in IEEE Transactions on Circuits and Systems II: Express Briefs. – Feb. 2016. – Vol. 63, No. 2. – P. 186-190. – DOI: 10.1109/TCSII.2015.2483422.
24. *Ifeachor E.C., Jervis B.W.* Digital Signal Processing: A Practical Approach. – 2nd ed. – Prentice Hall, – 2001. – 933 p. – ISBN 9785845907103.
25. *Matthew P.D.* CIC Filter Introduction. – URL: https://www.dspguru.com/files/cic.pdf (дата обращения: 20.05.2024).
26. Cyclone V Device Overview // Intel: official site. – 2018. – URL: https://www.intel.com/ content/dam/www/programmable/us/en/pdfs/literature/hb/cyclone-v/cv_51001.pdf (дата обращения: 12.04.2024).
27. ADC-SoC User Manual // Intel: official site. – 2016. – URL: https://www.intel.com/content/ dam/altera-www/global/en_US/portal/dsn/42/docus-dsnbk-42-2108361005542-adc-soc-user-manual-dsn.pdf (дата обращения: 06.05.2024).
28. *Jaymin P., Yash S., Lili H.* Design between AXI Lite and AHB Bus Protocol // Journal of Physics: Conference Series. – 2021. – Vol. 1993.
29. Патент № 2618046 Российская Федерация, МПК G 01 R 25/00. Способ измерения разности фаз и отношения уровней двух гармонических сигналов: № 2016109376: заявл. 15.03.2016: опубл. 10.05.2017, Бюл. №13 / Бабенко А.А., Гноевой А.В., Коротков К.С., Левченко А.С., Фролов Д.Р.; заявитель и патентообладатель Федеральное государственное бюджетное образовательное учреждение высшего образования "Кубанский государственный университет". – 12 с.
30. Свидетельство о государственной регистрации программы для ЭВМ №2022660602 Российская Федерация. Программа для определения комплексного коэффициента передачи и отражения СВЧ-смесителей методом суммы и разности: №2022619520: заявл. 26.05.2022: опубл. 07.06.2022 / Коротков К.С., Большов А.В., Гайденко И.В., Пожидаев Р.Б., Сердюков В.В.; правообладатель Федеральное государственное бюджетное образовательное учреждение высшего образования «Кубанский государственный университет» (ФГБОУ ВО «КубГУ»). Зарегистрировано в Реестре программ для ЭВМ.

REFERENCES

1. *Naniy O.E., Treshchikov V.N.* Analiz formatov modulyatsii dlya sistem DWDM so skorost'yu 40 Gbit/s [Analysis of modulation formats for DWDM systems with a speed of 40 Gbit/s], *Vestnik svyazi* [Vestnik svyazi], 2012, No. 1.
2. *Filippov A.V., Mal'tseva N.S.* Osobennosti razvitiya sistem DWDM [Features of the development of DWDM systems], *Tekhnicheskie sredstva sistem upravleniya i svyazi: Mater. V Mezhdunarodnoy konferentsii «Informatsionnye tekhnologii i tekhnicheskie sredstva upravleniya»* [Technical means of control and communication systems: Proceedings of the V International Conference "Information Technologies and Technical Means of Control"]. Astrakhan': Astrakhanskiy gosudarstvennyy tekhnicheskiy universitet, 2021, pp. 57-60.
3. *Velichko M.A., Naniy O.E., Sus'yan A.A.* Novye formaty modulyatsii v opticheskikh sistemakh svyazi [New modulation formats in optical communication systems], *LIGHTWAVE Russian Edition,* 2005, No. 4.
4. *Listvin V.N., Treshchikov V.N.* DWDM sistemy [DWDM systems]. Moscow: Tekhnosfera, 2015, 296 p.
5. *Golikov A.M.* Modulyatsiya, kodirovanie i modelirovanie v telekommunikatsionnykh sistemakh. Teoriya i praktika [Modulation, coding and modeling in telecommunication systems. Theory and practice]. Saint Petersburg: Lan', 2018, 419 p.
6. *Luchkovskiy D.P., Sukharev E.N., Kuklina A.I.* Kvadraturnaya amplitudnaya modulyatsiya [Quadrature amplitude modulation], *Molodezh'. Obshchestvo. Sovremennaya nauka, tekhnika i innovatsii* [Youth. Society. Modern science, technology and innovation], 2016, № 15, pp. 50-51.
7. *Li M., Hong W., Zhang X., Li S., Song W., Han C., Zhu M., Li W., Huang D.* Transparent wavelength conversion for a novel orthogonal FSK/IM modulation format at 40-Gb/s based on FWM effect of SOA, *Proceedings of SPIE,* 2008, No. 12.
8. *Gurkin N.V., Gurkin O.E. Naniy V.N.* Treshchikov Opticheskie kogerentnye DWDM sistemy svyazi s kanal'noy skorost'yu 100 Gbit/s [Optical coherent DWDM communication systems with a channel rate of 100 Gbit/s], *T-Comm: Telekommunikatsii i transport* [T-Comm: Telecommunications and transport], 2013, No. 4.

9. *Leonov A.V., Naniy O.E., Treshchikov V.N.* Sovershenstvovanie formatov modulyatsii v opticheskikh sistemakh svyazi DWDM [Improving modulation formats in DWDM optical communication systems], *Pervaya milya* [First Mile], 2019, No. 1, pp. 30-37.
10. *Proskurin A.A., Bliznyuk V.I., Ivkin V.B.* Realizatsionnye aspekty razrabotki otechestvennogo transpondera kogerentnykh volokonno-opticheskikh sistem peredachi so skorost'yu 100 Gbit/s [Implementation aspects of the development of a domestic transponder for coherent fiber-optic transmission systems with a speed of 100 Gbit/s], *Izvestiya Tul'skogo Gosudarstvennogo Universiteta. Tekhnicheskie nauki* [Bulletin of the Tula State University. Technical sciences], 2020, No. 2, pp. 64-69.
11. *Gurkin N.V., Treshchikov V.N., Novikov A.G., Naniy O.E.* Rossiyskoe DWDM oborudovanie s kanal'noy skorost'yu 100 Gbit/s [Russian DWDM equipment with a channel rate of 100 Gbit/s], *T-COMM: Telekommunikatsii i transport* [T-COMM: Telecommunications and transport], 2012, Vol. 6, No. 4, pp. 65-67.
12. *Juliano S., Martins F., Prata A., Lopes S.* High Performance Microwave Point-to-Point Link for 5G Backhaul with Flexible Spectrum Aggregation, *2015 IEEE MTT-S International Microwave Symposium, 17–22 May 2015*. Phoenix, USA, 2015.
13. *Root D.E.* A New Paradigm for Measurement, Modeling, and Simulation of Nonlinear Microwave and RF Components, *Berkeley Wireless Research Center. Presentation. Agilent Technologies,* Apr. 17, 2009. Available at: http://www.learningace.com/doc/3086134/0867637b6f4bb9900348a7da473cf467/bwrc_root_april09
14. *Root D.E., Horn J., Betts L., Gillease C., Verspecht J.* X–parameters: The new paradigm for measurement, modeling, and design of nonlinear RF and microwave components, *Microwave Engineering Europe,* December 2008, pp. 16-21.
15. *Vye D.* Fundamentally Changing Nonlinear Microwave Design, *Microwave Journal,* 2010, Vol. 53, No. 3, pp. 22-38.
16. *Nikolaev E.* X–parametry effektivnyy instrument dlya analiza elektricheskikh tsepey [X-parameters are an effective tool for analyzing electrical circuits], *Elektronika. Nauka. Tekhnologiya. Biznes* [Electronics. Science. Technology. Business], 2012, No. 1 (00115), pp. 7679.
17. *Sayapin V.Yu.* Opisanie nelineynykh tsepey na osnove X–parametrov i metodika ikh izmereniya [Description of nonlinear circuits based on X-parameters and methods of their measurement], *Doklady TUSURa* [Reports of TUSUR], 2012, No. 2 (26), Part 1, pp. 83-86.
18. *Scott J., Hoy M.* Group–Delay Measurement of Frequency–Converting Devices using a Comb Generator, *IEEE Transactions on Instrumentation and Measurement,* 2010, Vol. 59, Issue 11, pp. 3012-3017.
19. *Pavlov L.A., Pervova N.V.* Struktury i algoritmy obrabotki dannykh: uchebnik [Structures and algorithms of data processing: textbook]. 2nd ed. Saint Petersburg: Lan', 2020, 256 p.
20. *Bazhenov V.G., Bogdan G.A., Kravchenko M.V.* Tsifrovaya sistema izmereniya fazovykh sdvigov radioimpul'snykh signalov [Digital system for measuring phase shifts of radio pulse signals], *Mezhdunarodnyy nauchno-issledovatel'skiy zhurnal* [International research journal], 2016, No. 4, pp. 36-38.
21. *Naik R., Nadaf R.* Sine-Cosine Computation Using CORDIC Algorithm, *International Journal of Advanced Research in Computer and Communication Engineering,* 2015, Vol. 4, Issue 9.
22. *Jack Volder*. The CORDIC computing technique, *In Papers presented at the the March 3-5, 1959, western joint computer conference (IRE-AIEE-ACM '59 (Western)).* Association for Computing Machinery, New York, NY, USA, 1959, pp. 257-261. Available at: https://doi.org/10.1145/1457838.1457886.
23. *Garrido M., Källström P., Kumm M. and Gustafsson O.* CORDIC II: A New Improved CORDIC Algorithm, *in IEEE Transactions on Circuits and Systems II: Express Briefs,* Feb. 2016, Vol. 63, No. 2, pp. 186-190. DOI: 10.1109/TCSII.2015.2483422.
24. *Ifeachor E.C., Jervis B.W.* Digital Signal Processing: A Practical Approach. 2nd ed. Prentice Hall, 2001, 933 p. ISBN 9785845907103.
25. *Matthew P.D.* CIC Filter Introduction. Available at: https://www.dspguru.com/files/cic.pdf (accessed 20 May 2024).
26. Cyclone V Device Overview, *Intel: official site,* 2018. Available at: https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/hb/cyclone-v/cv_51001.pdf (accessed 12 April 2024).
27. ADC-SoC User Manual, *Intel: official site,* 2016. Available at: https://www.intel.com/content/dam/altera-www/global/en US/portal/dsn/42/docus-dsnbk-42-2108361005542-adc-soc-user-manual-dsn.pdf (accessed 06 May 2024).
28. *Jaymin P., Yash S., Lili H.* Design between AXI Lite and AHB Bus Protocol, *Journal of Physics: Conference Series,* 2021, Vol. 1993.
29. *Babenko A.A., Gnoevoy A.V., Korotkov K.S., Levchenko A.S., Frolov D.R.* Patent № 2618046 Rossiyskaya Federatsiya, MPK G 01 R 25/00. Sposob izmereniya raznosti faz i otnosheniya urovney dvukh garmonicheskikh signalov: № 2016109376: zayavl. 15.03.2016: opubl. 10.05.2017, Byul.

No. 13; zayavitel' i patentoobladatel' Federal'noe gosudarstvennoe byudzhetnoe obrazovatel'noe uchrezhdenie vysshego obrazovaniya "Kubanskiy gosudarstvennyy universitet" [Patent No. 2618046 Russian Federation, IPC G 01 R 25/00. Method for measuring the phase difference and the level ratio of two harmonic signals: No. 2016109376: decl. 15.03.2016: publ. 10.05.2017, Bull. No. 13; applicant and patent holder Federal State Budgetary Educational Institution of Higher Education "Kuban State University"], 12 p.

30. *Korotkov K.S., Bol'shov A.V., Gaydenko I.V., Pozhidaev R.B., Serdyukov V.V.* Svidetel'stvo o gosudarstvennoy registratsii programmy dlya EVM №2022660602 Rossiyskaya Federatsiya. Programma dlya opredeleniya kompleksnogo koeffitsienta peredachi i otrazheniya SVCh-smesiteley metodom summy i raznosti: №2022619520: zayavl. 26.05.2022: opubl. 07.06.2022; pravoobladatel' Federal'noe gosudarstvennoe byudzhetnoe obrazovatel'noe uchrezhdenie vysshego obrazovaniya «Kubanskiy gosudarstvennyy universitet» (FGBOU VO «KubGU»). Zaregistrirovano v Reestre programm dlya EVM [Certificate of state registration of computer program No. 2022660602 Russian Federation. Program for determining the complex transmission and reflection coefficient of microwave mixers using the sum and difference method: No. 2022619520: declared 05/26/2022: published 06/07/2022; copyright holder Federal State Budgetary Educational Institution of Higher Education "Kuban State University" (FGBOU VO "KubSU"). Registered in the Register of Computer].

Статью рекомендовал к опубликованию д.т.н., профессор К.Е. Румянцев.

Сердюков Владимир Владимирович – Кубанский государственный университет; e-mail: serdyukov-vladimir1996@mail.ru; г. Краснодар, Россия; тел.: 89528524156; преподаватель.

Коротков Константин Станиславович – e-mail: korsar@phys.kubsu.ru; тел.: 89182513055; д.т.н.; профессор.

Голан Антон Васильевич – e-mail: blazepro17@gmail.com; тел.: 89528668695; студент.

Маньшина Александра Тарасовна – e-mail: eteriliro@mail.ru; тел.: 89183596321; студентка.

Калюжная Софья Евгеньевна – e-mail: kaliuzhnaya.sophia.evgenevna@gmail.com; тел.: 89676747515; студентка.

Serdukov Vladinir Vladimirovih – Kuban State University; e-mail: serdyukov-vladimir1996@mail.ru; Krasnodar, Russia; phone: +79528524156; teacher.

Korotkov Konstantin Stanislavovich – e-mail: korsar@phys.kubsu.ru; phone: +79182513055; dr. of eng. sc.; professor.

Golan Anton Vasil'evich – e-mail: blazepro17@gmail.com; phone: +79528668695; student.

Manshina Aleksandra Tarasovna – e-mail: eteriliro@mail.ru;phone: +79183596321; student.

Kaliuzhnaya Sof'ya Evgen'evna – e-mail: kaliuzhnaya.sophia.evgenevna@gmail.com; phone: +79676747515; student.

УДК 621.396.677 DOI 10.18522/2311-3103-2024-4-169-181

К.М. Занин, Д.Д. Габриэльян, Ю.В. Кузнецов, С.Е. Мищенко

ПРОФИЛИРОВАНИЕ РЕФЛЕКТОРА ЗЕРКАЛЬНОЙ АНТЕННЫ ДЛЯ ФОРМИРОВАНИЯ ЗАДАННОГО КОНТУРА ГЛАВНОГО ЛЕПЕСТКА ДИАГРАММЫ НАПРАВЛЕННОСТИ

При построении комплексов спутниковой связи требуется обеспечение заданного уровня коэффициента усиления бортовой антенны в строго определенной области покрытия и значительно более низкого уровня вне этой области. Граница области покрытия при этом может иметь сложную, но неизменную в течение всего периода эксплуатации форму. Для удовлетворения данных требований применяются, в том числе, зеркальные антенны с профилированным рефлектором, в которых закон профилирования поверхности описывается гладкими аналитическими функциями. Однако при формировании контурной диаграммы направленности с более сложной формой сечения главного лепестка требуемое фазовое распределение может иметь разрывы при переходе через период 2π, которые не могут быть аппроксимированы гладкими функциями без искажений.

В данном случае известные подходы к профилированию рефлекторов зеркальных антенн не позволяют синтезировать диаграмму направленности с заданным качеством. Целью работы являлось построение рефлектора зеркальной антенны с одиночным облучателем, обеспечивающей формирование диаграммы направленности с заданными параметрами. Для достижения поставленной цели решены следующие задачи: 1. Разработка алгоритма определения профиля рефлектора зеркальной антенны с учетом требуемой формы границы обслуживаемой территории и заданного закона распределения коэффициента усиления; 2. Проведение численных исследований по построению профиля рефлектора. В ходе проведенных исследований разработан алгоритм определения профиля рефлектора зеркальной антенны, который позволяет формировать на апертуре антенны распределение поле, соответствующее диаграмме направленности с требуемыми параметрами. Для этого выполнен расчет распределения поля на плоскости, и синтезирована поверхность рефлектора на основании результатов расчета. Проведенные численные исследования подтвердили возможность построения зеркальной антенны, обеспечивающей формирование диаграммы направленности с заданными параметрами.

Зеркальная антенна; амплитудно-фазовое распределение; контурная диаграмма направленности.

K.M. Zanin, D.D. Gabrielyan, Yu.V. Kuznetsov, S.E. Mishenko

SHAPING OF CONTOURED-BEAM ANTENNA MAIN LOBE BY PROFILING OF REFLECTOR ANTENNA

In satellite communication complexes, it is required to ensure a given level of the gain of the space antenna in a given serviced area. A lower level of gain beyond this area is also required. The boundary of the coverage area may have a complex shape that does not change during the operation of the communication system. To meet these requirements, reflector antennas with a profiled reflector are used. The law of profiling the reflector surface is described by smooth analytical functions. However, when forming a contour lobe with a more complex shape, the required phase distribution may have discontinuities during the transition through the period 2π. These gaps cannot be eliminated by smooth functions without distortion. In this case, the known approaches to profiling reflectors of reflector antennas do not allow obtaining a radiation pattern with a given quality. The goal of the work was constructing a reflector of a reflector antenna, which provides the formation of a radiation pattern with specified parameters. To achieve this goal, the following tasks have been solved: 1. Development of an algorithm for profiling the reflector of a reflector antenna, taking into account the shape of the boundary of the serviced area and taking into account the given law of distribution of the gain; 2. Conducting numerical simulations on the construction of the reflector profile. In the course of the research, an algorithm has been developed that allows you to obtain the reflector profile of a reflector antenna. This reflector antenna generates a field distribution at the aperture corresponding to the radiation pattern with the required parameters. To do this, the calculation of the field distribution on the plane was performed, and the surface of the reflector was synthesized based on the calculation results. Numerical simulations have confirmed the possibility of constructing a reflector antenna that forms a radiation pattern with specified parameters.

Reflector antenna; amplitude-phase distribution; contoured main lobe; radiation pattern.

Введение. В ряде случаев при построении комплексов спутниковой связи требуется обеспечение заданного коэффициента усиления (КУ) бортовой антенны в строго определенной области покрытия и значительно более низкого значения КУ вне этой области, при котором ресурсы радиоканала системы связи расходуются незначительно. Граница области покрытия при этом может иметь сложную, но неизменную в течение всего периода эксплуатации форму. Следствием этого является требование к реализации диаграммы направленности (ДН) с соответствующей формой сечения главного лепестка. Для удовлетворения данных требований наиболее широкое распространение получили четыре основных варианта построения антенных систем спутниковой связи: активные фазированные антенные решетки (АФАР), гибридно-зеркальные антенны (ГЗА), зеркальные антенны (ЗА) с плоским рефлектором и ЗА с профилированным рефлектором.

АФАР позволяют реализовать требуемую ДН за счет управления комплексным коэффициентом передачи в каждом канале. Данные антенные системы имеют возможность оперативно перестраивать ДН и формировать различные рабочие области покрытия проектируемой системы связи. Однако имеющиеся возможности АФАР могут быть невос-

требованными в задачах, в которых область покрытия остается неизменной. Для таких задач наиболее обоснованным с технической точки зрения решением является использование ЗА или ГЗА [1–4], а также ЗА с плоским рефлектором, покрытым слоем диэлектрика [5–7].

ГЗА позволяют реализовать ДН со сложным контуром границы рабочей области при меньшей сложности и стоимости антенной системы в сравнении с АФАР. Однако конструкция облучающей системы может оказаться достаточно громоздкой, что приводит к снижению энергетических характеристик антенны.

В свою очередь, ЗА с одиночным облучателем и плоским рефлектором имеют упрощенную конструкцию облучателя. Реализация требуемого фазового распределения на апертуре достигается путем покрытия рефлектора набором отражающих элементов, каждый из которых обеспечивает требуемый коэффициент отражения по фазе [8–10]. К недостаткам данного варианта построения антенны можно отнести сниженный срок службы из-за деградации параметров диэлектрической подложки отражающих элементов.

В работах [11–14] рассмотрены вопросы применения принципов адаптивной оптики для профилирования металлического рефлектора ЗА с одиночным облучателем. Профиль поверхности при этом обеспечивает формирование требуемой ДН за счет изменения длин электрических путей сигналов, распространяющихся от фазового центра облучателя ЗА до апертуры антенны. Полученный закон профилирования поверхности рефлектора описывается гладкими аналитическими функциями, например, полиномами Цернике. Однако применительно к формированию ДН с более сложной формой сечения главного лепестка распределение требуемых разностей длин путей, определяемых распределением фаз на апертуре (ФР), может иметь более сложный вид. Фазовое распределение может иметь разрывы при переходе через период 2π, которые не могут быть устранены с помощью алгоритмов развертывания фазы. В данном случае описанные подходы не позволяют синтезировать поверхность рефлектора ЗА, формирующей ДН с заданным качеством.

Таким образом, для систем связи, в которых требуется обеспечение заданного ее качества в строго определенной области покрытия со сложной, но неизменной формой границы актуальной является задача поиска технического решения по построению антенны, свободного от выше отмеченных недостатков.

Целью работы является построение рефлектора ЗА с одиночным облучателем, обеспечивающей формирование ДН с заданными параметрами.

Решаемые задачи:

1. Разработка алгоритма определения профиля рефлектора ЗА с учетом требуемой формы границы обслуживаемой территории и заданного закона распределения КУ;
2. Проведение численных исследований по построению требуемого профиля рефлектора.

1. Разработка алгоритма. Решение первой задачи настоящей работы выполнено путем разделения ее на следующие основные этапы:

- синтез амплитудно-фазового распределения поля (АФР) на плоскости апертуры ЗА и определение исходной формы рефлектора по его результатам;
- уточнение ФР на плоскости апертуры ЗА с учетом АФР, которое формируется облучателем;
- расчет формы рефлектора ЗА, которая реализует требуемое АФР на апертуре.

Данная последовательность этапов может быть обобщена блок-схемой, приведенной на рис. 1. Далее ее этапы рассмотрены более подробно.

Для пошагового контроля и наглядной демонстрации промежуточных результатов расчета при рассмотрении предлагаемого алгоритма его реализация проведена на примере ДН со сложной формой главного лепестка.

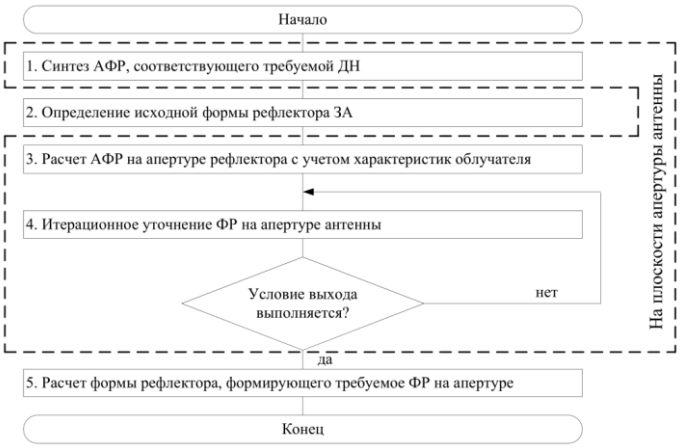


Рис. 1. Блок-схема предлагаемого алгоритма

В качестве исходных данных в настоящей работе были приняты следующие требования:

- задан рабочий сектор углов θ в сферической системе координат;
- установлен контур границы рабочей области ДН, например, показанный на рис. 2;
- задана ДН облучателя ЗА;
- установлены ограничения на максимальный размер апертуры и рабочая частота.

Синтез АФР, соответствующего требуемой ДН. На сегодняшний день наиболее распространены следующие методы синтеза АФР:

- метод парциальных диаграмм [15];
- метод синтеза антенн по заданной форме амплитудной ДН [16, 17];
- метод аппроксимационного синтеза [18] и др.

Приведенные методы хорошо себя показывают и широко применяются в многоканальных системах.

Для первого этапа синтеза применен упрощенный метод парциальных диаграмм. Данный метод позволяет синтезировать АФР с учетом координат расположения элементов, на которые делится апертура антенны, рабочей частоты и требуемой формы ДН. Расчет выполнен при равномерном шаге расположения элементов. Взаимное расположение парциальных лучей показано на рис. 3. Угловое расстояние между максимумами лучей выбрано равным половине ширины луча по уровню «нулей». Ширина таких лучей соответствует электрическим размерам апертуры антенны при условии равномерного возбуждения. Расчет выполнен в сферической системе координат. Для этого с учетом равномерного распределения парциальных лучей в заданной области определены угловые положения каждого p-го парциального луча, где $p = 0, 1, 2, \dots P - 1$.

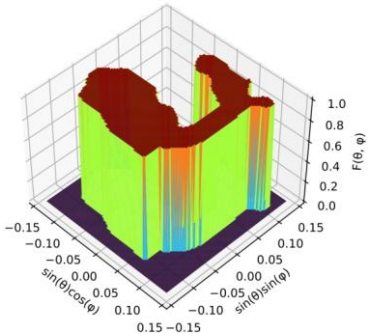


Рис. 2. Форма области покрытия требуемой ДН

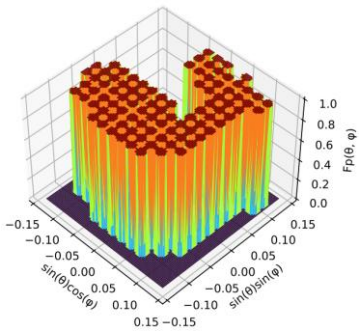


Рис. 3. Взаимное расположение парциальных лучей

Значения АФР на плоскости в дискретных точках могут быть вычислены следующим образом

$$J_{n,m} = \sum_{p} F\left(\theta_p, \varphi_p\right) exp\left[ikd\left(n\,sin\left(\theta_p\right) cos\left(\varphi_p\right) + m\,sin\left(\theta_p\right) sin\left(\varphi_p\right)\right)\right], \quad (1)$$

где n, m – номер строк и столбцов элементов апертуры;

θ, φ – координаты сферической системы координат;

$F(\theta, \varphi)$ – требуемая ДН;

k – волновое число;

d – период расположения элементов;

$J_{n,m}$ – значения комплексных амплитуд поля вектора E на плоскости.

Полученные распределения амплитуд и фаз приведены на рис. 4 и 5 соответственно. Данное АФР использовано в качестве начального приближения для последующих этапов.

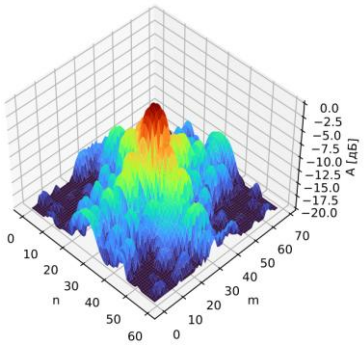


Рис. 4. Распределение амплитуд, полученное методом парциальных ДН

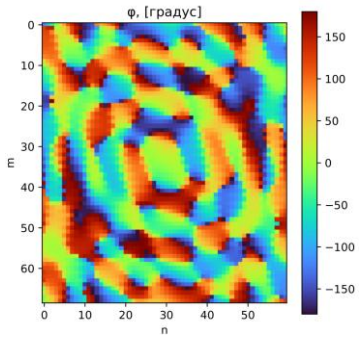


Рис. 5. Распределение фаз, полученное методом парциальных ДН

Определение исходной формы рефлектора ЗА. В качестве облучателя ЗА используется антенна с известной ДН. Геометрические параметры облучателя определяются дополнительными требованиями, такими как массогабаритные характеристики, требования по согласованию с питающей линией и т.д.

Амплитудное распределение (АР), приведенное на рис. 4, является сложно реализуемым в ЗА, поэтому оно заменяется более гладким. При этом уточняются параметры рефлектора из условия обеспечения требуемого спадания АР к краям его апертуры. В общем случае в подобных задачах требуемое спадание амплитуды к краям рефлектора может отличаться от оптимального, которое для ЗА составляет 12 дБ согласно [19]. Поэтому выбор расстояния от поверхности рефлектора до облучателя с учетом параметров облучателя является необходимым.

В качестве начального приближения выбран вариант конструкции ЗА, у которой форма рефлектора представляет собой поверхность параболоида вращения, а ось облучателя совмещена с осью параболоида. Середина раскрыва облучателя совмещена с точкой фокуса параболоида (рис. 6).

Как следует из рис. 4 максимум АР на поверхности рефлектора такой системы должен находиться в районе вершины параболоида, а также с учетом того, что габаритные размеры рефлектора заданы (его диаметр составляет $D = 36\lambda$, где λ – длина волны на средней рабочей частоте), может быть вычислено фокусное расстояние, при котором будет обеспечено требуемое спадание АР. Согласно АР, приведенному на рис. 4, его спадание к краям параболоида должно составлять $\Delta F \approx 14$ дБ. Тогда необходимое фокусное расстояние f может быть вычислено из системы уравнений (2).

$$\begin{cases} Fr(\theta) = -\Delta F \\ tg(\theta) = \dfrac{2fD}{4f^2 - 0{,}25D^2} \end{cases}, \quad (2)$$

где $Fr(\theta)$ – нормированная ДН рупорного облучателя.

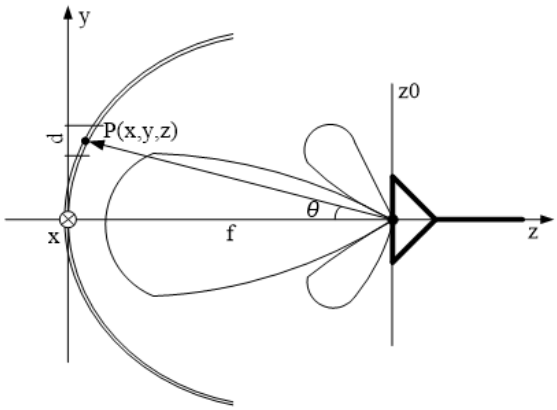


Рис. 6. Постановка задачи определения основных геометрических параметров ЗА и учета поля ее облучателя

В рассматриваемой ЗА $f \approx 21\lambda$.

Вычисленное расстояние f и известная ДН рупорного облучателя позволяют перейти к расчету АФР на апертуре рефлектора ЗА.

Расчет АФР на апертуре рефлектора с учетом характеристик облучателя. В качестве АР в точках апертуры рефлектора приняты значения ДН облучателя в соответствующих направлениях θ, φ. При расчете распределения фаз была использована фазовая ДН облучателя, которая задана относительно середины раскрыва, что позволило учесть смещение фазового центра облучателя относительно плоскости его апертуры. Полученные результаты расчета АФР приведены на рис. 7 и 8.

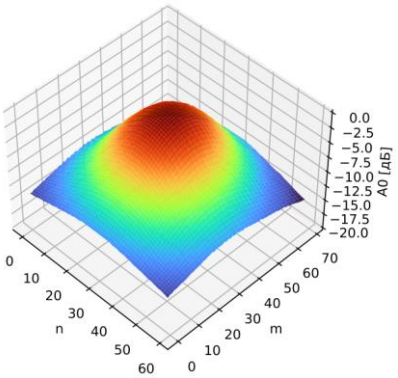


Рис. 7. Распределение амплитуд на апертуре рефлектора

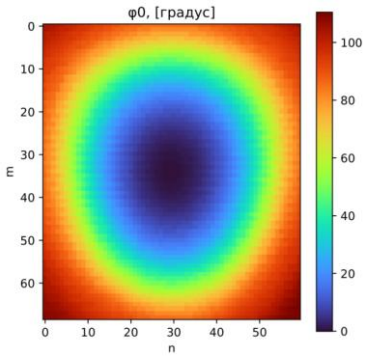


Рис. 8. Распределение фаз на апертуре рефлектора

Сравнивая АР, приведенные на рис. 4 и 7, можно сделать вывод о том, что АР, полученное методом парциальных диаграмм $A_{n,m}$, и АР, формируемое облучателем $A0_{n,m}$, имеют максимумы поля в середине апертуры, а величина спадания поля к краям сопоставима.

Распределение фаз на апертуре ЗА при условии, что фазовый центр облучателя совмещен с точкой фокуса параболоида вращения, должно быть равномерным. Однако из-за смещения фазового центра ФР на апертуре рефлектора не является равномерным, как показывает график на рис. 8 [20]. Для компенсации данного эффекта необходимо изменить форму рефлектора или сместить положение облучателя.

Итерационное уточнение ФР на апертуре антенны. На данном этапе при выбранном АР выполняется итерационное уточнение ФР на апертуре рефлектора. Для решения данной задачи применен метод синтеза ФР на плоскости по заданной амплитудной ДН при условии, что АР является известным [21]. Синтез выполнен в дискретных точках. Оси координат и координаты узлов поверхности использованы такие же, что при расчете методом парциальных диаграмм на первом шаге рассматриваемого алгоритма. В качестве заданной амплитудной ДН принята ДН, соответствующая требуемой области покрытия (см. рис. 2). В качестве АР использованы результаты, полученные для облучателя ЗА (рис. 7). В качестве начального ФР взяты результаты, полученные методом парциальных ДН (рис. 5). Тогда начальное приближение АФР имеет вид

$$B_{n,m} = A0_{n,m}\exp\,(i\varphi_{n,m}). \tag{4}$$

Исходная ДН $F(\theta,\varphi)$ разделена на дискретные участки, где нумерация отсчетов определяется индексами q и w, в которых определены соответствующие значения $F(\theta_q,\varphi_w)$. Итерационный метод синтеза реализован следующим образом. Вычислено очередное приближение к ДН и АФР по формулам:

$$\begin{gathered} F1_{q,w} = \sum_n \sum_m B_{n,m}\exp\,\Big[ikd\big[n\,sin(\theta_q)\,cos(\varphi_w) + m\ sin(\theta_q)\,sin\,(\varphi_w)\big]\Big], \\ C_{q,w} = \frac{F1_{q,w}}{|F1_{q,w}|}, \\ J1_{n,m} = \sum_q \sum_w F(\theta_q,\varphi_w)C_{q,w}\exp\Big[-ikd\big[n\,sin(\theta_q)\,cos(\varphi_w) + \\ + m\,sin(\theta_q)\,sin(\varphi_w)\big]\Big]. \end{gathered} \tag{5}$$

В (5) получено уточненное АФР, которое формирует ДН близкую к $F(\theta,\varphi)$. Распределение $|J1_{n,m}|$ в общем случае не совпадает с $A0_{n,m}$ и может быть по прежнему сложно реализуемым в рассматриваемой ЗА. Для уточнения АФР и приближения $|J1_{n,m}|$ к $A0_{n,m}$ переопределены исходные данные к новой итерации по формуле:

$$B_{n,m} = A0_{n,m}\,exp\big[i\,arg\big(J1_{n,m}\big)\big]. \tag{6}$$

Далее выполнены повторные итерации вычислений по формулам (5), при которых на каждом шаге уточнялись исходные данные для каждой следующей итерации по формуле (6). Критерием выхода из итерационного процесса служит условие, изложенное в [22].

После применения описанного итерационного подхода к решаемой задаче получено уточнено ФР и соответствующая ДН. Результаты приведены на рис. 9 и 10. Полученная ДН формируется плоской апертурой с равномерным шагом расположения элементов, в котором АР имеет вид $A0_{n,m}$ (см. рис. 8), а ФР имеет вид $arg\big(J1_{n,m}\big) = \Phi0_{n,m}$.

Полученное АФР необходимо реализовать на апертуре ЗА. Для этого выполнен переход к построению профилированного рефлектора, в котором учтено смещение фазового цента облучателя относительно середины его раскрыва путем внесения искажений в ФР, приведенное на рис. 9. Для этого вычислено

$$\Phi_{n,m} = \Phi0_{n,m} + \varphi0_{n,m}, \tag{7}$$

где $\varphi0_{n,m}$ – ФР на апертуре ЗА (см. рис. 8).

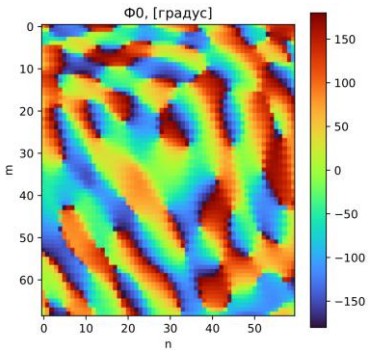


Рис. 9. Уточненное распределение фаз в апертуре ЗА

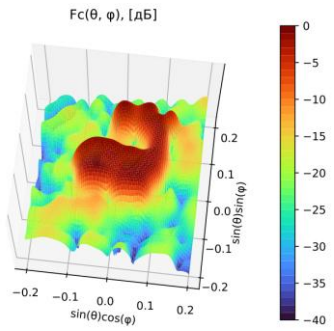


Рис. 10. ДН, соответствующая синтезированному АФР

Расчет поверхности рефлектора, формирующего требуемое ФР на апертуре. Как видно из рис. 9 полученное ФР имеет большое количество незамкнутых линий, где осуществляется переход от минус 180° до плюс 180° у соседних элементов апертуры. Следовательно, к такому распределению не может быть применен алгоритм сглаживания ФР [23] без существенных искажений формируемой ДН. Поэтому в данной работе для построения профиля рефлектора ЗА исходная форма рассматриваемого параболического рефлектора разделена на участки в соответствии с выбранным шагом расположения элементов d (см. рис. 6). Для формирования на апертуре требуемого ФР вида $\Phi_{n,m}$ выполнено смещение участков рефлектора P в направлении оси исходной фигуры. Для этого вычислена необходимая величина разности путей распространения сигнала, отраженного от поверхности рефлектора при заданной длине волны λ:

$$\Delta L_{n,m} = \lambda \frac{\Phi_{n,m}}{2\pi}. \tag{8}$$

Смещение выполнено таким образом, чтобы сохранить основное свойство параболоида – способность отражать лучи, падающие из фокуса, параллельно его оси. В противном случае, если для элемента поверхности точка фокуса будет вынесена из точки f, это повлечет за собой искажения в ФР, формируемом на апертуре антенны.

Для получения уравнения параболоида с учетом смещения $\Delta L_{n,m}$ вычислена новая форма параболической поверхности, у которой все локально параболические участки имеют точку фокуса, совпадающую с точкой f. Для этого рассмотрено уравнение параболы, построенной в координатах y, z и смещенной по z. Такое уравнение имеет вид: $z(y) = ay^2 + c$. Точка фокуса имеет координату $z_f = \frac{1}{4a} + c$. Пусть плоскость z_0 пересекает ось z в точке f (см. рис. 6). Тогда геометрическая длина пути от точки f до точки параболы с координатами y, z и от точки параболы до плоскости z_0 будет равна $L = {}^{1}/_{2a}$. Тогда, если первое уравнение параболы имеет коэффициент при старшей степени a_1, а второе – a_2, то рассматриваемая разность хода лучей имеет вид $\Delta L = \frac{1}{2a_1} - \frac{1}{2a_2}$. Отсюда искомый коэффициент при старшей степени с учетом $f = {}^{1}/_{4\mathrm{a}}$ может быть вычислен как

$$a_2 = \frac{1}{4f + 2\Delta L}. \tag{9}$$

С учетом (8) и (9) и с учетом длины пути распространения волны от фазового центра облучателя до поверхности рефлектора и от поверхности рефлектора до плоскости апертуры форма параболоида вращения, реализующего требуемое ФР в апертуре, имеет вид

$$z_{n,m} = \frac{{x_n}^2 + {y_m}^2}{4f + 2\Delta L_{n,m}} - \frac{\Delta L_{n,m}}{2}. \tag{10}$$

Уравнение (10) в данной работе определяет форму поверхности рефлектора, формирующего требуемое ФР на апертуре.

2. Проведение численных исследований. Объемная модель полученной ЗА приведена на рис. 11.

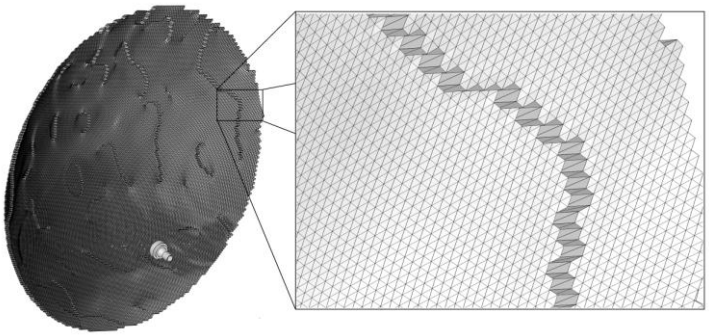


Рис. 11. Объемная модель ЗА

Для объемной модели, приведенной на рис. 11, методом электродинамического моделирования рассчитаны ДН и КНД. Результаты приведены на рис. 12.

Полученные данные показывают повторяемость результатов математического расчета (см. рис. 10) и результатов моделирования (рис. 12). Сравнение результатов приведено в табл. 1.

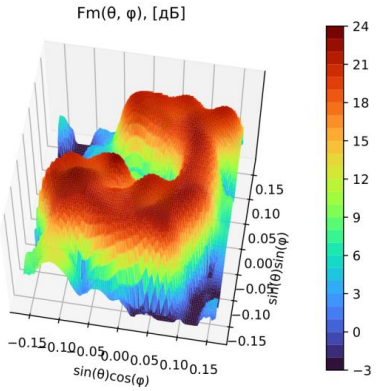


Рис. 12. ДН, полученная путем электродинамического моделирования ЗА

Таблица 1

Результаты расчета и моделирования ЗА

Исследуемый параметр	Максимум КНД в области $F(\theta, \varphi)$, [дБ]	Минимум КНД в области $F(\theta, \varphi)$, [дБ]	Средний КНД в области $F(\theta, \varphi)$, [дБ]	Область, где КНД не хуже минус 3 дБ от макс., [%]
Результаты математического расчета	24.4	21,3	23,1	97
Результаты электродинамического моделирования	23,2	20	21,9	94,7

Формируемая ЗА ДН имеет максимальный уровень боковых лепестков за пределами требуемой области минус 7 дБ относительно максимума.

Для формирования приведенной контурной ДН размер апертуры рефлектора увеличен в 9 раз в сравнении с размером апертуры антенны, реализующей ДН с круговым контуром границы, очерченным вокруг требуемой области. Эффективность данной антенны, определяемая в соответствии с [24], увеличивается по сравнению с круглой синфазно возбужденной апертурой на 3.5 дБ.

К достоинствам описанного алгоритма профилирования рефлектора можно отнести:

- возможность синтеза рефлекторов ЗА, формирующих на апертуре требуемое распределение амплитуд и фаз, которое соответствует заданной ДН;
- отсутствие поляризационной зависимости разрабатываемой структуры;
- синтезированная ЗА имеет улучшенные характеристики устойчивости к влияющим факторам космического пространства в сравнении с многоканальными системами.

Недостатками являются:

- алгоритм не позволяет синтезировать антенну, имеющую различные рабочие области на ортогональных поляризациях;
- рабочая область покрытия антенны является неизменной и не может быть перестроена во время работы.

Выводы. 1. Разработан алгоритм определения профиля рефлектора ЗА с одиночным облучателем, который позволяет формировать на апертуре антенны АФР, соответствующее ДН с требуемыми параметрами. На первом шаге выполнен предварительный расчет АФР на плоскости апертуры по заданной границе рабочей области. Для этого применен метод парциальных диаграмм. На основании данного расчета и с учетом известных характеристик облучателя ЗА на втором шаге выполнен расчет ее геометрических параметров, при которых на апертуре обеспечивается требуемое спадание АР. Также определена величина неравномерности ФР на апертуре рефлектора, вызванная смещением фазового центра облучателя относительно середины его раскрыва, которая скомпенсирована на последующих этапах. На четвертом этапе выполнено уточнение ФР на апертуре ЗА с помощью метода синтеза ФР по заданной амплитудной ДН при условии, что АР является известным. Для этого

- задано начальное приближение АФР, полученное на первом шаге предложенного алгоритма;
- выполнен итерационный процесс синтеза ФР, которое определяет требуемую ДН, с учетом того, что АР задано.

На заключительном этапе получено уравнение поверхности рефлектора ЗА, который реализует требуемое АФР на апертуре. Построение поверхности рефлектора выполнено на основании результатов расчета уточненного ФР. При получении уравнения данного параболоида использовались его оптические свойства.

2. Проведенные численные исследования синтезированной ЗА методами электродинамического моделирования показали возможность формирования ДН со сложной границей контура главного лепестка с использованием ЗА путем профилирования ее рефлектора.

Определены достоинства и недостатки предложенного алгоритма. Применение рассмотренного подхода позволяет выполнить расчет формы рефлектора ЗА, которая реализует ДН с заданными параметрами.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Калошин В.А., Ле Д.Т.* Синтез и анализ бифокальных двухзеркальных систем // Журнал радиоэлектроники. – 2018. – № 9. – DOI: 10.30898/1684-1719.2018.9.13.
2. *Калошин В.А., Фролова Е.В.* Синтез трехзеркальной системы с заданной формой главного зеркала по фазовому фронту и закону отображения // Радиотехника и электроника. – 2019. – Т. 64, № 9. – С. 887-890. – DOI: 10.1134/S0033849419080072.
3. *Калошин В.А., Попов А.П.* Геометрооптический синтез остронаправленных двухзеркальных антенн // Радиотехника и электроника. – 1981. – № 9. – С.1816-1818.

4. *JiXiang Wan, Tao Yan, Feng Wang.* A Hybrid Reflector Antenna for Two Contoured Beams With Different Shapes // IEEE Antennas and wireless propagation letters. – 2018. – Vol. 17, No. 7. – P. 1171-1175. – DOI: 10.1109/LAWP.2018.2836927.
5. *Былов А.А., Саломатов Ю.П.* Диаграммообразующая схема однозеркальной антенны Ku-диапазона на основе планарной фазированной антенной решетки отражательного типа // Доклады ТУСУРа. – 2016. – Т. 19, № 3. – С. 11-15. – DOI:10.21293/1818-0442-2016-19-3-11-15.
6. *Zhenqin Zheng, Long Zhang, Xianting Xie, Qingwei Ji, Hui Chen, Chunxu Mao.* Single-Feed Reflectarray Antenna with Eight Beams for Satellite Communications // IEEE 4th International Conference on Electronic Information and Communication Technology. – 2021. – P. 658-661. – DOI: 10.1109/ICEICT53123.2021.9531087.
7. *Банков С.Е, Калошин В.А., Фролова Е.А.* Оптимизация планарных зеркальных систем по совокупности показателей качества // Радиотехника и электроника. – 2014. – Т. 59, № 11. – С.1090-1101. – DOI: 10.7868/S0033849414100027.
8. *Yukhanov Y. and Privalova T.* Synthesis of a Twist Structure from a Given Multibeam Scattering Diagram // 2023 International Conference on Electromagnetics in Advanced Applications (ICEAA), Venice, Italy, 2023. – P. 135-138. – DOI: 10.1109/ICEAA57318.2023.10297754.
9. *Florencio R., Encinar J.A., Boix R.R., Losada V., and Toso G.* Reflectarray antennas for dual polarization and broadband telecom satellite applications // IEEE Transactions on antennas and propagation. – 2015. – Vol. 63, No. 4. – P. 1234-1246. – DOI: 10.1109/TAP.2015.2391279.
10. *Daniel R. Prado, Manuel Arrebola, Marcos R. Pino, George Goussetis.* Contoured-Beam Dual-Band Dual-Linear Polarized Reflectarray Design Using a Multiobjective Multistage Optimization // IEEE Transactions on antennas and propagation. – 2020. – Vol. 11, No. 68. – P. 7682-7687. – DOI: 10.1109/TAP.2020.2993014.
11. *Воронцов М.А., Шмальгаузен В.И.* Принципы адаптивной оптики. – М.: Наука, 1985. – 336 с.
12. *Узолин Е.Ю., Крюков И.Г.* Синтез профиля поверхности рефлекторов однозеркальных контурных антенн с использованием полиномов Цернике // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2012. – № 3. – С. 92-95.
13. *Мищенко Е.Н., Мищенко С.Е.* Рекуррентный алгоритм вычисления радиальных функций полиномов Цернике // Вестник РГУПС. – 2017. – № 1. – С. 121-127.
14. *Безуглов Д.А., Забродин Р.А.* Методика аппроксимации гибкого адаптивного зеркала ограниченным числом полиномов Цернике // Оптика атмосферы и океана. – 2006. – Т. 19, № 9. – С. 810-814.
15. *Мануилов Б.Д. Башлы П.Н. Безуглов Ю.Д.* Алгоритм управления многофункциональными антенными решетками на основе метода парциальных диаграмм // Антенны. – 2005. – № 9. – С. 72-81.
16. *Зелкин Е.Г., Соколов В.Г.* Методы синтеза антенн. – М.: Советское радио, 1980. – 295 с.
17. *Калошин В.А., Попов А.П.* О синтезе двухзеркальных антенн по амплитудной диаграмме направленности // Радиотехника и электроника. – 1982. – № 6. – С. 1110-1119.
18. *Зелкин Е.Г., Кравченко В.Ф., Гусевский В.И.* Конструктивные методы аппроксимации в теории антенн. – М.: САЙНС-ПРЕСС, 2005. – 512 с.
19. *Rafael A. Penchel, José R. Bergmann, Fernando J. S. Moreira.* Shaping Single Offset Reflector Antennas Using Local Axis-Displaced Confocal Quadrics // Hindawi Publishing Corporation International Journal of Antennas and Propagation. – 2016. – P. 1-9. – DOI: 10.1155/2016/4715681.
20. *Венецкий А.С., Калошин В.А.* Фазовое, лучевое и амплитудное распределение поля в осесимметричной двухзеркальной телескопической системе при смещении облучателя из фокуса // Доклады VI Всеросс. конф. «Радиолокация и радиосвязь». – М., 2012. – Т. 2. – С. 250-254.
21. *Semenov B.V., Chernykh N.I., Pleshchev V.M.* New method of reflector surface shaping to produce a prescribed contour beam // Ural mathematical journal. – 2017. – Vol. 3, No. 2. – P. 143-151. – DOI: 10.15826/umj.2017.2.015.
22. *Занин К.М.* Критерий эффективности применения антенн с контурными диаграммами направленности в составе бортовых комплексов систем спутниковой связи // Системы управления, связи и безопасности. – 2024. – № 2. – С. 1-13. – DOI: 10.24412/2410-9916-2024-2-001-013.
23. *Poulton G.T., Hay S.G.* Shaped reflector synthesis by successive projections // IEEE Antennas and propagation Society International Symposium – 1992 Digest. – P. 363-366. – DOI: 10.1109/APS.1992.221924.
24. *Шишлов А.В.* Зеркальные антенны с контурными диаграммами направленности – эффективность и предельные возможности // Радиотехника. – 2006. – № 4. – С.45-50.

REFERENCES

1. *Kaloshin V.A., Le D.T.* Sintez i analiz bifokal'nykh dvukhzerkal'nykh sistem [Synthesis and analysis of bifocal two-mirror systems], *Zhurnal radio-elektroniki* [Journal of Radio Electronics], 2018, No. 9. DOI: 10.30898/1684 1719.2018.9.13.
2. *Kaloshin V.A., Frolova E.V.* Sintez trekhzerkal'noy sistemy s zadannoy formoy glavnogo zerkala po fazovomu frontu i zakonu otobrazheniya [Synthesis of a three-mirror system with a given shape of the main mirror along the phase front and the law of mapping], *Radiotekhnika i elektronika* [Journal of Communications Technology and Electronics], 2019, Vol. 64, No. 9, pp. 887-890. DOI: 10.1134/S0033849419080072.
3. *Kaloshin V.A., Popov A.P.* Geometroopticheskiy sintez ostronapravlennykh dvukhzerkal'nykh antenn [Geometrooptic synthesis of highly directional two-reflector antennas], *Radiotekhnika i elektronika* [Journal of Communications Technology and Electronics], 1981, No. 9, pp. 1816-1818.
4. *JiXiang Wan, Tao Yan, Feng Wang.* A Hybrid Reflector Antenna for Two Contoured Beams With Different Shapes, *IEEE Antennas and wireless propagation letters,* 2018, Vol. 17, No. 7, pp. 1171-1175. DOI: 10.1109/LAWP.2018.2836927.
5. *Bylov A.A., Salomatov Yu.P.* Diagrammoobrazuyushchaya skhema odnozerkal'noy antenny Kudiapazona na osnove planarnoy fazirovannoy antennoy reshetki otrazhatel'nogo tipa [Diagramming diagram of a single-mirror Ku-band antenna based on a planar phased antenna array of reflective type], *Doklady TUSURa* [Proceedings of TUSUR University], 2016, Vol. 19, No. 3, pp. 11-15. DOI: 10.21293/1818 0442 2016 19 3 11 15.
6. *Zhenqin Zheng, Long Zhang, Xianting Xie, Qingwei Ji, Hui Chen, Chunxu Mao.* Single-Feed Reflectarray Antenna with Eight Beams for Satellite Communications, *IEEE 4th International Conference on Electronic Information and Communication Technology,* 2021, pp. 658-661. DOI: 10.1109/ICEICT53123.2021.9531087.
7. *Bankov S.E, Kaloshin V.A., Frolova E.A.* Optimizatsiya planarnykh zerkal'nykh sistem po sovokupnosti pokazateley kachestva [Optimization of planar reflector systems for a set of quality indices], *Radiotekhnika i elektronika* [Journal of Communications Technology and Electronics], 2014, Vol. 59, No. 11, pp. 1090-1101. DOI: 10.7868/S0033849414100027.
8. *Yukhanov Y. and Privalova T.* Synthesis of a Twist Structure from a Given Multibeam Scattering Diagram, *2023 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, Venice, Italy, 2023, pp. 135-138. DOI: 10.1109/ICEAA57318.2023.10297754.
9. *Florencio R., Encinar J.A., Boix R.R., Losada V., and Toso G.* Reflectarray antennas for dual polarization and broadband telecom satellite applications, *IEEE Transactions on antennas and propagation,* 2015, Vol. 63, No. 4, pp. 1234-1246. DOI: 10.1109/TAP.2015.2391279.
10. *Daniel R. Prado, Manuel Arrebola, Marcos R. Pino, George Goussetis.* Contoured-Beam Dual-Band Dual-Linear Polarized Reflectarray Design Using a Multiobjective Multistage Optimization, *IEEE Transactions on antennas and propagation,* 2020, Vol. 11, No. 68, P. 7682-7687. DOI: 10.1109/TAP.2020.2993014.
11. *Vorontsov M.A., SHmal'gauzen V.I.* Printsipy adaptivnoy optiki [Principles of adaptive optics]. Moscow: Nauka, 1985, 336 p.
12. *Uzolin E.Yu., Kryukov I.G.* Sintez profilya poverkhnosti reflektorov odnozerkal'nykh konturnykh antenn s ispol'zovaniem polinomov TSernike [Synthesis of the surface profile of reflectors of single-mirror contour antennas using Zernike polynomials], Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniya i radioelektroniki [Reports of Tomsk State University of Control Systems and Radioelectronics], 2012, No. 3, pp. 92-95.
13. *Mishchenko E.N., Mishchenko S.E.* Rekurrentnyy algoritm vychisleniya radial'nykh funktsiy polinomov Tsernike [Recurrent Algorithm for Calculation of the Radial Functions of Zernike Polynomials], *Vestnik RGUPS* [Bulletin of the Rostov State University of Railways], 2017, No. 1, pp. 121-127.
14. *Bezuglov D.A., Zabrodin R.A.* Metodika approksimatsii gibkogo adaptivnogo zerkala ogranichennym chislom polinomov Tsernike [Metodika approksimacii gibkogo adaptivnogo zerkala ogranichennym chislom polinomov Cernike], *Optika atmosfery i okeana* [Optics of the atmosphere and ocean], 2006, Vol. 19, No. 9, pp. 810-814.
15. *Manuilov B.D. Bashly P.N. Bezuglov Yu.D.* Algoritm upravleniya mnogofunktsional'nymi antennymi reshetkami na osnove metoda partsial'nykh diagramm [Algorithm for controlling multifunctional antenna arrays based on the method of partial diagrams], *Antenny* [Antennas], 2005, No. 9, pp. 72-81.
16. *Zelkin E.G., Sokolov V.G.* Metody sinteza antenn [Methods of antenna synthesis]. Moscow: Sovetskoe radio, 1980, 295 p.
17. *Kaloshin V.A., Popov A.P.* O sinteze dvukhzerkal'nykh antenn po amplitudnoy diagramme napravlennosti [On the synthesis of two-reflector antennas by amplitude pattern], *Radiotekhnika i elektronika* [*Journal of Communications Technology and Electronics*], 1982, No. 6, pp. 1110-1119.

18. *Zelkin E.G., Kravchenko V.F., Gusevskiy V.I.* Konstruktivnye metody approksimatsii v teorii antenn [Constructive methods of approximation in antenna theory]. Moscow: SAYNS-PRESS, 2005, 512 p.
19. *Rafael A. Penchel, José R. Bergmann, Fernando J. S. Moreira.* Shaping Single Offset Reflector Antennas Using Local Axis-Displaced Confocal Quadrics, *Hindawi Publishing Corporation International Journal of Antennas and Propagation,* 2016, pp. 1-9. DOI: 10.1155/2016/4715681.
20. *Venetskiy A.S., Kaloshin V.A.* Fazovoe, luchevoe i amplitudnoe raspredelenie polya v osesimmetrichnoy dvukhzerkal'noy teleskopicheskoy sisteme pri smeshchenii obluchatelya iz fokusa [Phase, radial and amplitude distribution of the field in an axisymmetric two-reflector telescopic system when the irradiator is displaced from focus], *Reports of the VI All-Russia conference «Radar and radio communication»*. Moscow, 2012, Vol. 2, pp. 250-254.
21. *Semenov B.V., Chernykh N.I., Pleshchev V.M.* New method of reflector surface shaping to produce a prescribed contour beam, *Ural mathematical journal,* 2017, Vol. 3, No. 2, pp. 143-151. DOI: 10.15826/umj.2017.2.015.
22. *Zanin K.M.* Kriteriy effektivnosti primeneniya antenn s konturnymi diagrammami napravlennosti v sostave bortovykh kompleksov sistem sputnikovoy svyazi [The Criterion of the Efficiency of Use of Antennas with Contour Beams in Satellite Communication Systems], *Sistemy upravleniya, svyazi i bezopasnosti* [*Systems of Control, Communication and Security*], 2024, No. 2, pp. 1-13. DOI: 10.24412/2410-9916-2024-2-001-013.
23. *Poulton G.T., Hay S.G.* Shaped reflector synthesis by successive projections, *IEEE Antennas and propagation Society International Symposium*, 1992 Digest, pp. 363-366. DOI: 10.1109/APS.1992.221924.
24. *Shishlov A.V.* Zerkal'nye antenny s konturnymi diagrammami napravlennosti – effektivnost' i predel'nye vozmozhnosti [Reflector antennas with contoured beams: efficiency and limiting capability], *Radiotekhnika* [Radio engineering], 2006, No. 4, pp. 45-50.

Статью рекомендовал к опубликованию д.т.н., профессор К.Е. Румянцев.

Занин Константин Михайлович – Ростовский-на-Дону научно-исследовательский институт радиосвязи; e-mail: konstantinzanin@rambler.ru; г. Ростов-на-Дону, Россия; тел.: +78632018463; аспирант; руководитель группы.

Габриэльян Дмитрий Давидович – e-mail: d.gabrielijan2011@yandex.ru; тел.: +78632508985; д.т.н.; профессор; зам. начальника НТК.

Кузнецов Юрий Викторович – тел.: +78632018463; e-mail: kuznetsov_yu_v@list.ru; руководитель группы.

Мищенко Сергей Евгеньевич – e-mail: mihome@yandex.ru; тел.: +78633067810; д.т.н.; профессор; в.н.с.

Zanin Konstantin Mikhailovich – Rostov-on-Don Research Institute of Radio Communication; e-mail: konstantinzanin@rambler.ru; Rostov-on-Don, Russia; phone: +78632018463; postgraduate student; head of group.

Gabrielyan Dmitry Davidovich – e-mail: d.gabrielijan2011@yandex.ru; phone: +78632508985; dr. of eng. sc.; professor; deputy Head of the Scientific and Technical complex.

Kuznetsov Yury Viktorovich – e-mail: kuznetsov_yu_v@list.ru; phone: +78632018463; head of group.

Mishenko Sergey Evgenievich – e-mail: mihome@yandex.ru; phone: +78633067810; dr. of eng. sc.; professor; lead research scientist.

УДК 621.396.674 DOI 10.18522/2311-3103-2024-4-181-188

И.А. Алшимайсаве, Ю.В. Юханов

ОПТИМИЗАЦИЯ ВСЕНАПРАВЛЕННОЙ АНТЕННЫ 2 × 2 MIMO ДЛЯ ВНУТРЕННИХ ПРИЛОЖЕНИЙ 2G, 3G, 4G И 5G

Из-за сосуществования нескольких типов сетей связи и растущей потребности в высокоскоростной передаче данных многочастотные и широкополосные системы связи приобрели популярность в качестве тем для изучения. Всенаправленные антенны могут обрабатывать больше отдельных диапазонов частот и полезны для различных устройств беспроводной связи благодаря

своей диаграмме направленности, которая облегчает эффективную передачу и прием с мобильного устройства. Однако для систем мобильной связи, поддерживающих приложения 2G, 3G, 4G и будущих 5G, использование антенны с высокой пропускной способностью может иметь решающее значение. Поскольку 5G предлагает своей обширной базе пользователей более высокую скорость передачи данных, большую надежность и снижение энергопотребления, были опубликованы многочисленные исследования широкополосных антенн 5G. Благодаря своим многочисленным преимуществам, таким как более высокая пропускная способность канала, лучшая производительность передачи и приема сигнала, возможность размещать большие антенны в крошечных пространствах и многое другое, MIMO стала важнейшей технологией для 5G. Недавно для мобильных телефонов было предложено несколько различных типов антенн 5G MIMO. В этом исследовании предлагается внутренняя система связи GSM/3G/LTE/5G с использованием широкополосной MIMO-антенны 2 × 2. В антенне используются два антенных элемента, равномерно расположенных вокруг центра, для формирования всенаправленной диаграммы направленности. Одновременно достигаются превосходные свойства всенаправленного излучения и широкая полоса пропускания. Полоса импеданса (0,7–5,3) ГГц может быть достигнута с обратными потерями до -23 по результатам моделирования и усилением до 6,5 дБ. Для моделирования антенны используется ANSYS HFSS (High Frequency Structure Simulator) 2020.

Всенаправленная; MIMO-антенна; сверхширокополосная.

I.A. Alshimaysawe, Yu.V. Yukhanov

OPTIMIZATION OMNI-DIRECTIONAL 2 × 2 MIMO ANTENNA FOR INDOOR 2G, 3G, 4G, AND 5G APPLICATIONS

Due to the cohabitation of multiple types of communication networks and the increasing need for high-speed data transmission, multi-frequency and broadband communication systems have gained popularity as study topics. Omnidirectional antennas can handle more individual frequency bands and are useful for a variety of wireless communications devices due to their radiation pattern, which facilitates effective transmission and reception from a mobile device. However, for mobile communication systems supporting 2G, 3G, 4G, and future 5G applications, the use of a high-bandwidth antenna may be crucial. Since 5G offers its vast user base higher data speed, greater dependability, and reduced power consumption, numerous studies on 5G broadband antennas have been published. Because of its many advantages, such as higher channel capacity, better signal transmission and reception performance, the ability to place big antennas in tiny spaces, and more, MIMO has emerged as a crucial technology for 5G. A number of different 5G MIMO antenna types have recently been suggested for cellphones. An indoor GSM/3G/LTE/5G communication system using a 2 × 2 wideband MIMO antenna is suggested in this study. The antenna uses two antenna elements evenly spaced around the centre to form an omnidirectional radiation pattern. Simultaneously, excellent omnidirectional emission properties and a broad bandwidth are obtained. An impedance bandwidth of (0.7-5.3) GHz can be accomplished with a return loss of up to -23 based on the simulation results, with a gain of up to 6.5 dB. ANSYS HFSS (High Frequency Structure Simulator) 2020 is used to simulate the antenna.

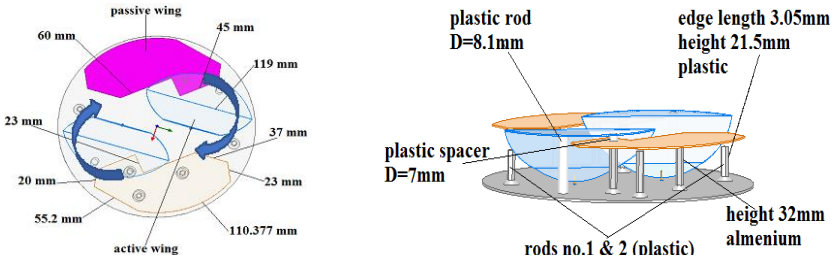
Omni-directional; MIMO antenna; ultra-wideband.

I. Introduction. Mobile base stations and antennas must adapt to support the new 5G frequency bands below 6 GHz, as well as 2G, 3G, and the current 4G (0.7-0.96 GHz and 1.7-2.7 GHz) as telecom service providers look to install 5G mobile systems around the world [1]. Compared to the present 4G system, fifth generation (5G) communication technology can offer a number of benefits, such as faster transfer rates and lower latency [2–4]. Research has demonstrated that in order to obtain greater transmission rates for 5G operations at frequencies lower than 6 GHz, a multiple-input multiple-output (MIMO) antenna system should be implemented [5–13]. A number of 5G MIMO smartphone antennas have recently been proposed [14–17]. In 2016, the European Commission (EC) revealed the bands between 3.4 and 3.8 GHz as part of its spectrum strategy for 5G testing. 2017 saw the official announcement by China's Ministry of Industry and Information Technology (MIIT) that the 3.4-3.6, 4.8-5, and 3.3-3.4 (indoor only) GHz bands would be reserved for 5G services [18]. For indoor base stations where there is limited space for placing antennas, a single antenna element covering the desired frequency bands is preferred over several antennas for different bands [1]. Many attempts have been made to use a single antenna covering the frequency bands of 0.7 to 0.96 GHz and from

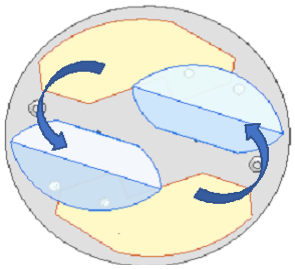
1.7 to 2.7 GHz to suit the needs of 2G, 3G, 4G, and 5G applications [1]. In [19], a wide-band multiple-input multiple-output (MIMO) antenna with dual-band (2.4 and 5 GHz) operation was proposed for premium indoor access points (IAPs), and its performance was compared with that of dipole and patch antennas. In [20], the explanation of a method for incorporating dual-band frequencies into a broad bandwidth single-layer board. In this study, two "L"-shaped and rectangular radiating elements are combined and implanted on a relatively compact single layer structure to construct a dual-band printed dipole antenna. Two IEEE WLAN standards can be completely supported by the printed dipole antenna, which operates in the frequency range of 2400–2500 MHz and 4900–5875 MHz. In [21], developed a brand-new, low-cross-polarization, wideband omni-directional antenna for indoor GSM1800/3G/LTE/5G communication systems. The printed log-periodical antenna elements of the suggested antenna are uniformly spaced around the centre to achieve the omni-directional radiation patterns. Additionally, a wide bandwidth and high omnidirectional radiation performance are realised concurrently between 1.7 and 3.8 GHz with a gain of around 1.5 dBi across the whole operating frequency thanks to the cooperation of the log-periodical antenna and annular parasitic patches. In [22], the proposed antenna is composed of three radiators above a ground plane, a coupling patch above the monopole with three shorting legs to increase the lowest operating frequencies, a monopole composed of three patches fed by a coaxial line and a top-loading disc on top of the coupling patch to further reduce the lowest operating frequencies. An enhanced impedance bandwidth of 9.23:1 was obtained with S11 -13.9 dB (for SWR 1.5) from 650 MHz to 6 GHz. In [23], an electrically compact upper-band Alford loop antenna and a lower-band omnidirectional loop antenna are combined on a single substrate to provide a small, horizontally polarised dual-band omnidirectional antenna, according to the technique presented. To reduce the gain variation in the azimuthal plane, a method for effectively extending the Alford loop's bandwidth was developed, the electrically small loop is fed by four symmetrical radial strips extended from a circular patch, for demonstration, the electrically small loop and Alford loop were created for the 2.4- and 5-GHz Wi-Fi bands. It covers the 2.4–2.5 GHz and 5.1–5.9 GHz Wi-Fi frequencies. In [1], the dual polarisation antenna consists of two orthogonal dipole antennas. Each dipole is composed of three different types of radiators are cat-ear-shaped arms for different bands, bowtie dipoles, and elliptical dipoles. Three broad bands, each with individually controlled fractional bandwidths of 31.3% (0.7-0.96 GHz), 55.3% (1.7-3 GHz), and 14% (3.3-3.8 GHz), are offered by the proposed antenna. In [24], three MED antennas were used in this submission to present a novel 3-D circular conformal MIMO antenna system. The single MED element was meticulously built with one main (lower-band) dipole and two auxiliary (upper-band) dipoles to achieve the dual band radiation. The MED element exhibits an impedance bandwidth of 54.2% (1.68 GHz-2.93 GHz) with a stable gain of 6.05±1.15dBi in the lower band and 9.2% (3.32 GHz-3.64 GHz) with 5.71±0.7dBi in the upper band, respectively. In this paper, an antenna designed in [25] is developed, operating at frequencies (698-960) MHz, (1710-2700) MHz and (2800-3800) MHz. Typical 2×2 MIMO antenna that are omnidirectional radiation pattern is developed by modification in the design Where work is done to change the location of two passive wings, Simulations are conducted for a number of designs, and the best results are obtained in two cases, the first case when change the angles through moving passive wings clockwise by 4 degrees, the results are better at first band (698-1) GHz, second bands (1.7-1.9, 2.4-3.1) GHz, and third band (3.4-3.8) GHz in S11, VSWR, gain, and efficiency, while the second case when change the angles through moving passive wings counter clockwise by 1 degree, the results are the best in the band (3.4-3.8) GHz and slightly lower in the band (698-0.9) GHz in terms of S11, VSWR, gain, and efficiency. The antenna must work in the whole WLAN frequency spectrum since the frequency covers both IEEE 802.11b/g and 802.11a/j, and because it has enough gain and beam coverage to apply to a premium access point. For mobile communication systems supporting 2G, 3G, 4G, and upcoming 5G applications, a broadband antenna can be useful. It is different from the previously listed antennas in that it has a large bandwidth, low return loss, and high gain, making it appropriate for usage in a variety of settings and with communication systems across all generations (2-5G).

II. The proposed of antenna design. The circular 2x2 MIMO antenna structure, as shown in figure 1, is designed for frequencies between 0.7 GHz and 5.3 GHz. The antenna consists of two active and two passive wings made of copper, four plastic rods and six aluminum rods. The antenna's ground is constructed of aluminum with dimensions of 213 mm, diameter, and 1.5 mm. As illustrated in figure referred to above, the antenna's size as simulated in ANSYS HFSS 2020.

In fig. 1(a) we have to change the location of two passive wings in [25] where moving passive wings clockwise by 4 degrees and the results obtained are better at first band (698-1) GHz, second bands (1.7-1.9, 2.4-3.1) GHz, and third band (3.4-3.8) GHz in S11, VSWR, gain, and efficiency, while in fig. 1(b) when moving passive wings in [25] where counter clockwise by 1 degree, the results obtained are the best in the band (3.4-3.8) GHz and slightly lower in the band (698-0.9) GHz in terms of S11, VSWR, gain, and efficiency.



(a) Rotate passive wings 4 degrees clockwise



(b) Rotate passive wings 1 degree counter clockwise

Fig. 1. MIMO 2×2 antenna circular shape

III. Simulation results and discussions. This suggesting antenna is created using ANSYS HFSS Version 2020 software. According to fig. 2–6, the lowest value for S11 is -23 in 1.9 GHz, the best value for VSWR is 1.12 in 1.9 GHz, and the maximum value for the Gain is 6.4 dB in 4 GHz. These results are obtained for the ultra-wide band (0.7–5.3 GHz) in the proposed antenna design shown in Figure 1(a), which illustrates how the antenna designed in [25] is modified by moving its passive wings clockwise by 4 degrees. The original antenna's characteristics are shown by red dashed curves, and the modifying antenna's characteristics are shown by solid green curves. The proposed antenna differs in that it can obtain better specifications at first range (698-1) GHz, S11 is -15, VSWR is 1.4, second range (1.7-1.9, 2.4-3.1) GHz, and third range (3.4-3.8) GHz, S11 is -12, VSWR is 1.6, gain is 6dB making it more efficient and suitable for use in 5G applications.

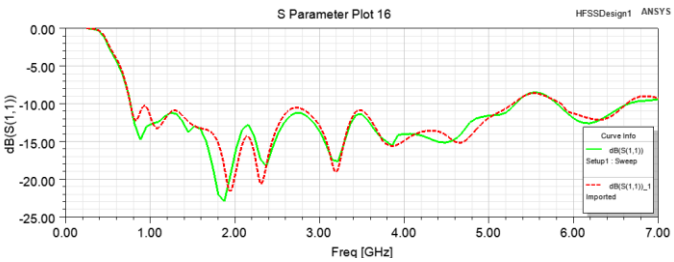


Fig. 2. Return Loss (S11)

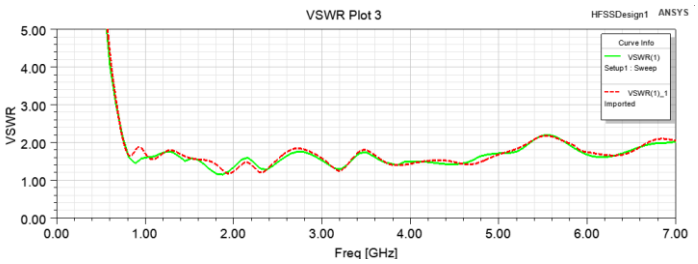


Fig. 3 The VSWR Simulated Results

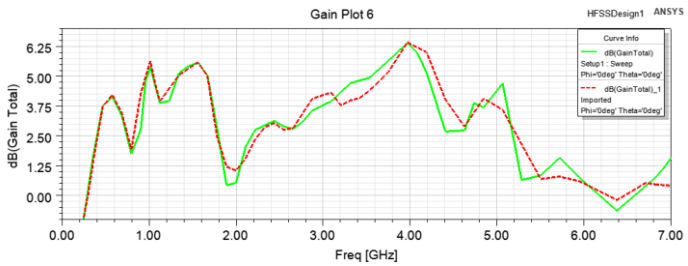


Fig. 4. The Gain vs. Frequency

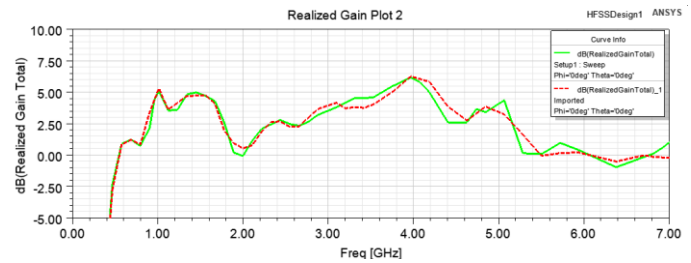


Fig. 5. The Realized Gain vs. Frequency

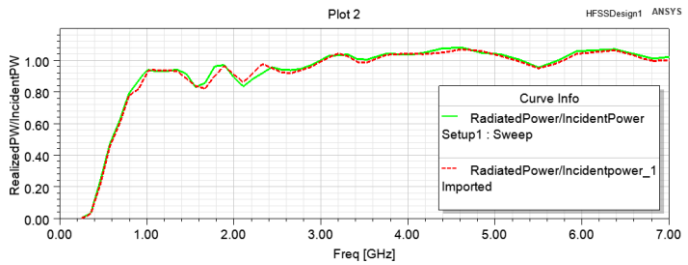


Fig. 6. The Radiated Power / Incident Power vs. Frequency

When modifying the antenna design [25] and studying several different cases, the best results are obtained through analysis by moving passive wings counter clockwise by 1 degree as in fig. 1(b), the results as shown in fig. 7-11 are -20 it's a minimum value for S11 in 1.9 GHz, 1.2 in 1.9 GHz it's the best value for VSWR, and 6.5 dB in 3.7 GHz represent the higher value of the Gain for the band (0.7–5.3 GHz), where (698-0.9) GHz with S11 is -12, VSWR is 1.6, gain is 5.4dB it is considered a little less compared to the antenna result referred to above, but with regard to the band (3.4-3.8) GHz with S11 is -13, VSWR is 1.5, gain is 6.4dB, the difference will be clearly for the better, which makes this design better for use, especially with fifth generation communications applications. Red dashed curves represent the characteristics of the original antenna and solid green curves represent the characteristics of the antenna after modifying the design.

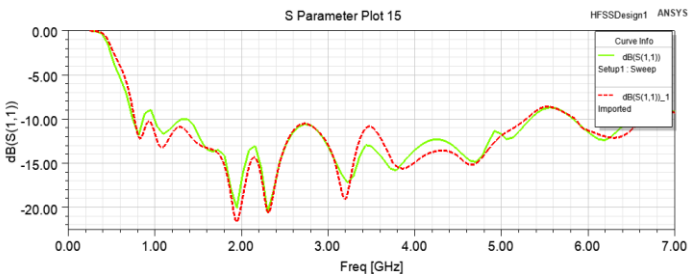


Fig. 7. Return Loss (S11)

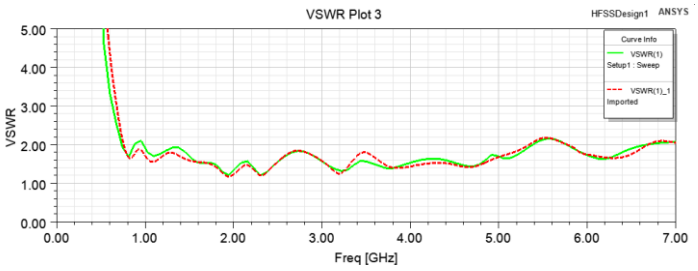


Fig. 8. The VSWR Simulated Results

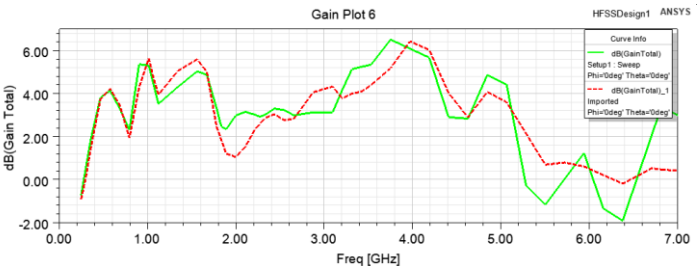


Fig. 9. The Gain vs. Frequency

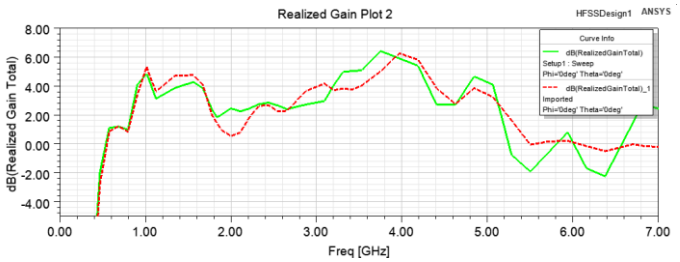


Fig. 10. The Realized Gain vs. Frequency

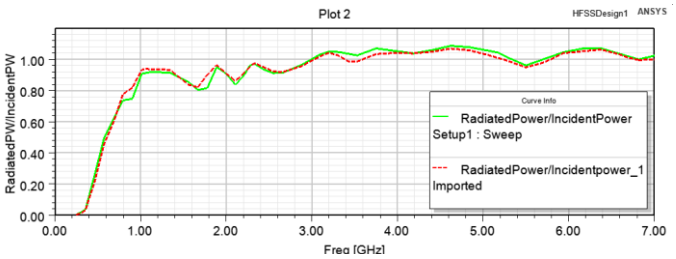


Fig. 11. The Radiated Power / Incident Power vs. Frequency

Conclusions. The rapid advancement of wireless technology and personal communications has led to a greater demand for antennas that can operate across a wide spectrum of frequencies. This makes them more appealing for a variety of applications and compatible with both present and future communication generations. This study describes the design of an omnidirectionally radiation-patterned wide-band MIMO antenna. The circular-shaped 2x2 MIMO antenna generated ultra-wide band frequencies (0.7-5.3GHz) with gain up to 6.5 dB and return losses approaching -23. It can be used for various applications across all mobile communication generations (2G, 3G, 4G, and 5G), all WLAN classes, and upcoming new applications.

REFERENCES

1. *Ahmed Alieldin, et al.,* A Triple-Band Dual-Polarized Indoor Base Station Antenna for 2G, 3G, 4G and Sub-6 GHz 5G Applications, *IEEE Access*, September 2018, Vol. 6, No. 2.
2. *Andrews J.G., et al.* What will 5G be?, *IEEE J. Sel. Areas Commun.*, Jun. 2014, Vol. 32, No. 6, pp. 1065-1082,.
3. Cho Y.S., Kim J., Yang W.Y., and Kang C.G. MIMO-OFDM wireless communications with MATHLAB. Wiley, New York, 2010.
4. *Anping Zhao, and Zhouyou Ren.* 'Size Reduction of Self-Isolated MIMO Antenna System for 5G Mobile Phone Applications. *IEEE Antennas and Wireless Propagation Letters*, November 2018.
5. *Ai-Hadi A.A., Iivonen J., Valkonen R., and Viikari V*. Eight-element antenna array for diversity and MIMO mobile terminal in LTE 3500 band, *Microwave and Optical Technol. Lett*., June 2014, Vol. 56, pp. 1323-1327.
6. *Wong K.- L., Lu J.Y., Chen L.Y., Li W.-Y., and Ban Y.L*. 8-antenna and 16-antenna arrays using the quad-antenna linear array as a building block for the 3.5GHz LTE MIMO operation in the smartphone, *Microwave and Optical Technol. Lett*., Jan. 2016, Vol. 58, pp. 174-181.
7. *Qin Z., Geyi W., Zhang M., and Wang J.* Printed eight-element MIMO system for compact and thin 5G mobile handset, *Electron. Lett.*, March 2016, Vol. 52, pp. 416-418.
8. *Ban Y.L., Li C., Sim C.Y.D., Wu G., and Wong K.-L.* 4G/5G multiple antennas for future multi-mode smartphone applications, *IEEE Access*, July 2016, Vol. 4, pp. 2981-2988.
9. *Li M.Y., Ban Y.L., Xu Z.Q., Wu G., Sim C., Kang K., and Yu Z.F*. Eight-port orthogonally dual-polarized antenna array for 5G smartphone applications, *IEEE Trans. Antennas Propag.*, Jun. 2016, Vol. 64, pp. 3820-3830.
10. *Wong K.-L., Tsai C.-Y., and Lu J.-Y.* Two asymmetrically mirrored gap-coupled loop antennas as a compact building block for eight antenna MIMO array in the future smartphone, *IEEE Trans. Antennas Propag.*, Apr. 2017, Vol. 65, pp. 1765-1778.
11. *Xu H., Zhou H., Gao S., Wang H., and Cheng Y*. Multimode decoupling technique with independent tuning characteristic for mobile terminals, *IEEE Trans. Antennas Propag.*, Dec. 2017, Vol. 65, pp. 6739-6751.
12. *Li Y., Sim C.-Y.-D., Luo Y. and Yang G*. 12-port 5G massive MIMO antenna array in sub-6GHz mobile handset for LTE bands 42/43/46 applications, *IEEE Access*, Feb. 2018, Vol. 6, pp. 344-354.
13. *Li Y., Sim C-Y.-D., Luo Y. and Yang G*. Multiband 10-antenna array for sub-6 GHz MIMO applications in 5-G smartphones, *IEEE Access*, June 2018, Vol. 6, pp. 28041-28253.
14. Ban Y.L., Chen Z.X., Chen Z., Kang K., and Li J.L.W. Decoupled closely spaced heptaband antenna array for wwan/lte smartphone applications, *IEEE Antennas Wirel. Propag. Lett*, 2014, Vol. 13, pp. 31-34.
15. *Li M.Y, Ban Y.L., Xu Z.Q., Wu G., Sim C.Y.D., Kang K., and Yu Z.F*. Eight-port orthogonally dual-polarized antenna array for 5g smartphone applications, *IEEE Trans. Antennas Propag*., Sept 2016, Vol. 64, No. 9, pp. 3820-3830.
16. *Ban Y.L., Li C., Sim C.Y.D., Wu G., and Wong K.L*. 4g/5g multiple antennas for future multi-mode smartphone applications, *IEEE Access*, 2016, Vol. 4, pp. 2981-2988.
17. *Li Y., Sim C.-Y.D., Luo Y., and Yang G*. 12-port 5G massive MIMO antenna array in sub-6GHz mobile handset for LTE bands 42/43/46 applications, *IEEE Access*, 2018, Vol. 6, pp. 344-354.
18. *Wu Q., Liang P., and Chen X.* A broadband ±45◦ dual-polarized multiple-input multiple-output antenna for 5G base stations with extra decoupling elements, *J. Commun. Inf. Netw.*, Mar. 2018, Vol. 3, No. 1, pp. 31-37.
19. *Insu Yeom, Young Bae Jung and Chang Won Jung*. Wide and Dual-Band MIMO Antenna with Omnidirectional and Directional Radiation Patterns for Indoor Access Points, Journal of Electromagnetic Engineering and Science, Jan. 2019, Vol. 19, No. 1, pp. 20-30.

20. *Tze-Meng O. and Geok T.K.* A dual-band omni-directional microstrip antenna, *Progress In Electromagnetics Research*, July. 2010, Vol. 106, pp. 363-376.
21. *Liangying Li, Wei Yan, Botao Feng and Li Deng.* A Wideband Omni-directional Antenna Based on Printed Log-Periodic Element, *IEEE 3rd International Conference on Electronic Information and Communication Technology*, November 2020, pp. 13-15.
22. *Lei Zhou, Yongchang Jiao, Yihong Qi*, *Zibin Weng, and Liang Lu.* Wideband Ceiling-Mount Omnidirectional Antenna for Indoor Distributed Antenna Systems, IEEE Antennas and Wireless Propagation Letters, April 2014, Vol. 13, pp. 836-839.
23. *Peng Fei Hu, Kwok Wa Leung, Yong Mei Pan, and Shao Yong Zheng.* Electrically Small, Planar, Horizontally Polarized Dual-band Omnidirectional Antenna and its Application in a MIMO System, *IEEE Transactions on Antennas and Propagation*, June 2021.
24. *Botao Feng, Kwok L. Chung, Jiexin Lai, and Qingsheng Zeng.* A Conformal Magneto-Electric Dipole Antenna with Wide H-Plane and Band-Notch Radiation Characteristics for MIMO Base-Station, *IEEE Access*, 2019.
25. *Alshimaysawe I.A*. Ultra wideband indoor omni-directional 2×2 MIMO antenna for 2G, 3G, 4G, and 5G applications, *Izvestiya SFedU. Engineering Sciences*, 2023, No. 3, pp. 266-275.

Статью рекомендовал к опубликованию д.т.н., профессор Б.Г. Коноплев.

Алшимайсаве Ихсан Абдлкарем – Южный федеральный университет; e-mail: alshimaysave@sfedu.ru; г. Таганрог, Россия; тел.: +79957566781; кафедра антенн и радиопередающих устройств; аспирант.

Юханов Юрий Владимирович – e-mail: yu_yukhanov@mail.ru; тел.: +78634371733; кафедра антенн и радиопередающих устройств; зав кафедрой; д.т.н.; профессор.

Alshimaysawe Ikhsan Abdlkarem – Southern Federal University; e-mail: alshimaysave@sfedu.ru; Taganrog, Russia; phone: +79957566781; the Department of Antennas and Radio Transmitting Devices; postgraduate student.

Yukhanov Yuriy Vladimirovich – e-mail: yu_yukhanov@mail.ru; phone: +78634371733; the Department of Antennas and Radio Transmitting Devices; head of Department; dr. of eng. sc.

УДК 621.396.67 DOI 10.18522/2311-3103-2024-4-188-196

Во Ба Ау, И.Н. Бобков, Ю.В. Юханов

НИЗКОПРОФИЛЬНАЯ АНТЕННАЯ РЕШЕТКА СИЛЬНО СВЯЗАННЫХ ДИПОЛЕЙ КРУГОВОЙ ПОЛЯРИЗАЦИИ

Рассмотрена конструкция низкопрофильной антенной решетки сильно связанных диполей круговой поляризации. Основной деталью конструкции являются два скрещенных диполя в печатном исполнении. Квадратурное возбуждение обеспечивается полосками в форме дуги окружности, соединяющими пары ортогонально расположенных плеч на верхнем и нижнем слое металлизации. Для обеспечения емкостной связи между элементами применяются металлические диски, гальванически соединенные с основанием при помощи металлических стержней. Для расширения полосы рабочих частот и улучшения характеристик излучения антенной решетки непосредственно над диполями расположен согласующий слой пластика Eccostock HiK. Представлены результаты численного исследования характеристик элементарной ячейки антенной решетки с периодическими граничными условиями на гранях в программном обеспечении ANSYS HFSS. Показана возможность работы в широкой полосе частот по заданному уровню согласования и коэффициента эллиптичности. Показана зависимость характеристик согласования и коэффициента эллиптичности от размеров полоска, обеспечивающего квадратурное питание плеч диполей. Расчетным путем установлено, что выбор радиуса полоска, обеспечивающего квадратурное возбуждение плеч диполей, представляет собой компромисс между широкой полосой рабочих частот и лучшим коэффициентом эллиптичности в центре диапазона. Показано, что использование расположенного непосредственно над слоем диполей согласующего слоя в решетках сильно связанных диполей круговой поляризации обеспечило согласование в широкой полосе частот при сохранении электри-

чески малой высоты профиля. На основе предложенного элемента разработаны модели конечных антенных решеток из 3×3, 4×4, 5×5 и 6×6 элементов. Показано влияние элементов, расположенных на краях, на характеристики антенной решетки. Исследована возможность улучшения характеристик за счёт подсоединения крайних элементов к согласованным нагрузкам.

Антенная решетка; дипольные антенны; поляризация; решетка сильносвязанных диполей.

Vo Ba Au, I.N. Bobkov, Y.V. Yukhanov

LOW-PROFILE CIRCULARLY POLARIZED TIGHTLY COUPLED DIPOLE ARRAY

The design of a low-profile antenna array of tightly coupled circularly polarized dipoles is considered. The main design detail is two crossed printed dipoles. Quadrature excitation is provided by arc-shaped strips connecting pairs of orthogonally located arms on the upper and lower metallization layers. To ensure capacitive coupling between the elements, metal disks are used, galvanically connected to the base using metal rods. To expand the operating frequency band and improve the radiation characteristics of the antenna array, a matching layer of Eccostock HiK plastic is located directly above the dipoles. The results of a numerical study of the characteristics of an elementary cell of an antenna array with periodic boundary conditions on the faces in the ANSYS HFSS software are presented. The possibility of operating in a wide frequency band at a given level of matching and ellipticity coefficient is demonstrated. The dependence of the matching characteristics and the ellipticity coefficient on the size of the strip that provides quadrature power to the dipole arms is shown. It was established by calculation that the choice of the strip radius, which ensures quadrature excitation of the dipole arms, is a compromise between a wide operating frequency band and a better ellipticity coefficient in the center of the range. It is shown that the use of a matching layer located directly above the dipole layer in arrays of tightly coupled circularly polarized dipoles ensured matching over a wide frequency band while maintaining an electrically low profile height. Based on the proposed element, models of finite antenna arrays of 3×3, 4×4, 5×5 and 6×6 elements have been developed. The influence of elements located at the edges on the characteristics of the antenna array is shown. The possibility of improving performance by connecting the outermost elements to matched loads was investigated.

Antenna arrays; dipole antennas; polarization; tightly coupled dipole array.

Введение. В настоящее время беспроводные системы находят широкое применение в различных сферах деятельности человека. К таким системам относятся системы навигации, системы широкополосного вещания, системы мобильной связи. Учитывая тенденцию к расширению полосы рабочих частот и уменьшению габаритов устройств связи, актуальным становится вопрос создания электрически малых, но при этом широкополосных антенных решеток (АР) [1].

Наиболее современным подходом к проектированию широкополосных АР является подход, предложенный Мунком [2]. Этот подход, который в отличие от традиционного подхода, в котором пытаются минимизировать связь между отдельно спроектированными элементами АР, предусматривает введение сильной емкостной связи между соседними элементами. Следствием применения такого подхода является малая высота профиля АР и широкая полоса рабочих частот, в отдельных случаях достигающая перекрытия 46:1 [3–16].

На основе подхода Мунка были разработаны различные АР сильно связанных дипольных излучателей [4–9]. В [8] описана АР, полоса рабочих частот которой достигает 2.8:1 (1.74–4.9 ГГц). При этом для питания АР применяется симметрирующий трансформатор.

Для обеспечения работы некоторых систем связи предпочтительными являются широкополосные антенные решетки круговой поляризации, поскольку круговая поляризация помогает снизить влияние негативных эффектов многолучевого распространения радиоволн и повысить эффективность передачи [9]. В [10] описана АР сильно связанных дипольных излучателей для базовых станций мобильной связи, работающая в диапазонах частот 700 МГц, 1.7 ГГц и 2.6 ГГц, при этом высота профиля элементов АР является электрически большой, что ограничивает её применение.

В [11] описана АР круговой поляризации с полосой пропускания по уровню КСВН < 3 от 2.06 до 6.46 ГГц (3.14:1) и полосой рабочих частот по уровню коэффициента эллиптичности < 3 дБ от 2.35 до 5.6 ГГц (2.38:1). Однако в описанной АР нет металлического основания-рефлектора. Для подавления заднего излучения был применен радиопоглощающий материал, что привело к потере не менее чем половины подводимой ко входу АР мощности.

В настоящей статье предложена конструкция АР с металлическим основанием-рефлектором. Для обеспечения емкостной связи между элементами АР применяются металлические диски, гальванически соединенные с основанием АР.

Конструкция элемента антенной решетки. На рис. 1 показан общий вид элемента предлагаемой АР. Основной деталью конструкции являются два скрещенных диполя в печатном исполнении. Плечи диполей размещаются на обеих сторонах диэлектрической подложки толщиной 1 мм из материала ФЛАН-2.8 (ε_r = 2.8, tan(δ) = 0.0015).

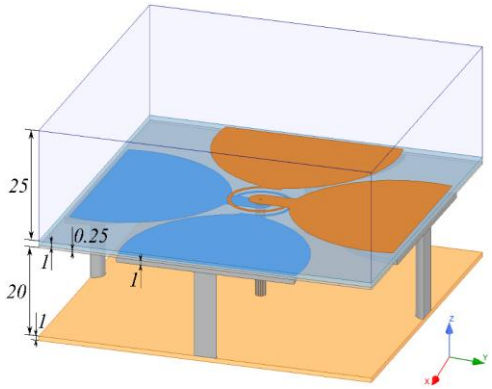


Рис. 1. Общий вид элемента антенной решетки. Диэлектрические материалы показаны полупрозрачными для ясности

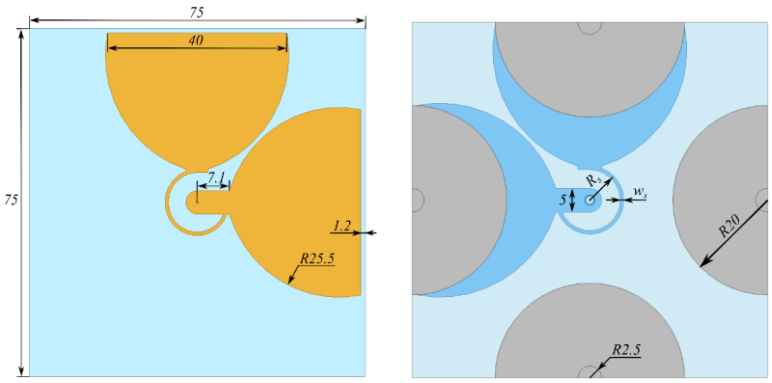
Питание ячейки осуществляется одним коаксиальным кабелем, подводимым снизу сквозь отверстие в металлическом основании. Расстояние от металлического основания (рефлектора) до слоя диполей составляет примерно $\lambda/10$ на центральной частоте рабочего диапазона длин волн и уточняется на этапе проектирования в электромагнитных САПР. Жила коаксиального кабеля подпаивается к отводу одного из плеч диполей на верхнем слое металлизации диэлектрической подложки. Экран кабеля аналогичным образом подпаивается к отводу на нижнем слое металлизации.

Механизм формирования круговой поляризации в предлагаемой АР был сформулирован Брауном [17] ещё в 1935 году. Его суть заключается в возбуждении двух диполей с общим центром, расположенных ортогонально друг к другу, равными амплитудами, но со сдвигом фаз 90°. Антенны такого типа Браун назвал турникетными антеннами [18].

Квадратурное возбуждение в предлагаемой АР обеспечивается полосками в форме дуги окружности, соединяющими пары ортогонально расположенных плеч на верхнем и нижнем слое металлизации [9, 14] (рис. 2). Электрическая длина дуги окружности выбирается равной $\lambda/4$ на центральной частоте рабочего диапазона длин волн и уточняется в процессе настройки АР в САПР.

Для обеспечения емкостной связи между элементами АР в конструкцию добавлены металлические диски, гальванически соединенные с основанием при помощи металлических стержней [19]. Для изоляции нижнего слоя диполей от емкостных дисков между ними установлен дополнительный слой диэлектрика ФЛАН-2.8 толщиной 0.25 мм.

Расширить полосу рабочих частот и улучшить характеристики излучения антенной решетки удалось путем размещения непосредственно над диполями слоя (суперстрата) пластика Eccostock HiK (ε_r = 3.5, tan(δ) = 0.002) [19], электрическая толщина которого составила $\lambda/4$ на центральной частоте рабочего диапазона длин волн [20].



а б

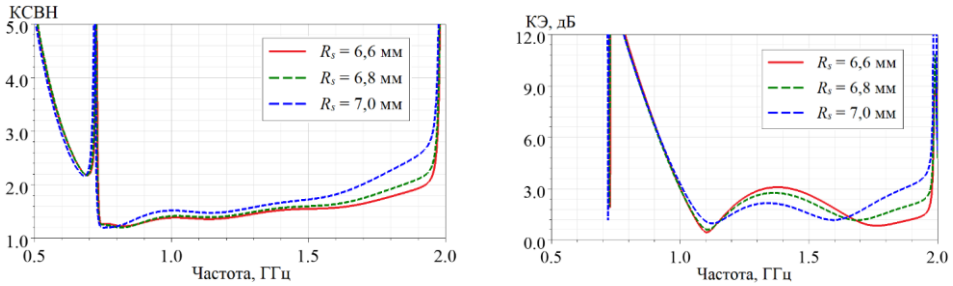
Рис. 2. Элемент антенной решетки. Суперстрат и рефлектор-основание не показаны: а – вид сверху, б – вид снизу

Особенности расчёта характеристик АР. Проектирование и разработка предлагаемой АР были выполнены при помощи программного обеспечения ANSYS HFSS. Выполнялся расчёт характеристик одной ячейки АР с периодическими граничными условиями на гранях. Таким образом, приводятся характеристики одной ячейки в составе АР бесконечных размеров, в которой не учитываются эффекты конкатенации АР. На основе предлагаемой модели элемента АР были построена конструкция конечной АР и было приведено сравнение характеристик АР различных размеров.

Результаты численного исследования одной ячейки. На рис. 3,а,б показаны зависимости КСВН и коэффициента эллиптичности от частоты при различных значениях радиуса полоска в форме дуги окружности предлагаемой низкопрофильной АР. Из анализа результатов можно сделать вывод, что меньший радиус полоска, обеспечивающего квадратурное возбуждение плеч диполей, обеспечивает лучшее согласование и лучший коэффициент эллиптичности на верхних частотах диапазона рабочих длин волн. Но в то же время уменьшение радиуса полоска приводит к увеличению коэффициента эллиптичности в области центра диапазона рабочих частот.

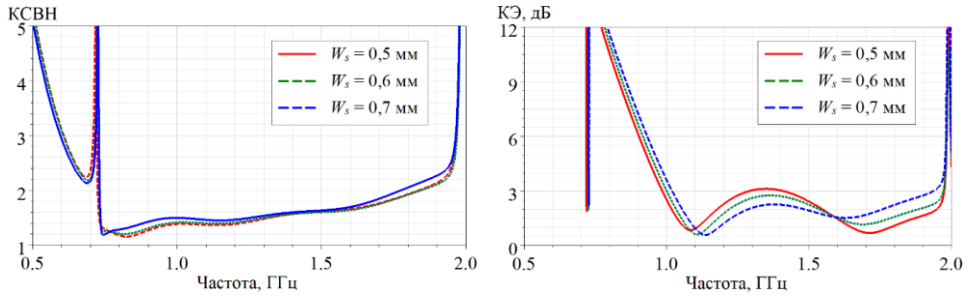
Таким образом, выбор радиуса полоска, обеспечивающего квадратурное возбуждение плеч диполей, представляет собой компромисс между широкой полосой рабочих частот и лучшим коэффициентом эллиптичности в центре диапазона.

На рис. 4,а,б показаны зависимости КСВН и коэффициента эллиптичности от частоты для различных значений ширины полоска, обеспечивающего квадратурное возбуждение плеч диполей. Изменение КСВН является слабо выраженным, но путём уменьшения ширины полоска можно добиться расширения полосы рабочих частот по критерию КЭ < 3 дБ в области верхних частот диапазона рабочих длин волн.



а б

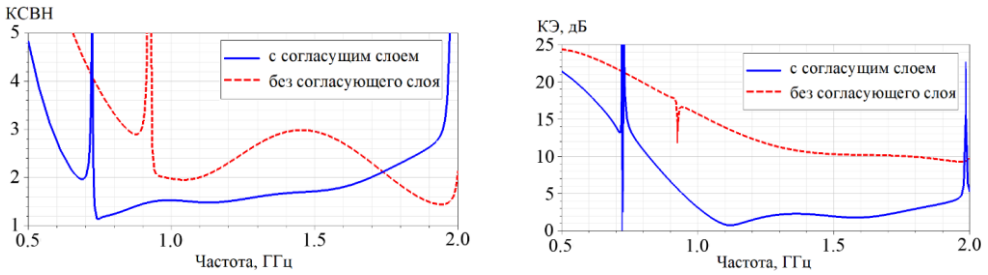
Рис. 3. Влияние радиуса полоска в форме дуги окружности на характеристики антенной решетки: а – зависимость КСВН элемента АР от частоты, б – зависимость коэффициента эллиптичности от частоты



а б

Рис. 4. Влияние ширины полоска в форме дуги окружности на характеристики антенной решетки: а – зависимость КСВН элемента АР от частоты, б – зависимость коэффициента эллиптичности от частоты

На рис. 5,а,б показаны зависимости КСВН и коэффициента эллиптичности от частоты для модели элемента АР со слоем суперстрата и без. Применение суперстрата в антенных решетках сильно связанных диполей круговой поляризации является необходимым условием для получения широкой полосы рабочих частот при сохранении малой высоты профиля АР, также как и в АР, работающих на линейной поляризации [20].



а б

Рис. 5. Влияние наличия суперстрата на характеристики антенной решетки: а – зависимость КСВН элемента АР от частоты; б – зависимость коэффициента эллиптичности от частоты

На рис. 6 показана зависимость эффективности излучения от частоты для варианта исполнения АР с R = 6.8 мм и w = 0.6 мм. Эффективность излучения более 70% получена в диапазоне частот от 0.7 до 2.0 ГГц (2.85:1).

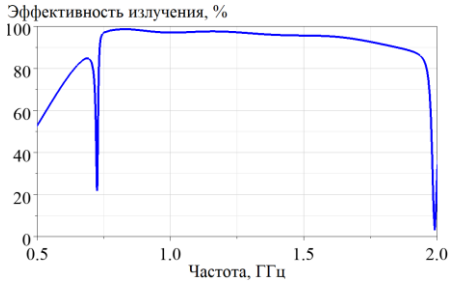


Рис. 6. Эффективность излучения предлагаемой антенной решетки

На рис. 7 показан реализованный коэффициент усиления элемента предлагаемой антенной решетки в сравнении с теоретически достижимым коэффициентом усиления апертуры того же размера. Видно, что мощность, подводимая ко входу элемента АР излучается в свободное пространство и не переходит в потери.

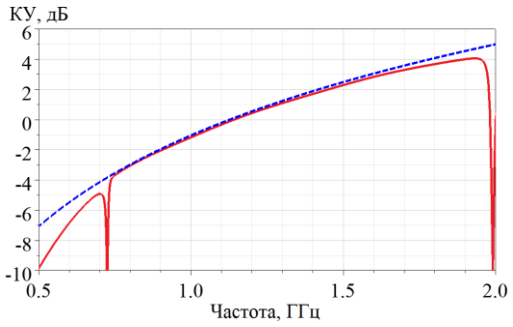


Рис. 7. Реализованный коэффициент усиления элемента АР

Расчет конечной антенной решетки. На основе предлагаемой конструкции элемента АР были разработаны электродинамические модели АР конечных размеров из 3×3, 4×4, 5×5 и 6×6 элементов. Для наглядного примера на рис. 8 представлена АР размерностью 3×3.

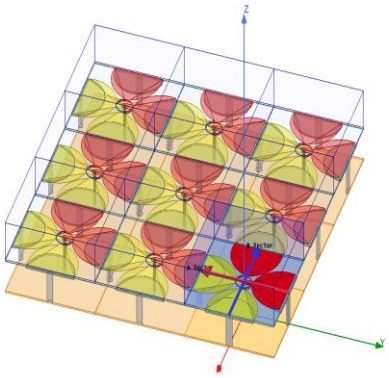


Рис. 8. Конфигурация конечной АР 3×3

Сравнение коэффициента эллиптичности АР с различным количеством элементов показано на рис. 9. Коэффициент эллиптичности АР конечных размеров отличается от коэффициента эллиптичности элементарной ячейки в худшую сторону. При этом в центре диапазона КЭ > 3 дБ, и поляризация становится эллиптической. Причина проблемы заключается во влиянии крайних элементов на характеристики конечных АР.

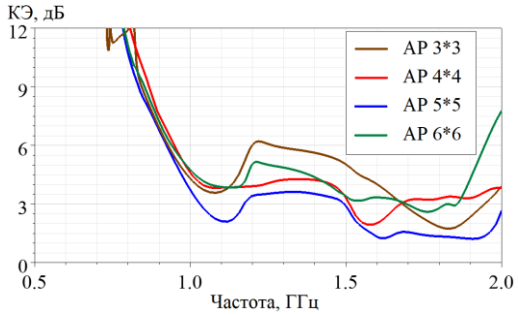


Рис. 9. Коэффициент эллиптичности конечных АР различных размерностей

Для проверки того, как влияют крайние элементы на характеристики конечных АР, были исследованы 2 конфигурации размером 5×5 и 6×6, при этом возбуждались только центральные ячейки (3×3 и 4×4 соответственно), все крайние элементы были подключены к согласованным нагрузкам.

Рассчитанные значения коэффициента эллиптичности показаны на рис. 10. В сравнении с рис. 9 в центре диапазона отмечается значительное улучшение поляризационных свойств АР, что подтверждает сильное влияние крайних элементов на характеристики конечной АР.

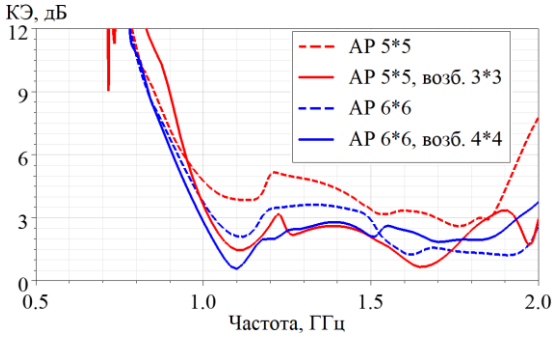


Рис. 10. Коэффициент эллиптичности АР конечных размеров

На рис. 11 показано, что эффективность излучения более 70% получена в диапазоне рабочих частот от 0,73 до 1.95 ГГц (2.67:1).

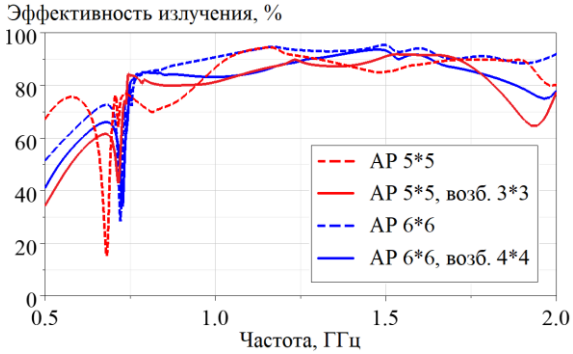


Рис. 11. Эффективность излучения АР конечных размеров

Заключение. В работе предложена конструкция низкопрофильной антенной решетки сильно связанных диполей круговой поляризации. Применены гальванически соединенные с основанием металлические диски для обеспечения емкостной связи между элементами антенной решетки. Показана зависимость характеристик согласования и коэффициента эллиптичности АР от размеров полоска, обеспечивающего квадратурное питание плеч диполей. Показано, что использование расположенного непосредственно над слоем диполей суперстрата в решетках сильно связанных диполей круговой поляризации обеспечило согласование в широкой полосе частот при сохранении электрически малой высоты профиля. Показано влияние элементов, расположенных на краях, на характеристики антенной решетки. Исследована возможность улучшения характеристик за счёт подсоединения крайних элементов к согласованным нагрузкам.

Благодаря широкой полосе рабочих частот, малой высоте профиля, простой схеме питания из одного коаксиального кабеля, не требующей применения сложных и громоздких симметрирующих трансформаторов, предложенная конструкция антенной решетки может найти применение в различных системах связи.

Источник финансирования. Данное исследование выполнено при финансовой поддержке Российского научного фонда (проект №22-19-00537, https://rscf.ru/project/22-19-00537/), Южный федеральный университет, Таганрог и был выполнен в Совместном научном центре ”Прикладная электромагнетика и антенные измерения".

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Holland S.S., Schaubert D.H. and Vouvakis M.N.* A 7–21 GHz Dual-Polarized Planar Ultrawideband Modular Antenna (PUMA) Array, *IEEE Transactions on Antennas and Propagation*, Oct. 2012, Vol. 60, No. 10, pp. 4589-4600.
2. *Munk B.* A low-profile broadband phased array antenna, *IEEE Antennas and Propagation Society Int. Symp*., June 2003, Vol. 2, pp. 448-451.
3. *Johnson A.D., Zhong J., Venkatakrishnan S.B., Alwan E.A. and Volakis J.L.* Phased Array With Low-Angle Scanning and 46:1 Bandwidth, *IEEE Transactions on Antennas and Propagation*, Dec. 2020, Vol. 68, No. 12, pp. 7833-7841.
4. *Novak M. and Volakis J.L.* Dual polarized Tightly Coupled Dipole Array (TCDA) for UHF to millimeter wave applications, *2014 IEEE Antennas and Propagation Society International Symposium (APSURSI)*, 2014, pp. 819-819.
5. *Papantonis D.K. and Volakis J.L.* Dual-Polarized Tightly Coupled Array With Substrate Loading, *IEEE Antennas and Wireless Propagation Letters*, 2016, Vol. 15, pp. 325-328.
6. *Johnson A.D., Zhong J., Livadaru M., Alwan E.A. and Volakis J.L.* Tightly Coupled Dipole Array with Wideband Differential Feeding Network, *2018 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, 2018, pp. 1987-1988.
7. *Islam M.R., Nichols M.W., Johnson A.D., Venkatakrishnan S.B., Alwan E.A. and Volakis J.L.* Wideband Tightly Coupled Dipole Array (TCDA) With 5.4:1 Bandwidth on a Miura-Ori Pattern, *2020 IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting*, 2020, pp. 1685-1686.
8. *Ling X., Wang Y., Zhao L., Wang S. and Chen S.* A new UWB tightly coupled antenna array, *2013 International Conference on Anti-Counterfeiting, Security and Identification (ASID)*, 2013, pp. 1-4.
9. *He Y., He W. and Wong H.* A Wideband Circularly Polarized Cross-Dipole Antenna, *IEEE Antennas and Wireless Propagation Letters*, 2014, Vol. 13, pp. 67-70.
10. *Novak M., Abdelrahman A.H., Yi H., Piazzi L. and Ma Z.* Tightly Coupled Dipole Array for 5G Multiband Cellular Base Station Antennas, *2020 IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting*, 2020, pp. 1681-1682.
11. *Zhang L., Gao S., Luo Q., Li W., He Y. and Li Q.* A Wideband Circularly Polarized Tightly Coupled Array, *IEEE Transactions on Antennas and Propagation*, Nov. 2018, Vol. 66, No. 11, pp. 6382-6387.
12. *Yang W.J., Pan Y.M. and Zheng S.Y.* A Low-Profile Wideband Circularly Polarized Crossed-Dipole Antenna With Wide Axial-Ratio and Gain Beamwidths, *IEEE Transactions on Antennas and Propagation*, July 2018, Vol. 66, No. 7, pp. 3346-3353.
13. *Yang W., Pan Y., Zheng S. and Hu P.* A Low-Profile Wideband Circularly Polarized Crossed-Dipole Antenna, *IEEE Antennas and Wireless Propagation Letters*, 2017, Vol. 16, pp. 2126-2129.
14. *Ta S.X., Park I. and Ziolkowski R.W.* Crossed Dipole Antennas: A review, *IEEE Antennas and Propagation Magazine*, Oct. 2015, Vol. 57, no. 5, pp. 107-122.
15. *Tran H.H., Ta S.X., and Park I.* Single-Feed, Wideband, Circularly Polarized, Crossed Bowtie Dipole Antenna for Global Navigation Satellite Systems, *Journal of electromagnetic engineering and science*, Vol. 14, No. 3. Korean Institute of Electromagnetic Engineering and Science, Sep 2014, pp. 299-305.
16. *Yang W.J., Pan Y.M. and Zheng S.Y.* A Compact Broadband Circularly Polarized Crossed-Dipole Antenna With a Very Low Profile, *IEEE Antennas and Wireless Propagation Letters*, Oct. 2019, Vol. 18, No. 10, pp. 2130-2134.
17. *Brown G.* The turnstile, *Electron*, Apr. 1936, Vol. 9, pp. 14-17.
18. *Brown G.* Antenna system, U.S. Patent 2086976, Jul., 13, 1937.
19. *Merola C.S. and Vouvakis M.N.* UHF planar ultra-wideband modular antenna (PUMA) arrays, 2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting, 2017, pp. 1803-1804.
20. *Holland S.S. and Vouvakis M.N.* The Planar Ultrawideband Modular Antenna (PUMA) Array, *IEEE Transactions on Antennas and Propagation*, Jan. 2012, Vol. 60, No. 1, pp. 130-140.

Статью рекомендовал к опубликованию д.т.н., профессор И.И. Турулин.

Во Ба Ау – Южный федеральный университет; e-mail: bvo@sfedu.ru; г. Таганрог, Россия; тел.: 88634371634; кафедра антенн и радиопередающих устройств; аспирант.

Бобков Иван Николаевич – e-mail: ibobkov@sfedu.ru; кафедра антенн и радиопередающих устройств; аспирант.

Юханов Юрий Владимирович – e-mail: yu_yukhanov@mail.ru; кафедра антенн и радиопередающих устройств; зав. кафедрой; д.т.н.; профессор.

Vo Ba Au – Southern Federal University; e-mail: bvo@sfedu.ru; Taganrog, Russia; phone: +78634371634; the Department Antenna and Radio Transmitter; postgraduate.

Bobkov Ivan Nikolaevich – e-mail: ibobkov@sfedu.ru; the Department Antenna and Radio Transmitter; postgraduate.

Yukhanov Yury Vladimirovich – e-mail: yu_yukhanov@mail.ru; the Department Antenna and Radio Transmitter; head of the department; dr. of eng. sc.; professor.

УДК 621.376.4 DOI 10.18522/2311-3103-2024-4-196-203

В.В. Сердюков, К.С. Коротков

К ВОПРОСУ ОПРЕДЕЛЕНИЯ ФАЗОВЫХ СДВИГОВ В СМЕСИТЕЛЕ

Целью исследования является оценка проблемы влияния на результаты измерений абсолютных фазовых сдвигов, возникающих в смесителе, нелинейности фазовых сдвигов гармоник при умножении частоты и погрешностей в различных методах измерений этих сдвигов в смесителе при гетеродинном преобразовании частоты входного высокочастотного- (ВЧ) или сверхвысокочастотного- (СВЧ) сигнала. Ввиду того, что сигнал на входе ВЧ- или СВЧ- смесителя и сигнал промежуточной частоты на его выходе лежат в разных диапазонах частот, невозможно традиционными способами измерить сдвиг фаз, вносимый нелинейным элементом смесителя в сигнал промежуточной частоты, в процессе гетеродинного преобразования частоты входного ВЧ- или СВЧ- сигнала. Проблемой, рассмотренной в данном исследовании, является выявление погрешности измерений абсолютных фазовых сдвигов, возникающих в смесительном диоде при гетеродинном преобразовании частоты из-за его нелинейности. Эта погрешность может оказать значительное влияние на точность измерений, и поэтому ее учет и компенсация являются важными задачами в радиотехнике и связи. В данной статье демонстрируется важная разница фазовых сдвигов гармоник, умноженных на фазовый сдвиг другого сигнала, используемых в методах измерений абсолютных фазовых сдвигов смесителей. Нами также был предложен метод, лишенный указанных погрешностей измерений, который позволяет учесть нелинейность смесительного диода и обеспечить снижение погрешности измерения. Результаты данного исследования имеют большое значение для повышения точности измерений в радиотехнике и связи. Предложенный метод, лишенный указанных погрешностей, может существенно повысить точность измерений абсолютных фазовых сдвигов смесителей при гетеродинном преобразовании частоты. Это инновационное решение позволяет учесть нелинейность смесительного диода и обеспечить увеличение точности измерения, что может быть весьма полезно при создании приборов, способных измерять сдвиг фаз испытуемого смесителя СВЧ и векторных вольтметров на базе программируемых логических интегральных схем (ПЛИС).

Смеситель сигналов; сдвиг фаз; гетеродинное преобразование частоты; определение фазы.

V.V. Serdukov, K.S. Korotkov

ON THE ISSUE OF DETERMINING PHASE SHIFTS IN THE MIXER

The aim of the study is to solve the problem of the influence of the nonlinearity of phase shifts of harmonics during frequency multiplication on the measurement results of absolute phase shifts occurring in mixers and errors in various measurement methods of these shifts in the mixer during heterodyne frequency conversion of the input ultrahigh frequency (microwave) signal. Since the signal at the input of the microwave mixer and the intermediate frequency signal at its output lie in different frequency ranges, it is impossible by traditional methods to measure the phase shift introduced by the nonlinear element of the mixer into the intermediate frequency signal during the heterodyne frequency conversion of the input microwave signal. The problem that we have considered in this study is to identify the measurement error of absolute phase shifts that occur in a mixing diode during heterodyne frequency conversion due to its nonlinearity. This error can have a significant impact on the accuracy of measurements, and therefore its accounting and compensation are important tasks in radio engineering and communications. This scientific article demonstrates the important inequality of the phase shifts of harmonics multiplied by the phase shift of the multiplied signal used in the measurement methods of absolute phase shifts of mixers. We also proposed an innovative method devoid of these measurement errors, which allows us to take into account

the nonlinearity of the mixing diode and provide more accurate measurements. The results of this study are of great importance for accurate measurements in radio engineering and communications. The proposed method, devoid of these errors, can significantly increase the accuracy of measurements of absolute phase shifts of mixers with heterodyne frequency conversion. This innovative solution allows you to take into account the nonlinearity of the mixing diode and provide accurate measurements, which can be very useful when creating devices capable of measuring the phase shift of the microwave mixer under test and vector voltmeters based on programmable logic integrated circuits (FPGAs).

Signal mixer; phase shift; heterodyne frequency conversion; phase detection.

Введение. В последнее время, для передачи информации широкое распространение получил способ фазовой модуляции радиосигналов, как наиболее крипто- и помехозащищенный. При этом подавляющее число устройств для приема таких радиосигналов использует принцип гетеродинного преобразования частоты с применением смесителей радиосигналов, построенных на основе диодов, которые вносят дополнительный сдвиг фаз и в принципе являющийся паразитным, искажающим передаваемую информацию. Это связано с тем, что каждый диод имеет эквивалентную индуктивность, объем полупроводника L_s, по которому протекает сигнал, и барьерную емкость $C_б$ p-n перехода, которые, как всякая реактивность, вносят в протекающий через них сигнал время задержки, что эквивалентно дополнительному сдвигу фаз, особенно в диапазоне СВЧ.

Физика процесса сдвига фаз в смесителе. Теоретический расчет сдвига фаз, происходящий в смесителе, весьма сложен и не всегда возможен из-за отсутствия информации о молекулярной структуре материала полупроводника. Поэтому широкое распространение получили инструментальные методы измерений сдвигов фаз, вносимых смесительным диодом в сигнал промежуточной частоты (ПЧ) при гетеродинном преобразовании частоты, например, патент РФ № 2524049 [1]. Однако, из-за разных частот сигналов на входе и выходе смесителя, измерить вносимый ими сдвиг фаз традиционными способами не представляется возможным. Поэтому были найдены и разработаны инструментальные способы, специально предназначенные для измерения сдвига фаз, возникающего в смесителе при гетеродинном преобразовании частоты, физика которого состоит в следующем: общий фазовый сдвиг, вносимый полупроводниковым смесительным диодом в сигнал ПЧ — $\varphi_{общ.}$ В процессе гетеродинного преобразования входного сигнала с частотой ω_1, как известно [2], образуется из фазовых сдвигов: входного сигнала φ_1 и фазового сдвига гетеродина φ_2 с частотой ω_2 в виде их суммы $(\varphi_1 + \varphi_2)$ или разности $(\varphi_1 - \varphi_2)$ в зависимости от типа преобразования «вверх» или «вниз» по частоте. Одновременно с этим в сигнал ПЧ, с частотой $\omega_{пч}$ также протекающий через смесительный диод, вносится дополнительный фазовый сдвиг – $\Delta\varphi_{пч}$, определяемый его реактивностями индуктивностью L_s и барьерной емкостью $C_б$, в результате чего общий фазовый сдвиг сигнала ПЧ – $\varphi_{пч\ общ.}$ – складывается из фазовых сдвигов входного сигнала φ_1, гетеродина φ_2 и дополнительного фазового сдвига – $\Delta\varphi_{пч}$, и определяется выражением (1):

$$\varphi_{пчобщ} = (\varphi_1 \pm \varphi_2) + \Delta\varphi_{пч}. \quad (1)$$

Для выяснения причин и последствий возникновения дополнительного фазового сдвига – $\Delta\varphi_{пч}$, проанализируем электрическую схему однодиодного смесителя с эквивалентной схемой полупроводникового смесительного диода приведенную на рис. 1 [3].

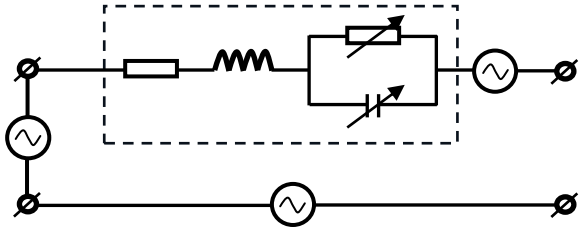


Рис. 1. Эквивалентная схема однодиодного смесителя

где (1-1') вход смесителя, Ес – ЭДС входного сигнала, 1 – эквивалентная схема полупроводникового смесительного диода, в которой: rs – сопротивление объема полупроводника, Ls – индуктивность объема полупроводника, rn – активное сопротивление p-n перехода, $C_б$ – барьерная емкость p-n перехода, Епч – эквивалентная ЭДС сигнала ПЧ, 2-2' выход смесителя, Ег – ЭДС сигнала гетеродина.

Комплексное сопротивление смесительного диода Zд на рис. 1 описывается выражением (2) вида:

$$Z_д = \frac{r_n \cdot \frac{1}{j\omega C_б}}{r_n + \frac{1}{j\omega C_б}} + j\omega L_s r_s = \frac{r_n + r_s(r_n^2\omega^2 C_б^2)}{r_n^2\omega^2 C_б^2 + 1} + j\,\frac{\omega L_s(r_n^2\omega^2 C_б^2 + 1) - r_n^2\omega C_б}{r_n^2 + \omega^2 C_б^2 + 1}. \quad (2)$$

Это сопротивление вносит сдвиг фаз в протекающие через них сигналы любых частот ω_1, в том числе и сигнал ПЧ с частотой $\omega_{пч}$, величина которого рассчитывается по формуле (3), полученной из (2) как отношение её действительной и мнимой частей:

$$\Delta\varphi_{пч} = arctg\,\frac{\omega_{пч} L_s(r_n^2\omega_{пч}^2 C_б^2 + 1) - r_n^2\omega_{пч}^2 C_б^2}{r_n + r_s(r_n^2\omega_{пч}^2 C_б^2 + 1)}. \quad (3)$$

Из формулы (3) видно, что сдвиг фаз, вносимый смесительным диодом, должен носить нелинейный характер, так как функция tg в математике принципиально нелинейна и определяется его реактивностями L_s и $C_б$. При этом, если L_s величина постоянная и не зависит от амплитуды сигнала, то величина барьерной емкости $C_б$ зависит от амплитуды сигнала, приложенного к p-n переходу.

Как показали лабораторные инструментальные исследования абсолютных сдвигов фаз смесительных диодов, основную величину этих сдвигов вносит барьерная емкость $C_б$ p-n перехода.

Комплексное сопротивление p-n перехода Z_n с параллельным соединением емкости $C_б$ и активного сопротивления Z_n имеет вид [4]:

$$Z_n = \frac{r_n \cdot \frac{1}{j\omega c_б}}{r_n + \frac{1}{j\omega c_б}} = \frac{\frac{r_n}{\omega_{пч}^2 c_б} + j\frac{r_n}{\omega_{пч} c_б}}{r_n^2 + \frac{1}{\omega_{пч} c_б}} = \frac{r_n}{r_n^2\omega_{пч}^2 C_б^2 + 1} - j\,\frac{r_n\omega_{пч} c_б}{r_n^2\omega_{пч}^2 C_б^2 + 1}. \quad (4)$$

Таким образом, главным источником переменного сдвига фаз $\Delta\varphi_{пч\ пер.}$ на ПЧ является параллельное соединение барьерной ёмкости $C_б$ и барьерного сопротивления rn p-n перехода, который описывается выражением (5) полученным из (4) как отношение её действительной и мнимой частей в виде:

$$\tan^{-1}\Delta\varphi_{пч\ пер.} = \frac{-r_n^2 c_б \omega_{пч}}{r_n} = r_n\omega c_б = 2\pi f r_n. \quad (5)$$

В работе [5] показано, что сдвиг фаз, вносимый барьерной ёмкостью – есть величина нелинейная.

Известно [6, 7], что инженерный расчёт величины барьерной ёмкости весьма трудоёмкий и сложный процесс, требующий знания параметров материала, из которого изготовлен диод, которые известны только на заводе-изготовителе.

Поэтому в работе [6] для оценочных расчётов барьерной ёмкости, как для диодов с p-n переходом, так и с барьером Шоттки рекомендована формула:

$$c_б = 10^{-12} a_k \left(\frac{1}{\varphi_k - U}\right)^n, \quad (6)$$

в которой a_k – зависит от типа диода, φ_k – термический потенциал, U – напряжение приложенное к p-n переходу. Для иллюстрации поведения барьерной ёмкости в работе [8, с. 265] при $U > \varphi_k$, на рис. 2 и 3 приведены её графики в диапазоне изменений напряжений «ноль» и «плюс 1В» приложенного к p-n переходу. Из графиков виден достаточно нелинейный характер ёмкости в зависимости от напряжения U приложенного к p-n переходу.

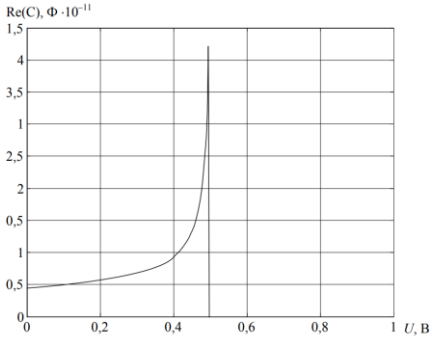


Рис. 2. График поведения барьерной ёмкости $C_б = 10^{-12} a_k (\frac{1}{\varphi_k - U})^n$ *в зависимости от напряжения U*

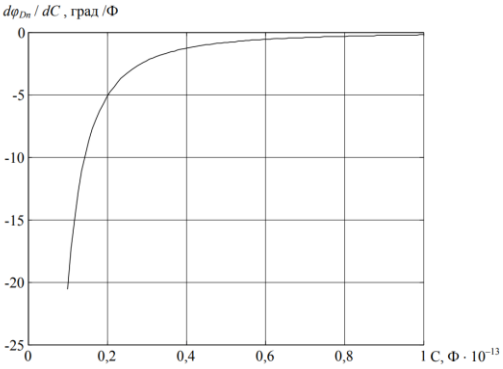


Рис. 3. График измерения $\frac{\partial \varphi_{Dn}}{\partial C_б}$

Из проведённого анализа следует главный вывод – при использовании диодов для умножения частоты, сдвиг фаз гармоник не равен произведению номера гармоники, умноженной на фазовый сдвиг умножаемого сигнала, требуется коррекция [9].

Анализ погрешностей измерения при определении сдвигов фаз смесителей известными методами. Как было показано выше, в настоящее время разработаны шесть инструментальных способов определения (измерения с последующим вычислением результатов) сдвигов фаз, возникающих в смесителе при гетеродинном преобразовании частоты.

Метод трёх смесителей [7], когда последовательно измеряются сдвиги фаз трёх смесителей, один из которых испытуемый, а два других вспомогательные с последующим вычислением по результатам измерений сдвига фаз испытуемого смесителя [10, с. 557].

Метод двойного преобразования частоты [10, с. 534], в котором с помощью векторного анализатора цепей (ВАЦ) измеряют общий сдвиг фаз двух последовательно соединённых испытуемых смесителя, включённых один в прямом, а второй в обратном направлениях с общим для обеих гетеродином, а полученный результат делят пополам и приписывают каждому из испытуемых смесителей.

Метод отражения [10, с. 557], при котором через испытуемый смеситель пропускается сигнал в прямом направлении, который на его выходе отражается от нагрузки с известным коэффициентом отражения и опять возвращается через смеситель на его вход. На входе смесителя и измеряют сдвиг фазы между прямым и отражённым сигналами, который, в свою очередь, равен сдвигу фаз испытуемого смесителя [10, с. 560].

Метод фазовой опоры, в котором используют ВАЦ-содержащий генератор комбинационных частот вырабатывающих гармоники путём умножения частоты опорного кварцевого генератора 10 МГц с дискретностью гармоник 10 МГц. Эти гармоники используют для получения испытательного сигнала исследуемого смесителя и опорного сигнала ПЧ, который сравнивают по фазе с выходным сигналом ПЧ этого же смесителя, в результате чего и измеряют их фазовый сдвиг [10, с. 564].

Сигнал гетеродина также сравнивают по фазе с соответствующей ему гармоникой и измеряют фазовый сдвиг между ними. Затем вычитают этот фазовый сдвиг из измеренного фазового сдвига сигнала ПЧ, в результате чего получают фазовый сдвиг смесителя, вносимый им в сигнал ПЧ при гетеродинном преобразовании частоты входного сигнала.

Сигналы всех гармоник когерентны между собой. Поэтому считают, что их фазовые сдвиги соответствуют фазе исходного сигнала опорного кварцевого генератора умноженной на номер гармоники

Но как было показано ранее, фазы гармоник отличаются друг от друга на величину $\Delta\varphi_{\text{п}}$, разную для каждой отдельной гармоники, что является прямой погрешностью измерений, так как измеряется не истинный фазовый сдвиг испытуемого смесителя, а его реплика, отличающаяся от истинного сдвига фаз на величину $\Delta\varphi_{\text{п}}$. Поэтому необходима таблица индивидуальной коррекции сдвигов фаз с дискретностью, равной частоте опорного кварцевого генератора.

Все вышеперечисленные методы испытаний смесителей имеют существенные недостатки [11, 12].

В методе трёх смесителей главная и существенная погрешность при измерении фазы возникает из-за согласования входных сопротивлений смесителей, собранных в пару между собой и входами каждой пары смесителей и входами ВАЦ, к которым они присоединяются для испытаний. Как показывает практика, погрешность измерения фазы при КСВ входов смесителей 1,5–2,0 может достигать 10–15 градусов.

Метод отражения применим для измерения параметров смесителей, в том числе и фазы, только в случае, когда потери преобразования смесителя в прямом и обратном направлении равны и, кроме того, не превышают 10 дБ.

Метод двойного преобразования частоты пригоден только для измерений двух совершенно идентичных по параметрам смесителей и совершенно непригоден для двух последовательно соединённых смесителей с разными потерями преобразования и, следовательно, с разными собственными сдвигами фаз [13]. Кроме того, даже при двух совершенно идентичных по параметрам смесителях, соединённых последовательно в пару, амплитуда испытательного сигнала на входе второго смесителя пары будет меньше, чем амплитуда испытательного сигнала на входе первого смесителя пары на величину его потерь преобразования. Это неизбежно приводит к разным сдвигам фаз у первого и второго смесителей пары.

В методе фазовой опоры главный недостаток состоит в том, что из-за нелинейности фазовых сдвигов между гармониками фазовой опоры, производятся измерения не самого фазового сдвига вносимого смесителем при гетеродинном преобразовании в сигнал ПЧ, а его реплики, отличающейся от истинного сдвига фаз на величину $\Delta\varphi_{\text{П}}$, разную для разных гармоник.

Все испытательные сигналы в таком методе являются когерентными, так как получены на основе частоты опорного кварцевого генератора 10 МГц.

В то же время известен метод измерения абсолютного сдвига фаз смесительного диода, лишенный вышеперечисленных недостатков. Этот метод [14] состоит в том, что одновременно измеряют сумму и разность сдвигов фаз двух смесителей, один из которых опорный, а другой испытуемый при общем гетеродине для обоих смесителей. При измерении суммы сдвига фаз испытуемой и опорной смесители соединяются вместе последовательно выходами ПЧ. На вход испытуемого смесителя подают испытательный сигнал,

который в результате прямого преобразования частоты в испытуемом смесителе и обратного преобразования частоты в опорном смесителе снимают с его выхода с частотой равной частоте входного испытательного сигнала. Далее его подают на один из двух входов фазометра, а на другой вход поступает входной испытательный сигнал, в результате чего измеряют общий сдвиг фаз (сумму) двух смесителей в виде:

$$\sum \varphi = (\varphi_1 + \varphi_2),$$

где φ_1 – сдвиг фаз испытуемого, а φ_2 – сдвиг фаз опорного смесителей. Затем измеряют разность сдвигов фаз между включёнными параллельно смесителями в диапазоне их промежуточных частот, в результате чего получают $\Delta\varphi$:

$$\Delta\varphi = \varphi_1 - \varphi_2.$$

После чего вычисляют истинные сдвиги фаз испытуемого смесителя φ_1, по формуле:

$$\varphi_1 = \frac{\sum \varphi + \Delta\varphi}{2}, \quad (7)$$

и опорного смесителя φ_2 по формуле:

$$\varphi_2 = \frac{\sum \varphi - \Delta\varphi}{2}. \quad (8)$$

Заключение Проанализирован сдвиг фаз, вносимый полупроводниковым смесительным диодом в процессе гетеродинного преобразования частоты входного сигнала в сигнал ПЧ. Показано, что этот сдвиг является нелинейным в заданном диапазоне частот, а величина сдвига определяется постоянным активным и индуктивным сопротивлениями объёма полупроводника и переменной барьерной ёмкостью p-n перехода, теоретический инженерный расчёт которой весьма сложен и не всегда возможен.

Рассмотрены известные инструментальные методы измерения сдвигов, вносимых смесительным диодом в сигнал ПЧ при гетеродинном преобразовании частоты, и показаны их недостатки [9, 19, 20].

Приведено описание нового способа измерений, который позволяет определять сдвиг фаз смесительного диода при преобразовании частоты без описанных ранее недостатков [19, 21].

Показано, что в случае применения диода для умножения частоты, сдвиги фаз каждой из гармоник не равны сдвигу фаз исходного сигнала, и их величина индивидуальна для каждой гармоники.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Патент РФ на изобретение № 2524049. Устройство для измерения комплексных коэффициентов передачи и отражения СВЧ–устройств с преобразованием частоты / *Коротков К.С., Левченко А.С., Мильченко Д.Н., Фролов Д.Р.;* опубликовано: 27.07.2014 г.
2. Радиоприемные устройства: учебник / под общей ред. В.И. Сифорова. – М.: Советское радио, 1974. – С. 502-507.
3. *Стриха В.И., Бузанева Е.В., Радзиевский Е.А.* Полупроводниковые приборы с барьером Шоттки. – М.: Советское радио, 1974. – С. 130-134.
4. *Валиев К.А. Паминцев Ю.И. Петров Г.В.* Применение контакта металл–полупроводник в электронике. – М.: Радио и связь, 1981. – 86 с.
5. *Медведев В.В. Пругло В.П. Уэдзин Р.И.* Методы и аппаратура для измерения параметров СВЧ-смесителей // Сб. ВИИФТРИ. Серия «Радиотехнические измерения». – М., 1973.
6. Полупроводниковые диоды. Параметры, методы измерений / под ред. Н.Н. Горюнова, Ю.Ю. Носова. – М.: Советское радио, 1968. – С. 196-198.
7. *Епифанов Е.И., Новиков В.А., Чемес Е.А.* Экспериментальные исследования радиочастотных смесителей // Радиотехника. – 1988. – № 10.
8. *Коротков К.С.* Методы определения комплексных коэффициентов передачи и отражения СВЧ–устройств и преобразованием частоты: дисс. … д-ра техн. наук. – Таганрог, 2002.
9. Диоды полупроводниковые СВЧ. Методы измерения электрических параметров. ГОСТ 19656.9-79. – М., 1987.

10. *Дансмор Джоэль П.* Настольная книга инженера. Измерения параметров СВЧ-устройств с использованием передовых методик векторного анализа цепей: пер. с англ. и научная редакция Е.Ю. Харитонова [и др.]. – М.: Техносфера, 2019. – 735 с. – (Мир радиоэлектроники / Минпромторг России; XVII–32). – ISBN 978-5-94836-505-3.
11. А.с. СССР № 1475347 G01R 27/28 с приоритетом от 13.12.86. Способ определения сдвига фаз четырехполюсников с преобразователем частоты / *Коротков К.С. Малышков В.Е. Суровенный В.Г.;* Зарегистр. 22.12.86.
12. *Chacran Jay, Tenenhltr R*. Recent advances in microwave mixers // IRE Intern. Conv. Rec. – 1961. – Vol. 9, No. 3.
13. А.с. 1475347 СССР, МКИ G01 R 27/28. Способ определения сдвига фаз четырёхполюсников с преобразованием частоты / *Коротков К.С., Малышков В.Е., Суровенный В.Г.*; заявл. 13.12.86.
14. Радиоизмерительные приборы 90/91. Каталог. Изделия промышленности средств связи. – 26-е изд. – М.: ЭКОС, 1989.
15. Панорамный измеритель разности фаз коэффициентов передачи и разности коэффициентов передачи четырехполюсников. РРК4–001/РРК4–003. Техническое описание и инструкция по эксплуатации ХВ1.405.032. – Краснодар, 1984.
16. Double Balanced Waveguide Mixer // Marki Microwave Company. E-mail mixers@markimicrowave.com, 2000.
17. Broad Band Mixers // Lorch Microwave Company. E-mail lorchsales@lorch.com, 2000.
18. Accurate Mixer Measurements Using the ENA RF Networks Analyzers' Frequency–Offset Mode / Application Note 5989–1420EN // Keysight Technologies. – 2014.
19. *Korotkov K.S., Frolov D.R.* A New Method of Measuring The Complex Parameters of Microwave // Measurement Techniques. – 2014. – Vol. 57, Issue 6.
20. Patent Application US 2012/0295548 A1 United States of America, Int. Cl. H 04 B 17/00. System For Characterizing Mixer or Converter Response / *Joel P. Dunsmore* (USA); assignee Agilent Technologies Inc. – 13/110,129; filed May 18, 2011; pub. Date Nov 22, 2012.
21. *Vyzulin S.A., Korotkov K.S.* To the New Phase Shifts in Mixer Measurement Method // Nato Advanced Research Workshop Microwave Physics and Technique. 30 sept – 5 oct, Sozopol Bulgaria, 1996.
22. Agilent 8510 Network Analyzer. Microwave Component Measurements – Amplitude and Phase Measurements of Frequency Translation Devices Using the Agilent 8510C Network Analyzer / Product Note 8510–7A // Agilent Technologies, Inc. – 2006.

REFERENCES

1. *Korotkov K.S., Levchenko A.S., Mil'chenko D.N., Frolov D.R.* Patent RF na izobretenie № 2524049. Ustroystvo dlya izmereniya kompleksnykh koeffitsientov peredachi i otrazheniya SVCh–ustroystv s preobrazovaniem chastity [Patent RF for invention No. 2524049. A device for measuring complex transmission and reflection coefficients of microwave devices with frequency conversion]; published: 07/27/2014.
2. Radiopriemnye ustroystva: uchebnik [Radio receivers. textbook], under the general ed. V.I. Siforova. Moscow: Sovetskoe radio, 1974, pp. 502-507.
3. *Strikha V.I., Buzaneva E.V., Radzievskiy E.A.* Poluprovodnikovye pribory s bar'erom Shottki [Semiconductor devices with a Schottky barrier]. Moscow: Sovetskoe radio, 1974, pp. 130-134.
4. *Valiev K.A. Pamintsev Yu.I. Petrov G.V.* Primenenie kontakta metall–poluprovodnik v elektronike [Application of metal–semiconductor contact in electronics]. Moscow: Radio i svyaz', 1981, 86 p.
5. *Medvedev V.V. Pruglo V.P. Uedzin R.I.* Metody i apparatura dlya izmereniya parametrov SVCh-smesiteley [Methods and equipment for measuring parameters of microwave mixers], *Sb. VIIFTRI. Ser. Radiotekhnicheskie izmereniya* [Collection of VIIFTRI. Series "Radio engineering measurements"]. Moscow, 1973.
6. Poluprovodnikovye diody. Parametry, metody izmereniy [Semiconductor diodes. Parameters and measurement methods], ed. by N.N. Goryunova, Yu.Yu. Nosova. Moscow: Sovetskoe radio, 1968, pp. 196-198.
7. *Epifanov E.I., Novikov V.A., Chemes E.A.* Eksperimental'nye issledovaniya radiochastotnykh smesiteley [Experimental studies of radio frequency mixers], *Radiotekhnika* [Radio engineering], 1988, No. 10.
8. *Korotkov K.S.* Metody opredeleniya kompleksnykh koeffitsientov peredachi i otrazheniya SVCh–ustroystv i preobrazovaniem chastoty: disc. … d-ra tekhn. nauk [Methods for determining the complex coefficients of transmission and reflection of microwave devices and frequency conversion: dr. of eng. sc. diss.]. Taganrog, 2002.
9. Diody poluprovodnikovye SVCh. Metody izmereniya elektricheskikh parametrov. GOST 19656.9-79 [Semiconductor microwave diodes. Methods of measuring electrical parameters. GOST 19656.9-79]. Moscow, 1987.

10. *Dansmor Dzhoel' P.* Nastol'naya kniga inzhenera. Izmereniya parametrov SVCh-ustroystv s ispol'zovaniem peredovykh metodik vektornogo analiza tsepey [The Engineer's Handbook. Measurements of parameters of microwave devices using advanced techniques of vector circuit analysis]: translated from English and scientific edition by E.Y. Kharitonov [et al.]. Moscow: Tekhnosfera, 2019, 735 p. (Mir radioelektroniki, Minpromtorg Rossii; XVII–32 [World of Radioelectronics / Ministry of Industry and Trade of Russia; XVII-32]). ISBN 978-5-94836-505-3.
11. *Korotkov K.S. Malyshkov V.E. Surovennyy V.G.* A.s. SSSR № 1475347 G01R 27/28 s prioritetom ot 13.12.86. Sposob opredeleniya sdviga faz chetyrekhpolyusnikov s preobrazovatelem chastoty [A.S. USSR No. 1475347 G01R 27/28 with priority from 12/13/86]; Register. 22.12.86.
12. *Chacran Jay, Tenenhltr R*. Recent advances in microwave mixers, *IRE Intern. Conv. Rec*., 1961, Vol. 9, No. 3.
13. *Korotkov K.S., Malyshkov V.E., Surovennyy V.G.* A.s. 1475347 SSSR, MKI G01 R 27/28. Sposob opredeleniya sdviga faz chetyrekhpolyusnikov s preobrazovaniem chastoty [A.S. 1475347 USSR, MKI G01 R 27/28. A method for determining the phase shift of four-pole frequency conversion]; application 13.12.86.
14. Radioizmeritel'nye pribory 90/91. Katalog. Izdeliya promyshlennosti sredstv svyazi [Radio measuring devices 90/91. Catalog. Products of the communications industry]. 26th ed. Moscow: EKOS, 1989.
15. Panoramnyy izmeritel' raznosti faz koeffitsientov peredachi i raznosti koeffitsientov peredachi chetyrekhpolyusnikov. RRK4–001/RRK4–003. Tekhnicheskoe opisanie i instruktsiya po ekspluatatsii HV1.405.032 [Panoramic meter of the phase difference of the transmission coefficients and the difference of the transmission coefficients of four-poles. RRK4–001/RRK4–003. Technical description and operating instructions HV1.405.032]. Krasnodar, 1984.
16. Double Balanced Waveguide Mixer, *Marki Microwave Company*. E-mail mixers@markimicrowave.com, 2000.
17. Broad Band Mixers, *Lorch Microwave Company*. E-mail lorchsales@lorch.com, 2000.
18. Accurate Mixer Measurements Using the ENA RF Networks Analyzers' Frequency–Offset Mode, Application Note 5989–1420EN, *Keysight Technologies,* 2014.
19. *Korotkov K.S., Frolov D.R.* A New Method of Measuring The Complex Parameters of Microwave // Measurement Techniques. – 2014. – Vol. 57, Issue 6.
20. *Joel P. Dunsmore* (USA). Patent Application US 2012/0295548 A1 United States of America, Int. Cl. H 04 B 17/00. System For Characterizing Mixer or Converter Response; assignee Agilent Technologies Inc. – 13/110,129; filed May 18, 2011; pub. Date Nov 22, 2012.
21. *Vyzulin S.A., Korotkov K.S.* To the New Phase Shifts in Mixer Measurement Method, *Nato Advanced Research Workshop Microwave Physics and Technique. 30 sept – 5 oct, Sozopol Bulgaria, 1996.*
22. Agilent 8510 Network Analyzer. Microwave Component Measurements – Amplitude and Phase Measurements of Frequency Translation Devices Using the Agilent 8510C Network Analyzer, Product Note 8510–7A, *Agilent Technologies, Inc*., 2006.

Статью рекомендовал к опубликованию д.т.н., профессор И.И. Турулин.

Сердюков Владимир Владимирович – Кубанский государственный университет; e-mail: serdyukov-vladimir1996@mail.ru; г. Краснодар, Россия; тел.: 89528524156; преподаватель.

Коротков Константин Станиславович – e-mail: korsar@phys.kubsu.ru; тел.: 89182513055; д.т.н.; профессор.

Serdukov Vladinir Vladimirovih – Kuban State University; e-mail: serdyukov-vladimir1996@mail.ru; Krasnodar, Russia; phone: +79528524156; teacher.

Korotkov Konstantin Stanislavovich – e-mail: korsar@phys.kubsu.ru; phone: +79182513055; dr. of eng. sc.; professor.

ПРАВИЛА ОФОРМЛЕНИЯ РУКОПИСЕЙ

1. Объем статьи должен быть не менее 12 и не более 18 страниц. Формат (А 4). Редактор ***Word 7 for Windows***, шрифт Times New Roman, размер 14, интервал 1,5. Авторы представляют в редакцию 1 экз. статьи и идентичный электронный вариант.

2. Названию статьи предшествует индекс УДК, соответствующий заявленной теме.

3. Текст статьи начинается с названия статьи (на русском и английском языках), фамилии, имени и отчества автора (полностью) и снабжается аннотацией на русском и английском языках объёмом ***не менее 250-300 слов***. В тексте аннотации указывается цель, задачи исследования и краткие выводы. В аннотации ***не следует*** давать ссылки на номер публикации в списке литературы к статье. После аннотаций приводятся ключевые слова (словосочетания), несущие в тексте основную смысловую нагрузку (на русском и английском языках).

4. В тексте статьи следует использовать минимальное количество таблиц и иллюстраций. Рисунок должен иметь объяснения значений всех компонентов, порядковый номер, название, расположенное под рисунком. В тексте на рисунок дается ссылка. Таблица должна иметь порядковый номер, заголовок, расположенный над ней. Данные таблиц и рисунков не должны дублировать текст. Формулы должны быть набраны ***в редакторе формул Word 7 for Windows.***

5. Цитаты тщательно сверяются с первоисточником и визируются автором на обратной стороне последней страницы: "Цитаты и фактический материал сверены". Подпись, дата.

6. Наличие пристатейного библиографического списка на русском и английском языках обязательно. ***Ссылок должно быть не менее 20-ти,*** из них на зарубежные источники – не менее 35 %. В тексте ссылки должны быть в квадратных скобках.

Примеры оформления литературы: а) для книг: фамилия, инициалы автора(ов), полное название книги, место, год издания, страницы; б) для статей: фамилия и инициалы автора(ов), полное название сборника, книги, газеты, журнала, где опубликована статья, место и год издания (сборника, книги), номер (для журнала), год и дата (для газеты), выпуск, часть (для сборника), страницы, на которых опубликована статья. Иностранная литература оформляется по тем же правилам.

Ссылки на неопубликованные работы не допускаются.

7. Рукопись должна быть тщательно вычитана. Редакционная коллегия оставляет за собой право при необходимости сокращать статьи, редактировать и отсылать авторам на доработку.

8. Статьи сопровождаются сведениями об авторе(ах) (фамилия, имя, отчество, ученое звание, должность, место работы, адрес, электронный адрес и номер телефона) на русском и английском языках.

9. Плата с аспирантов за публикацию рукописей не взымается.

Адрес журнала в Интернете: http://izv-tn.tti.sfedu.ru/.