



№2-2021

ISSN 1999-9429

ИЗВЕСТИЯ ЮФУ

ТЕХНИЧЕСКИЕ НАУКИ

- Алгоритмы обработки информации
- Связь, навигация и наведение
- Моделирование процессов и систем
- Информационный анализ
и распознавание образов
- Системы управления

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 связи, информационных технологий
и массовых коммуникаций

Научно-технический и прикладной журнал

Издается с 1995 года, до середины 2007 года под названием «Известия ТРТУ»

Подписной индекс 41970

№ 2 (219). 2021 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».

Редакционный совет

Каляев И.А. (гл. редактор); Курейчик В.В. (зам. гл. редактора); Курейчик В.М. (зам. гл. редактора); Бородянский И.М. (ученый секретарь); Абрамов С.М.; Агеев О.А.; Бабенко Л.К.; Веселов Г.Е.; Гонкальвес Ж.; Колесников А.А.; Коноплев Б.Г.; Левин И.И.; Макаревич О.Б.; Маркович И.И.; Микрин Е.А.; Никитов С.А.; Обуховец В.А.; Осипов Г.С.; Панатов Г.С.; Панич А.Е.; Петров В.В.; Петровский А.Б.; Пшихопов В.Х.; Редько В.Г.; Румянцев К.Е.; Саламах М.; Солдатов А.В.; Стемпковский А.Л.; Сухинов А.И.; Сысоев В.В.; Тарасов С.П.; Фрадков А.Л.; Хашемипур М.; Чаплыгин Ю.А.; Чередниченко Д.И.; Четверушкин Б.Н.; Чичков Б.Н.

Учредитель Южный федеральный университет.

Издатель Южный федеральный университет.

Ответственные за выпуск: Клевцов С.И., Самойлов А.Н.

Технический редактор Ярошевич Н.В.

Оригинал-макет выполнен Ярошевич Н.В.

Дата выхода в свет 26.07.2021 г. Формат 70×108 $\frac{1}{16}$. Бумага офсетная.

Офсетная печать. Усл. печ. л. – 17,5. Уч.-изд. л. – 14,9.

Заказ № 8105. Тираж 250 экз.

Адрес издателя: 344091, г. Ростов-на-Дону, пр. Стачки, 200/1. Тел. 8(863)2478051.

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 (863) 247-80-51.

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)

© Южный федеральный университет, 2021

СОДЕРЖАНИЕ

РАЗДЕЛ I. АЛГОРИТМЫ ОБРАБОТКИ ИНФОРМАЦИИ

Л.К. Бабенко, Е.А. Толоманенко ГИБРИДНОЕ ШИФРОВАНИЕ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ СИММЕТРИЧНЫХ И ГОМОМОРФНЫХ ШИФРОВ	6
Н.А. Будко, М.Ю. Медведев, А.Ю. Будко РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДА ВЕКТОРНОГО АНАЛИЗА ЭМГ ПРЕДПЛЕЧЬЯ ДЛЯ ПОСТРОЕНИЯ ЧЕЛОВЕКО-МАШИННЫХ ИНТЕРФЕЙСОВ	18
В.И. Данильченко, Е.В. Данильченко, В.М. Курейчик МНОГОМЕРНЫЙ ПОИСК В ЗАДАЧЕ РАЗМЕЩЕНИИ ЭЛЕМЕНТОВ СБИС НА ОСНОВЕ ГЕНЕТИЧЕСКОГО АЛГОРИТМА.....	31
В.В. Курейчик, Вл.Вл. Курейчик ПРОГРАММНАЯ ПОДСИСТЕМА ДЛЯ РЕШЕНИЯ NP-СЛОЖНЫХ КОМБИНАТОРНО-ЛОГИЧЕСКИХ ЗАДАЧ НА ГРАФАХ.....	39
В.В. Курейчик, Вл.Вл. Курейчик, А.Э. Саак ЭВОЛЮЦИОННЫЙ АЛГОРИТМ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ДИСПЕТЧЕРИЗАЦИИ	50
Н.К. Полюянович, М.Н. Дубяго АЛГОРИТМ ОБУЧЕНИЯ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ФАКТОРНОГО ПРОГНОЗИРОВАНИЯ РЕСУРСА ИЗОЛЯЦИОННЫХ МАТЕРИАЛОВ СИЛОВЫХ КАБЕЛЬНЫХ ЛИНИЙ.....	59

РАЗДЕЛ II. СВЯЗЬ, НАВИГАЦИЯ И НАВЕДЕНИЕ

А.А. Ваганова, Н.Н. Кисель, А.И. Паньчев НАПРАВЛЕННЫЕ И ПОЛЯРИЗАЦИОННЫЕ СВОЙСТВА МИКРОПОЛОСКОВОЙ РЕКОНФИГУРИРУЕМОЙ АНТЕННЫ, ПЕРЕСТРАИВАЕМОЙ ПО ЧАСТОТЕ И ПОЛЯРИЗАЦИИ.....	74
А.Н. Зикий, А.С. Кочубей ГЕНЕРАТОР ДЕЦИМЕТРОВОГО ДИАПАЗОНА	84
А.О. Касьянов РЕЗУЛЬТАТЫ ЧИСЛЕННОГО ИССЛЕДОВАНИЯ ХАРАКТЕРИСТИК РАССЕЯНИЯ АНТЕННЫХ ОБТЕКАТЕЛЕЙ НА ОСНОВЕ МЕТАЛЛОДИЭЛЕКТРИЧЕСКИХ ДИФРАКЦИОННЫХ РЕШЕТОК	91
П.А. Курапов ГИБРИДНОЕ ИСПОЛНЕНИЕ ЗАПРОСОВ К АНАЛИТИЧЕСКИМ БАЗАМ ДАННЫХ.....	105
П.А. Пантелюк ДЕТЕКТИРОВАНИЕ ВЫБРОСОВ В МЕТОДЕ ПРЯМОЙ СТЕРЕО-ВИЗУАЛЬНОЙ ОДОМЕТРИИ НА БАЗЕ ИЕРАРХИЧЕСКОЙ КЛАСТЕРИЗАЦИИ	113

РАЗДЕЛ III. МОДЕЛИРОВАНИЕ ПРОЦЕССОВ И СИСТЕМ

В.А. Деркачев МОДЕЛЬ РАССЕЯНИЯ РАДИОЛОКАЦИОННЫХ СИГНАЛОВ ОТ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ	120
В.В. Янчич ИССЛЕДОВАНИЕ ВЛИЯНИЯ КРЕПЛЕНИЯ ЭЛЕКТРОМЕХАНИЧЕСКОГО ПРЕОБРАЗОВАТЕЛЯ НА ХАРАКТЕРИСТИКИ ПЬЕЗОЭЛЕКТРИЧЕСКОГО ДАТЧИКА УСКОРЕНИЯ.....	130

РАЗДЕЛ IV. ИНФОРМАЦИОННЫЙ АНАЛИЗ И РАСПОЗНАВАНИЕ ОБРАЗОВ

А.Н. Каркищенко, В.Б. Мнухин ОЦЕНКА ПОГРЕШНОСТИ ПРИ МНОЖЕСТВЕННОМ СРАВНЕНИИ ЗАШУМЛЕННЫХ ИЗОБРАЖЕНИЙ.....	141
Ю.А. Кравченко, А.М. Мансур, Ж.Х. Мохаммад ВЕКТОРИЗАЦИЯ ТЕКСТА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ	154
А.К. Мельников ОГРАНИЧЕНИЕ КОЛИЧЕСТВА РАЗЛИЧНЫХ ОПРОБУЕМЫХ ВЕКТОРОВ ДЛЯ ПОЛУЧЕНИЯ ВСЕХ РЕШЕНИЙ СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ ВТОРОЙ КРАТНОСТИ НА МНОГОПРОЦЕССОРНОЙ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЕ	167

РАЗДЕЛ V. СИСТЕМЫ УПРАВЛЕНИЯ

З.Р. Майрансаев, А.Б. Чернышев ОБОБЩЕННЫЙ КРУГОВОЙ КРИТЕРИЙ АБСОЛЮТНОЙ УСТОЙЧИВОСТИ РАСПРЕДЕЛЕННЫХ СИСТЕМ.....	182
Д.Е. Чикрин, А.А. Егорчев СРАВНЕНИЕ МЕТОДОЛОГИЙ ПРОЕКТИРОВАНИЯ СВЕРХУ–ВНИЗ И СНИЗУ–ВВЕРХ ПРИ РАЗРАБОТКЕ СИСТЕМ ADAS	189

CONTENT

SECTION I. INFORMATION PROCESSING ALGORITHMS

L.K. Babenko, E.A. Tolomanenko HYBRID ENCRYPTION BASED ON SYMMETRIC AND HOMOMORPHIC CIPHERS	6
N.A. Budko, M.Yu. Medvedev, A.Yu. Budko DEVELOPMENT AND RESEARCH OF THE METHOD OF VECTOR ANALYSIS OF EMG OF THE FOREARM FOR CONSTRUCTION OF HUMAN-MACHINE INTERFACES	19
V.I. Danilchenko, Y.V. Danilchenko, V.M. Kureichik AUTOMATED STRUCTURAL-PARAMETRIC SYNTHESIS OF A STEPPED DIRECTIONAL RESPONDER ON CONNECTED LINES BASED ON A GENETIC ALGORITHM	31
V.V. Kureichik, VI.VI. Kureichik SOFTWARE SUBSYSTEM FOR SOLVING NP-COMPLEX COMBINATORIAL LOGIC PROBLEMS ON GRAPHS	40
V.V. Kureichik, VI.VI. Kureichik, A.E. Saak AN EVOLUTIONARY ALGORITHM FOR SOLVING THE DISPATCHING PROBLEM.....	51
N.K. Poluyanovich, M.N. Dubyago ALGORITHM FOR TRAINING THE ARTIFICIAL NEURAL NETWORK OF FACTOR PREDICTING THE POWER CABLE LINES INSULATING MATERIALS LIFE	60

SECTION II. COMMUNICATION, NAVIGATION AND GUIDANCE

A.A. Vaganova, N.N. Kisel, A.I. Panychev DIRECTIONAL AND POLARISATION PROPERTIES OF A MICROSTRIP RECONFIGURABLE ANTENNA WITH TUNABLE FREQUENCY AND POLARISATION	74
A.N. Zikiy, A.S. Kochubey DECIMETER RANGE GENERATOR	84
A.O. Kasyanov RESULTS OF NUMERICAL STUDY OF SCATTERING CHARACTERISTICS IN ANTENNA RADOMES BASED ON METAL-DIELECTRIC GRATINGS	92
P.A. Kurapov HYBRID EXECUTION OF QUERIES TO ANALYTICAL DATABASES	105
P.A. Panteliuk OUTLIERS DETECTION IN THE STEREO-VISUAL ODOMETRY BASED ON HIERARCHICAL CLUSTERIZATION.....	113

SECTION III. MODELING OF PROCESSES AND SYSTEMS

V.A. Derkachev MODEL OF SCATTERING OF RADAR SIGNALS FROM UAV	120
V.V. Yanchich RESEARCH OF ELECTROMECHANICAL TRANSDUCER'S FASTENING INFLUENCE TO THE ACCELERATION PIEZOSENSOR'S CHARACTERISTICS .	130

SECTION IV. INFORMATION ANALYSIS AND PATTERN RECOGNITION

A.N. Karkishchenko, V.B. Mnukhin ERROR ESTIMATION FOR MULTIPLE COMPARISON OF NOISY IMAGES.....	141
Yu.A. Kravchenko, A.M. Mansour, J.H. Mohammad TEXT VECTORIZATION USING DATA MINING METHODS.....	155

A.K. Melnikov	
LIMITING THE NUMBER OF DIFFERENT TEST VECTORS TO OBTAIN ALL SOLUTIONS OF A SYSTEM OF THE SECOND MULTIPLICITY LINEAR EQUATIONS ON MULTIPROCESSOR COMPUTER SYSTEM.....	168

SECTION V. CONTROL SYSTEMS

Z.R. Mayransaev, A.B. Chernyshev	
GENERALIZED CIRCULAR CRITERION FOR THE ABSOLUTE SUSTAINABILITY OF DISTRIBUTED SYSTEMS	182
D.E. Chickrin, A.A. Egorchev	
TOP-DOWN VS BOTTOM-UP METHODOLOGIES FOR ADAS SYSTEM DESIGN.....	190

Раздел I. Алгоритмы обработки информации

УДК 004.021

DOI 10.18522/2311-3103-2021-2-6-18

Л.К. Бабенко, Е.А. Толоманенко

ГИБРИДНОЕ ШИФРОВАНИЕ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ СИММЕТРИЧНЫХ И ГОМОМОРФНЫХ ШИФРОВ*

Целью данной работы является разработка и исследование гибридного алгоритма шифрования на основе совместного применения симметричного алгоритма шифрования Кузнечик и гомоморфного шифрования (Схемы Джендри или схемы BGV). Такой алгоритм шифрования может быть востребован в ситуациях ограниченных вычислительных ресурсов. Смысл заключается в том, что при правильном выражении основных операций симметричного алгоритма шифрования через булевы функции, появляется возможность на передающей стороне зашифровать данные симметричным шифром, а секретный ключ шифрования – гомоморфным. В таком случае на стороне приема можно провести манипуляции так, чтобы исходное зашифрованное сообщение оказалось также зашифровано только гомоморфным шифром. При этом симметричное шифрование снимается, но информация остается недоступной обрабатывающему ее узлу. Такое свойство секретности позволяет проводить ресурсоемкие операции на мощном вычислительном узле, предоставляя гомоморфно зашифрованные данные для малоресурсного узла с целью их последующей обработки в зашифрованном виде. В статье представлен разработанный гибридный алгоритм. В качестве симметричного алгоритма шифрования использован алгоритм шифрования Кузнечик, являющийся частью стандарта ГОСТ Р 34.12 – 2015. Для того, чтобы иметь возможность применять гомоморфное шифрование к данным, зашифрованным шифром Кузнечик, S-блок замены алгоритма Кузнечик представлен в булевом виде с использованием полинома Жегалкина. Также линейное преобразование L представлено в виде последовательности выполнения простейших операций сложения и умножения над преобразуемыми данными. Первичное моделирование разрабатываемого алгоритма было проведено на упрощенной версии алгоритма Кузнечик S-KNI.

Криптография; блочный шифр; симметричный шифр; алгоритм Кузнечик; гомоморфное шифрование; гибридное шифрование; полином Жегалкина; булева функция.

L.K. Babenko, E.A. Tolomanenko

HYBRID ENCRYPTION BASED ON SYMMETRIC AND HOMOMORPHIC CIPHERS

The purpose of this work is to develop and research a hybrid encryption algorithm based on the joint application of the symmetric encryption algorithm Kuznyechik and homomorphic encryption (Gentry scheme or BGV scheme). Such an encryption algorithm can be useful in situations with limited computing resources. The point is that with the correct expression of the basic operations of the symmetric encryption algorithm through Boolean functions, it becomes possible on the transmitting side to encrypt the data with a symmetric cipher, and the secret encryption key - with a homomorphic one. In this case, manipulations can be carried out on the receiving side so that the original encrypted message is also encrypted only with a homomorphic cipher. In this case, symmetric encryption is removed, but the information remains inaccessible to the node that processes it. This property of secrecy makes it possible to carry out resource-intensive operations on

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта РФФИ №19-37-90148.

a powerful computing node, providing homomorphically encrypted data for a low-resource node for the purpose of their subsequent processing in encrypted form. The article presents the developed hybrid algorithm. As a symmetric encryption algorithm, Kuznyechik encryption algorithm is used, which is part of the GOST R34.12 - 2015 standard. In order to be able to apply homomorphic encryption to data encrypted with the Kuznyechik cipher, the Kuznyechik algorithm S-boxes is presented in a boolean form using the Zhegalkin polynomial. Also, the linear transformation L is presented in the sequence form of performing the simplest operations of addition and multiplication on the transformed data. The primary modeling of the developed algorithm was carried out on a simplified version of the KuznyechikS-KN1 algorithm.

Cryptography; block cipher; symmetric cipher; Kuznyechik cipher; homomorphic encryption; hybrid encryption; Zhegalkin polynomial; Boolean function.

Введение. В последнее десятилетие стремительно развиваются информационные технологии. Активно развивается отрасль интернета вещей (IoT), робототехнические системы, сенсорные сети. Так как подобные устройства, как правило обладают ограниченным вычислительным ресурсом, то в первую очередь при разработке таких средств внимание уделяется вопросам безопасности, и в частности вопросам шифрования данных. Об актуальности проблем, связанных с обеспечением криптографической надежности, говорит, например, ряд докладов, представленных на ежегодной научной конференции «РусКрипто-2021» [1–3].

Область применения подобных устройств разнообразна, затрагивает различные сферы деятельности человека. Это и контроль экологических параметров окружающей среды, и контроль медицинских показателей пациентов, и контроль доступа к различным удаленным системам мониторинга, геологоразведочные работы, и многое другое. Часто подобные устройства обрабатывают и передают информацию, которая может носить конфиденциальный характер (например, медицинские показатели пациента, которые составляют врачебную тайну) или информацию, передача которой в открытом виде может быть скомпрометирована или изменена с целью кибератаки (например, при мониторинге параметров объекта в режиме реального времени). Это явным образом показывает необходимость разработки и внедрения алгоритмов и протоколов, направленных на защиту передаваемых и обрабатываемых данных.

Частично данный вопрос решается с использованием легковесной криптографии [4]. Тем не менее, легковесная криптография не позволяет обрабатывать данные в зашифрованном виде. Это означает, что устройство должно иметь механизмы для зашифрования и расшифрования информации с целью ее обработки. Кроме того, секретный ключ шифрования должен быть у всех узлов, участвующих в обработке информации. Это в свою очередь накладывает необходимость применения стойких, но при этом ресурсонезатратных протоколов формирования и распределения ключа.

Одним из вариантов решения сложившейся проблемы казалось появление нового вида шифров, названных гомоморфными [5–9]. Впервые гомоморфная схема шифрования была предложена в 2009 году сотрудником ИВМ Крейгом Джендри. Данная схема представляет собой асимметричную полностью гомоморфную схему шифрования на основе идеальных решеток [10]. В схеме Джендри имелся явный недостаток. При выполнении многократных действий в данных накапливалась ошибка. При превышении определенного порогового значения данные становились недешифруемыми. В 2012 году Джендри показал, как можно выполнять самокоррекцию шифров [11] для того, чтобы данные по-прежнему могли подвергаться обратному преобразованию. Гомоморфные шифры позволяют обрабатывать информацию в зашифрованном виде. В частично гомоморфной схеме доступен только один вид операций для обработки. Как правило, это сложение или умножение. В полностью гомоморфной схеме доступны обе операции для выполнения преобразований [12]. Однако у гомоморфных шифров есть ряд особенностей – это большая объем памяти, который требуется для выполнения гомоморфных опера-

ций. Данный факт напрямую связан с ограничением применения гомоморфных шифров для защиты информации в малоресурсной отрасли [13, 14]. Решением сложившейся проблемы может являться использование гибридных схем шифрования, сочетающих в себе использование симметричных и гомоморфных шифров. В 2012 году Джентри показал, как можно использовать схему полностью гомоморфного шифрования совместно с алгоритмом шифрования AES [15].

В настоящей статье предлагается рассмотреть возможность создания гибридной схемы шифрования на основе использования отечественного стандарта шифрования данных ГОСТ Р 34.12-2015 и схемы гомоморфного шифрования BGV.

Статья организована следующим образом. В первой главе приводится краткое описание тех алгоритмов шифрования, которые предполагается использовать при создании гибридной схемы. Во второй главе описывается общая формулировка протокола обмена данными между двумя узлами сети при использовании разрабатываемой гибридной схемы шифрования. В третьей главе представлена разработка алгоритмов для представления базовых алгоритмов шифрования Кузнечик в виде конъюнктивной нормальной формы (КНФ). В главе 4 представлен разработанный алгоритм для гибридной схемы.

Постановка задачи. Разработать гибридный алгоритм шифрования на основе использования симметричного и гомоморфного видов шифров для возможности осуществления операций в малоресурсных системах. В качестве симметричного шифра предполагается использовать алгоритм шифрования Кузнечик, входящий в состав отечественного стандарта шифрования данных ГОСТ Р 34.12-2015, а в качестве алгоритма гомоморфного шифрования схему BGV.

1. Краткое описание используемых алгоритмов шифрования

1.1. Алгоритм шифрования Кузнечик. Алгоритм шифрования Кузнечик представляет собой симметричный блочный шифр со следующими параметрами: длина блока – 128 бит, длина ключа – 256 бит, количество раундов шифрования – 9, схема построения – SP-сеть. Каждый раунд шифрования содержит три базовые операции: сложение данных с раундовым ключом посредством операции XOR, замена байтов с помощью S-блока замены (операция S) и линейное преобразование (операция L). Схематичное описание одного раунда шифрования для алгоритма Кузнечик приведено на рис. 1.

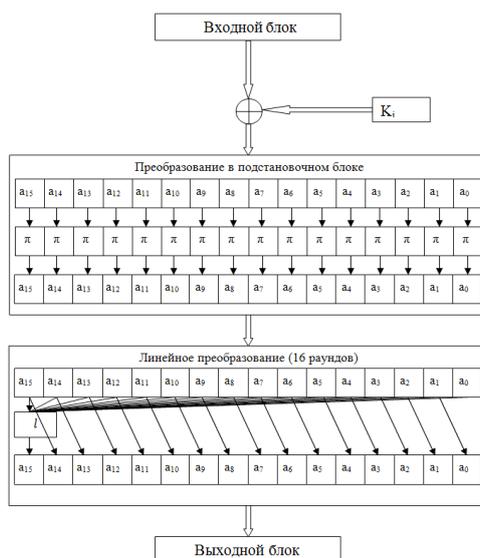


Рис. 1. Один раунд шифрования

Для SP-сети всегда требуется раундовых ключей на 1 больше, чем используемое количество раундов шифрования, поэтому всего необходимо 10 раундовых ключей. Они вырабатываются из исходного 256-битного мастер-ключа с применением схемы Фейстеля как показано на рис. 2. В каждом раунде используется функция F, состоящая из трех операций: хог преобразуемых данных с раундовой константой, преобразование с помощью S-блока замены и линейное преобразование L, которые изображены на рис. 3. С подробным описанием работы алгоритма Кузнечик можно познакомиться в [16, 17].

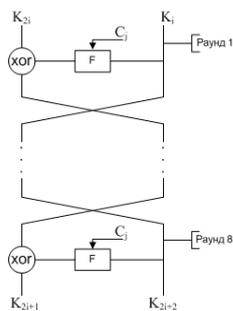


Рис. 2. Сеть Фейстеля для выработки раундовых ключей

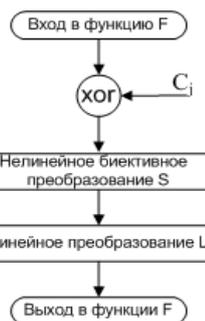


Рис. 3. Функция F, используемая в сети Фейстеля

1.2. Упрощенный алгоритм шифрования Кузнечик S-KN1. Упрощенный алгоритм Кузнечик (S-KN1) был разработан коллективом кафедры БИТ ИКТИБ ЮФУ и впервые представлен на международной конференции SIN-2017 [18]. Схема S-KN1 полностью повторяет структуру оригинального шифра Кузнечик и представлена на рис. 4.

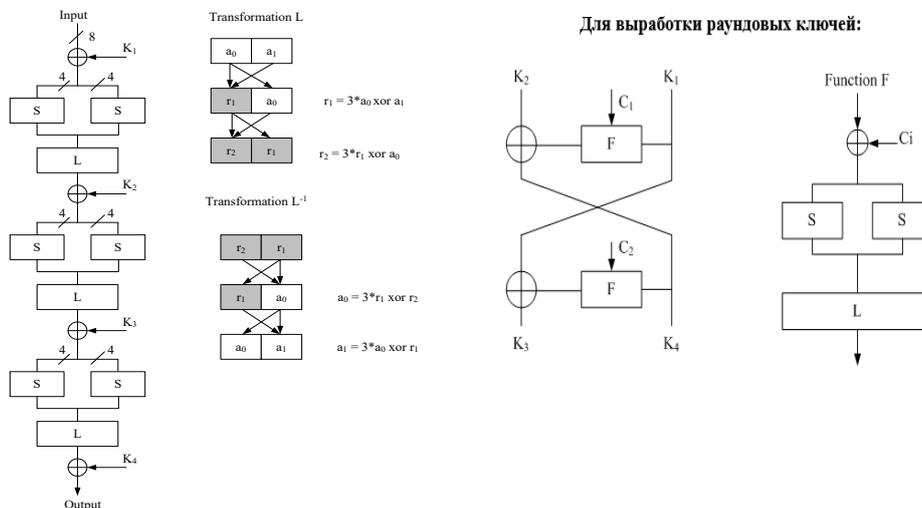


Рис. 4. Схема упрощенного алгоритма S-KN1

S-KN1 представляет собой симметричный блочный шифр, основанный на SP-сети и преобразовании информации в трех раундах. Длина преобразованного блока 8 бит; длина секретного ключа составляет 16 бит. Алгоритм начинает работу со сложения исходных данных с подключом K1. За этой операцией следуют три раунда шифрования. В каждом раунде сообщение разбивается на два полубайта, ка-

ждый из которых проходит через блок замены S , затем части объединяются и проходят через линейное преобразование L . Каждый раунд заканчивается сложением с соответствующим раундовым подключом.

1.3. Схема BGV(Brakerski-Gentry-Vaikuntanathan). Схема BGV была предложена тройкой ученых, в состав которых входит автор первого алгоритма гомоморфного шифрования Крейг Джентри [19]. Схема BGV впервые предложила решить проблему недешифруемости текстов без использования самокоррекции, вместо этого используются две другие операции: смена ключа и смена модуля. В схеме BGV множество открытых текстов представляется в виде кольца R_p , а множество зашифрованных текстов – в виде кольца R_q . При этом числа p и q являются простыми и связаны соотношением: $q \equiv 1 \pmod{p}$. Для реализации схемы BGV используется пять базовых операций.

1. Операция генерации ключа KeyGen

$s', e \leftarrow \chi$;
 $a \leftarrow R_q$;
 $b \leftarrow -(as' + pe)$;
 $sk \leftarrow s = (1, s')$ // приватный ключ
 $pk \leftarrow k = (b, a)$ // публичный ключ

2. Операция шифрования Enc (выполняется с использованием публичного ключа получателя)

$r \leftarrow R_p$ // случайное число
 $e \leftarrow \chi^2$ // распределение хи-квадрат
 $m \leftarrow (m, 0)$ // шифруемое значение
 $c \leftarrow m + kr + pe$ // результат шифрования

3. Операция расшифрования Dec (выполняется с использованием приватного ключа получателя)

$m \leftarrow [[\langle c|s \rangle]_q]_p$ // скалярное произведение зашифрованного сообщения и приватного ключа берется сначала по модулю q , а затем берется по модулю p .

4. Операция смены ключа KeySwitch. Применяется для коррекции чрезмерного роста длин ключей: $[[\langle c|s \rangle]_q]_p = [[\langle c'|s' \rangle]_q]_p$

5. Операция смены модуля ModulusSwitch. Применяется для уменьшения зашумленности шифртекста. Заменяет пару c, q на пару c', q' так, что: $[[\langle c|s \rangle]_q]_p = [[\langle c'|s \rangle]_{q'}]_p$

При этом для двух текстов m_1, m_2 зашифрованных как c_1, c_2 выполняются соотношения: $m_1 + m_2 = [[\langle c_1 + c_2 | s \rangle]_q]_p$ и $m_1 \otimes m_2 = [[c_1 \otimes c_2 | s \otimes s]_q]_p$.

Таким образом, сложению текстов соответствует сложение шифртекстов, а умножению текстов — тензорное произведение шифртекстов и изменение секретного ключа s на $s \otimes s$. Таким образом, в процессе гомоморфного шифрования на каждом уровне по умножению производится смена как модулей, так и ключей.

2. Протокол обмена данными между двумя узлами сети при использовании разрабатываемой гибридной схемы шифрования. На основании идеи Джентри, а также с использованием особенностей сенсорных сетей, был разработан протокол обмена данными между двумя узлами сети при использовании разрабатываемой гибридной схемы шифрования. Разработанный протокол предназначен для задач сети, в которой есть узел, обладающий потенциально мощной вычислительной мощностью (База) и узлы с небольшим вычислительным ресурсом (например, сенсоры). Пример такой схемы представлен на рис. 5.

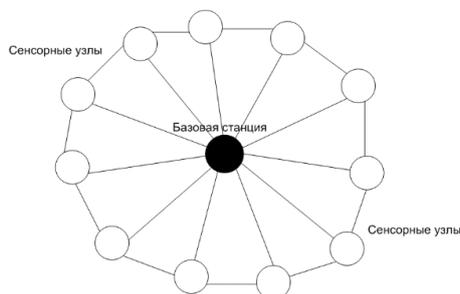


Рис. 5. Пример конфигурации сети, предназначенной для использования разработанного протокола

Важно отметить, что предложенный протокол будет практичным только в том случае, если ни база, ни сенсоры не владеют информацией о секретном ключе. То есть в данном случае база выступает в качестве мощного вычислительного ресурса. Она производит вычисления над данными, не зная самих данных (по аналогии с облачными вычислениями). Например, такие сенсоры нужны для секретной разведки. Тогда все базы оперируют открытыми ключами, не зная саму информацию. А вскрыть данные может только доверенное лицо, у которого есть секретный ключ. В противном случае теряется весь смысл разработанного протокола.

Протокол 1:

На стороне **Сенсорного узла:**

1. Определяются стартовые данные Data
2. Генерируется секретный ключ K
3. Данные шифруются с использованием алгоритма Кузнечик: $Enc_{Kuzn}(Data, K)$
4. Ключ K шифруется с использованием полностью гомоморфного шифрования $Enc_FHE(K)$
5. От сенсорного узла к базовой станции передаются значения $Enc_{Kuzn}(Data, K)$ и $Enc_FHE(K)$

На стороне **Базовой станции:**

1. Каждый бит зашифрованного сообщения $Enc_{Kuzn}(Data, K)$ шифруется с использованием полностью гомоморфного шифрования $Enc_FHE(Enc_{Kuzn}(Data, K))$
2. Выполняются функции расшифрования симметричного шифра с использованием гомоморфных операций $Enc_FHE(Enc_{Kuzn}(Data, K)) = Enc_FHE(Data)$. Данные остаются зашифрованы только гомоморфным шифрованием. Данные возвращаются Сенсорному узлу.

В результате сенсорные узлы могут производить вычисления над данными, передаваемыми друг другу. Если требуется производить коррекцию зашифрованных данных (например, если сенсорные датчики выполняют измерение температуры, то можно таким же образом обработать и передать сенсорному узлу корректирующие дельты для измеряемых параметров). Основное преимущество разработанной модели состоит в том, что она уменьшает размер зашифрованных данных, передаваемых между сенсорным узлом и базовой станцией. А также это сокращает количество ресурсов, требуемых на стороне сенсора.

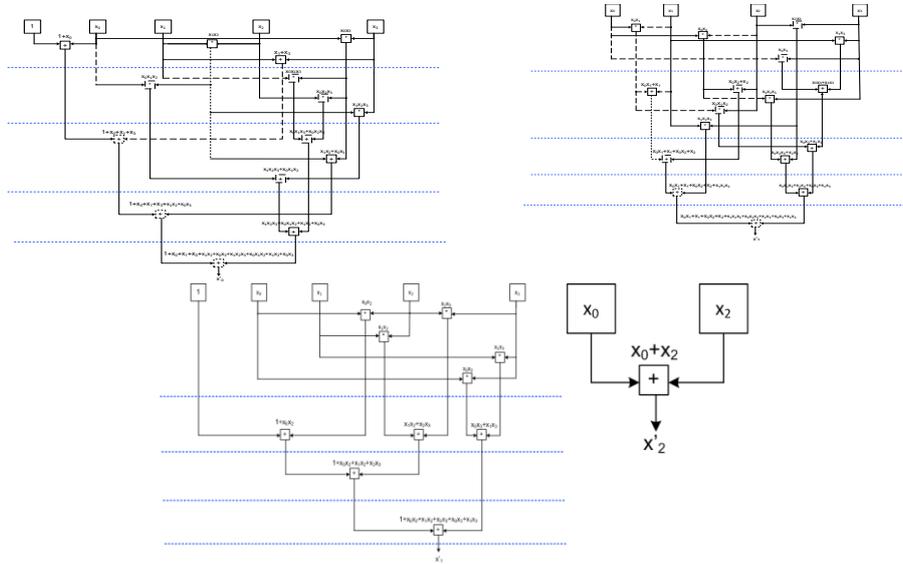


Рис. 8. Схемы реализации полиномов для алгоритма S-KN1

Так, из рис. 8 видно, что для бита x_0 требуется 2 уровня умножения. Сложение по сравнению с умножением может вызвать переполнение только в одном разряде. Поэтому в данном случае глубиной сложения ($=5$) можно пренебречь. Для x_1 требуется 1 уровень умножения, а для x_2 требуется только сложение. Аналогичным образом строятся схемы и для других полиномов.

2.2 Алгоритм построения КНФ для линейного преобразования L. Линейное преобразование L заключается в полиномиальном перемножении констант со значениями данных в текущем блоке. Так как в случае гомоморфного шифрования каждый бит будет зашифрован по отдельности, то нам необходимо представить результат перемножения полиномов A и B по модулю F в виде сумм для битовых произведений (рис. 9). Можно видеть, что для получения каждого следующего значения слагаемого нам необходимо произвести сдвиг влево на 1 разряд в регистре множителя и добавить нормализующий полином, умноженный на старший разряд, вытолкнутый из регистра. В таком случае нормализация произойдет только в том случае, если вытолкнутый разряд был равен единицы. Каждый бит результата необходимо домножить на очередной бит множителя. Таким образом, для 4-разрядного полинома каждый бит результата будет получен в результате 7 умножений и 6 сложений (рис. 9).

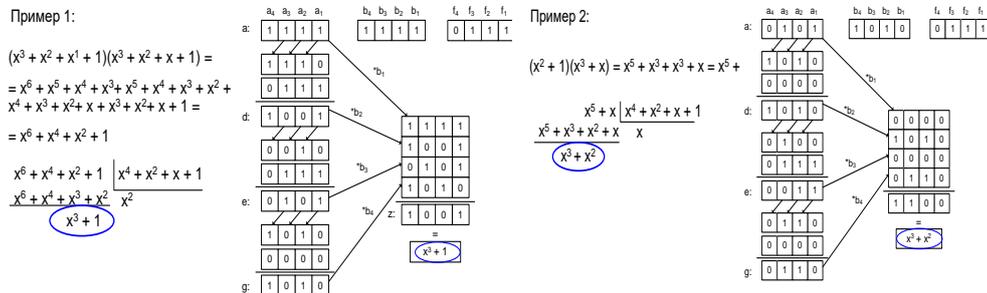


Рис. 9. КНФ для преобразования L в алгоритме S-KN1

Таким образом получаем 4 уровня умножения. Так как для применения операции L требуется применить полиномиальное перемножение дважды, то итоговый уровень умножения составит 8 шагов (рис. 10). Аналогичным образом формируются формулы для описания работы преобразования L оригинального шифра Кузнечик.

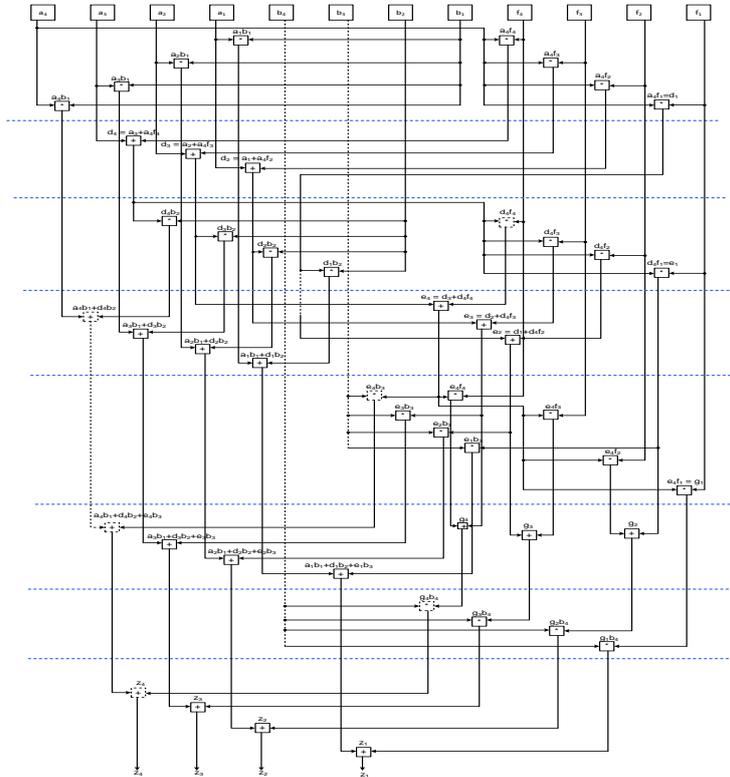


Рис. 10. Уровни умножения для побитного выполнения преобразования L

3. Алгоритмы работы разработанной гибридной схемы. На основании использования разработанных битовых умножений была разработана гибридная схема для расшифрования симметричного алгоритма с использованием гомоморфного шифрования. Схема состоит из двух частей: схемы выработки раундовых ключей и самой схемы расшифрования. На рис. 11 представлена схема выработки раундовых ключей. Каждый квадратик схемы – это один бит, зашифрованный гомоморфно. В результате применения разработанных схем мы получаем 4 раундовых ключа $Key1$, $Key2$, $Key3$, $Key4$.

Аналогичным образом была разработана схема для расшифрования данных S-KN1, зашифрованных гомоморфно. Каждый квадратик схемы – это один бит, зашифрованный гомоморфно. В результате применения разработанной схемы мы получаем 8 битов с $z1$ по $z8$, которые представляют собой открытые данные сенсорного узла, зашифрованные гомоморфно (рис. 12). Далее узел может обрабатывать эти данные (предпочтительно складывать в силу ограниченности ресурса сенсора) до тех пор, пока доверенное лицо не захочет эти данные считать и расшифровать. Аналогичным образом формируется алгоритм для оригинального шифра Кузнечик.

Раздел I. Алгоритмы обработки информации

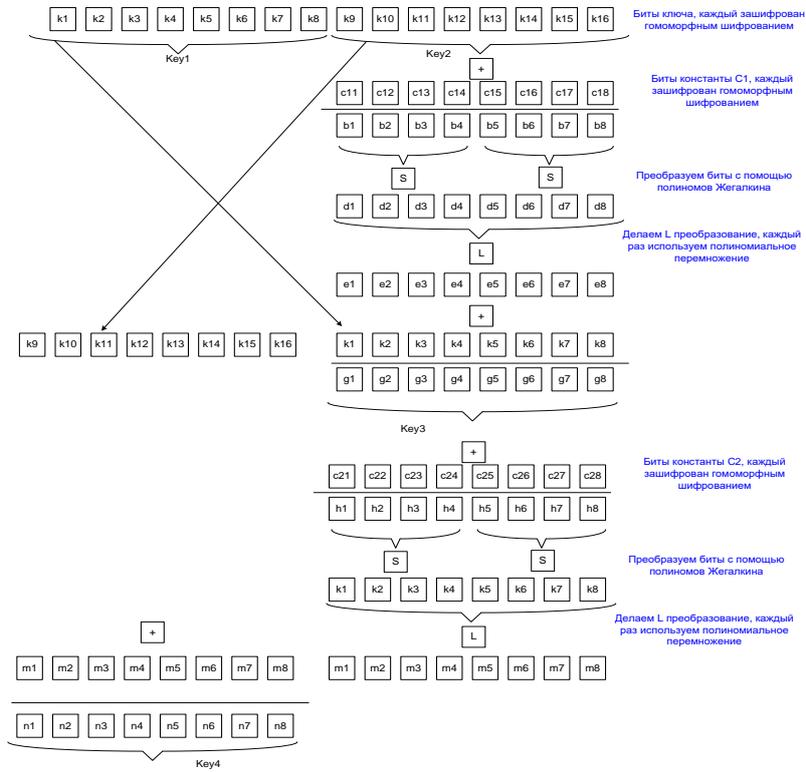


Рис. 11. Схема выработки раундовых ключей

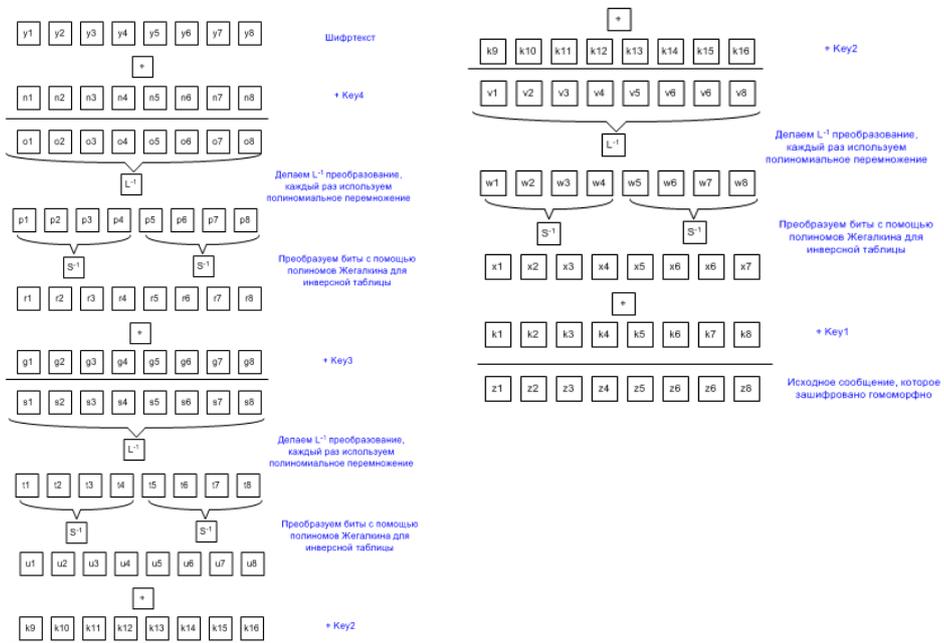


Рис. 12. Схема расшифрования данных

Выводы. В данной статье описана разработка алгоритма гибридного шифрования на основе использования симметричного алгоритма шифрования Кузнечик и гомоморфной схемы BGV для возможности безопасной обработки и передачи информации в малоресурсных системах. Описаны алгоритмы выработки полиномов Жегалкина для S-блока замены алгоритма Кузнечик и S-KN1, схема представления преобразования L в виде последовательности простейших арифметических операций, гибридная схема для расшифрования симметричного алгоритма с использованием гомоморфного шифрования, схема выработки ключей. В результате применения предложенной гибридной возможна передача симметрично зашифрованных данных и гомоморфно зашифрованного ключа шифрования, при этом на принимающей стороне зашифрованное сообщение остается также зашифровано только гомоморфным шифром. При снятии симметричного шифрования информация будет недоступна обрабатывающему узлу. Все вышеперечисленное обеспечивает передачу гомоморфно зашифрованных данных между ресурсоемким вычислительным узлом и малоресурсными узлами для последующей обработки в зашифрованном виде.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Ноздронов В.* Об уязвимостях протокола интернета вещей NB-Fi в новом проекте национального стандарта // Ежегодная международная научно-практическая конференция «РусКрипто'2021». – Режим доступа: https://www.ruscrypto.ru/resource/archive/rc2021/files/02_nozdrunov.pdf (дата обращения: 07.05.2021).
2. *Поликарпов А.* Особенности внедрения СКЗИ в РТК с БлЛА МД // Ежегодная международная научно-практическая конференция «РусКрипто'2021». – Режим доступа: https://www.ruscrypto.ru/resource/archive/rc2021/files/11_polikarpov.pdf (дата обращения: 07.05.2021).
3. *Полегенько А.* Способ сопряжения сетей разного уровня «открытости», организованных робототехническими комплексами и системами // Ежегодная международная научно-практическая конференция «РусКрипто'2021». – Режим доступа: https://www.ruscrypto.ru/resource/archive/rc2021/files/11_polegenko.pdf (дата обращения: 07.05.2021).
4. *Жуков А.* Легковесная криптография // Вопросы кибербезопасности. – 2015. – № 1 (9). – С. 26-44.
5. *Дерябин М.А., Кучеров Н.Н.* Обзор безопасных методов шифрования для облачных вычислений // Новости науки в АПК. – 2019. – № 3 (12). – С. 298-303.
6. *Mark A. Will, Ryan K.L.* The Cloud Security Ecosystem Chapter 5 - A guide to homomorphic encryption // Technical, Legal, Business and Management Issues. – 2015. – P. 101-127. – Режим доступа: <https://www.sciencedirect.com/science/article/pii/B9780128015957000057> (дата обращения: 07.05.2021).
7. *Naehrig M., Lauter K.* Can Homomorphic Encryption be Practical? // Proceedings of the 3rd ACM Cloud Computing Security Workshop, CCSW 2011, Chicago, USA, 2011. – P. 113-124.
8. *Halevi S., Shoup V.* Design and implementation of HELib: a homomorphic encryption library. – Режим доступа: <https://eprint.iacr.org/2020/1481> (дата обращения: 07.05.2021).
9. *Гаража А.А., Герасимов И.Ю., Николаев М.В., Чижов И.В.* Об использовании библиотек полностью гомоморфного шифрования // International Journal of Open Information Technologies. – 2021. – Vol. 9, No. 3. – P. 11-22.
10. *Gentry C.* Fully Homomorphic encryption using ideal lattices // Proceedings of 41-th ACM symposium on theory of computing (STOC). Bethesda, 2009. – P. 169-178.
11. *Gentry C., Halevi S., Smart N.P.* Better Bootstrapping in Fully Homomorphic Encryption Gentry // Public Key Cryptography – PKC 2012. Vol 7293 – 2012, Springer. – P. 1-16.
12. *Араkelов Г.Г., Грибов А.В., Михалев А.В.* Прикладная гомоморфная криптография: примеры // Фундаментальная и прикладная математика. – 2016. – Т. 21, № 3. – С. 25-38.
13. *Alam S., De D.* Analysis of Security Threats in Wireless Sensor Network // International Journal of Wireless & Mobile Networks. – 2014. – Vol. 6, No. 2, – P. 1-12.
14. *Borghain T., Sanyal S.* Survey of Operating Systems for the IoT Environment // International Journal of Advanced Networking and Applications. – 2015. – Vol. 6. – P. 1-5.

15. Gentry C., Halevi S., Smart N.P. Homomorphic Evaluation of the AES Circuit // *Advances in Cryptology - CRYPTO 2012*. Vol 7417, Springer. – P. 850-867.
16. Криптографическая защита информации Блочные шифры – ГОСТ Р 34.12-2015. – URL: <https://www.tc26.ru/standard/gost/GOST R 3412-2015.pdf> (дата обращения: 07.05.2021).
17. Бабенко Л.К., Ицуква Е.А., Толочаненко Е.А. Дифференциальный анализ шифра Кузнецик // *Известия ЮФУ. Технические науки.* – 2017. – № 5 (190). – С. 25-37.
18. Ishchukova E.A., Babenko L.K. Two simplified versions of Kuznyechik cipher (GOST R 34.12-2015) // *Proceedings of the 10th International Conference on Security of Information and Networks.* – SIN '17. – New York, NY, USA: ACM, 2017.
19. Brakerski Z., Gentry C., Vaikuntanathan Vinod. Fully homomorphic encryption without bootstrapping. // *Cryptology ePrint Archive, Report 2011/277.* – 2011. – Режим доступа: <https://eprint.iacr.org/2011/277> (дата обращения: 07.05.2021)
20. Babenko L.K., Tolochanenko E.A. Development of algorithms for data transmission in sensor networks based on fully homomorphic encryption using symmetric Kuznyechik algorithm // *Journal of Physics: Conference Series.* – 2021. – Vol. 1812. – P. 246-251.

REFERENCES

1. Nozdrunov V. Ob uyazvimostyakh protokola interneta veshchey NB-Fi v novom proekte natsional'nogo standarta [About vulnerabilities of the Internet of Things protocol NB-Fi in the new draft of the national standard], *Ezhegodnaya mezhdunarodnaya nauchno-prakticheskaya konferentsiya «RusKripto'2021»* [Annual International scientific and practical Conference "RusKripto'2021"]. Available at: https://www.ruscrypto.ru/resource/archive/rc2021/files/02_nozdrunov.pdf (accessed 07 May 2021).
2. Polikarpov A. Osobennosti vnedreniya SKZI v RTK s BpLA MD [Features of the implementation of SKZI in RTC with UAV MD], *Ezhegodnaya mezhdunarodnaya nauchno-prakticheskaya konferentsiya «RusKripto'2021»* [Annual International scientific and practical Conference "RusKripto'2021"]. Available at: https://www.ruscrypto.ru/resource/archive/rc2021/files/11_polikarpov.pdf (accessed 07 May 2021).
3. Polegen'ko A. Sposob sopryazheniya setey raznogo urovnya «otkrytosti», organizovannykh robototekhnicheskimi kompleksami i sistemami [The method of interfacing networks of different levels of "openness", organized by robotic complexes and systems], *Ezhegodnaya mezhdunarodnaya nauchno-prakticheskaya konferentsiya «RusKripto'2021»* [Annual International scientific and practical conference "RusKripto'2021"]. Available at: https://www.ruscrypto.ru/resource/archive/rc2021/files/11_polegenko.pdf (accessed 07 May 2021).
4. Zhukov A. Legkovesnaya kriptografiya [Lightweight Cryptography], *Voprosy kiberbezopasnosti* [Cybersecurity issues], 2015, No. 1 (9), pp. 26-44.
5. Deryabin M.A., Kucherov N.N. Obzor bezopasnykh metodov shifrovaniya dlya oblachnykh vychisleniy [Review of secure encryption methods for cloud computing], *Novosti nauki v APK* [Science news in the agro-industrial complex], 2019, No. 3 (12), pp. 298-303.
6. Mark A. Will, Ryan K.L. The Cloud Security Ecosystem Chapter 5 - A guide to homomorphic encryption, *Technical, Legal, Business and Management Issues*, 2015, pp. 101-127. Available at: <https://www.sciencedirect.com/science/article/pii/B9780128015957000057> (accessed 07 May 2021).
7. Naehrig M., Lauter K. Can Homomorphic Encryption be Practical?, *Proceedings of the 3rd ACM Cloud Computing Security Workshop, CCSW 2011, Chicago, USA, 2011*, pp. 113-124.
8. Halevi S., Shoup V. Design and implementation of HELib: a homomorphic encryption library. Available at: <https://eprint.iacr.org/2020/1481> (accessed 07 May 2021).
9. Garazha A.A., Gerasimov I.Yu., Nikolaev M.V., Chizhov I.V. Ob ispol'zovanii bibliotek polnost'yu gomomorfno shifrovaniya [On the use of fully homomorphic encryption libraries], *International Journal of Open Information Technologies*, 2021, Vol. 9, No. 3, pp. 11-22.
10. Gentry C. Fully Homomorphic encryption using ideal lattices, *Proceedings of 41-th ACM symposium on theory of computing (STOC)*. Bethesda, 2009, pp. 169-178.
11. Gentry C., Halevi S., Smart N.P. Better Bootstrapping in Fully Homomorphic Encryption Gentry, *Public Key Cryptography – PKC 2012*. Vol 7293 – 2012, Springer, pp. 1-16.
12. Arakelov G.G., Gribov A.V., Mikhalev A.V. Prikladnaya gomomorfnyaya kriptografiya: primery [Applied homomorphic cryptography: examples], *Fundamental'naya i prikladnaya matematika* [Fundamental and applied mathematics], 2016, Vol. 21, No. 3, pp. 25-38.

13. Alam S., De D. Analysis of Security Threats in Wireless Sensor Network, *International Journal of Wireless & Mobile Networks*, 2014, Vol. 6, No. 2, pp. 1-12.
14. Borgohain T., Sanyal S. Survey of Operating Systems for the IoT Environment, *International Journal of Advanced Networking and Applications*, 2015, Vol. 6, pp. 1-5.
15. Gentry C., Halevi S., Smart N.P. Homomorphic Evaluation of the AES Circuit, *Advances in Cryptology - CRYPTO 2012. Vol 7417, Springer*, pp. 850-867.
16. Kriptograficheskaya zashchita informatsii Blochnye shifry – GOST R 34.12-2015 [Cryptographic protection of information Block ciphers-GOST R 34.12-2015]. Available at: <https://www.tc26.ru/standard/gost/GOST R 3412-2015.pdf> (accessed 07 May 2021).
17. Babenko L.K., Ishchukova E.A., Tolomanenko E.A. Differential'nyy analiz shifra Kuznechik [Differential analysis of the Grasshopper cipher], *Izvestiya YuFU. Tekhnicheskie nauki [Izvestiya SFedU. Engineering Sciences]*, 2017, No. 5 (190), pp. 25-37.
18. Ishchukova E.A., Babenko L.K. Two simplified versions of Kuznyechik cipher (GOST R 34.12-2015), *Proceedings of the 10th International Conference on Security of Information and Networks. – SIN '17. New York, NY, USA: ACM*, 2017.
19. Brakerski Z., Gentry C., Vaikuntanathan Vinod. Fully homomorphic encryption without bootstrapping, *Cryptology ePrint Archive, Report 2011/277*, 2011. Available at: <https://eprint.iacr.org/2011/277> (accessed 07 May 2021)
20. Babenko L.K., Tolomanenko E.A. Development of algorithms for data transmission in sensor networks based on fully homomorphic encryption using symmetric Kuznyechik algorithm, *Journal of Physics: Conference Series*, 2021, Vol. 1812, pp. 246-251.

Статью рекомендовала к опубликованию к.т.н. Е.А. Ищукова.

Бабенко Людмила Климентьевна – Южный федеральный университет; e-mail: lkbabenko@sfedu.ru; г. Таганрог, Россия; тел.: 88634312018; кафедра безопасности информационных технологий; профессор.

Толоманенко Екатерина Алексеевна – e-mail: kat.tea@mail.ru; тел.: 88634371905; кафедра безопасности информационных технологий; аспирант.

Babenko Lyudmila Klimentevna – Southern Federal University; e-mail: lkbabenko@sfedu.ru; Taganrog, Russia; phone: +78634312018; the department of security of information technologies; professor.

Tolomanenko Ekaterina Alekseevna – e-mail: kat.tea@mail.ru; phone: +78634371905; the department of security of information technologies; postgraduate student.

УДК 612.743, 612.817.2

DOI 10.18522/2311-3103-2021-2-18-31

Н.А. Будко, М.Ю. Медведев, А.Ю. Будко

РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДА ВЕКТОРНОГО АНАЛИЗА ЭМГ ПРЕДПЛЕЧЬЯ ДЛЯ ПОСТРОЕНИЯ ЧЕЛОВЕКО-МАШИННЫХ ИНТЕРФЕЙСОВ

Рассматриваются проблемы увеличения глубины и повышения долговременной устойчивости каналов связи в интерфейсах человек-машина, построенных на основе данных об электрической активности мышц предплечья. Возможным вариантом решения является применение метода анализа сигналов электромиограмм (ЭМГ), совмещающий векторное и командное управление. В виду возможности случайного смещения положения электродов в процессе эксплуатации, построена математическая модель для векторного анализа ЭМГ в сферических координатах, инвариантная к пространственному расположению электродов на предплечье. Командное управление осуществляется на основе распознавания жестов посредством предварительно обученной искусственной нейронной сети (ИНС). Векторное управление заключается в решении задачи калибровки каналов датчиков ЭМГ по пространственному расположению электродов и расчета результирующего вектора мы-

шечных усилий, используемого в качестве дополнительного информационного канала для задания направления движения рабочей точки объекта управления. Предложенный метод апробирован на реально регистрируемых сигналах ЭМГ. Исследовано влияние длительности обрабатываемых фрагментов сигнала на процесс извлечения информации о вращательном движении кисти. Поскольку изменение положения электродов между сеансами эксплуатации различно, представлен алгоритм переназначения и калибровки усиления каналов ЭМГ, позволяющий в дальнейшем использовать единожды обученную ИНС для распознавания и классификации жестов. Практическое применение результатов работы возможно при разработке алгоритмов калибровки, распознавания жестов и управления техническими объектами на основе электромиографических интерфейсов человек-машина.

Интерфейс человек-машина; векторно-командное управление; электромиограмма.

N.A. Budko, M.Yu. Medvedev, A.Yu. Budko

DEVELOPMENT AND RESEARCH OF THE METHOD OF VECTOR ANALYSIS OF EMG OF THE FOREARM FOR CONSTRUCTION OF HUMAN-MACHINE INTERFACES

The paper deals with the problems of increasing the depth and increasing the long-term stability of communication channels in human-machine interfaces, built on the basis of data on the electrical activity of the forearm muscles. A possible solution is to use the method of analysis of electromyogram (EMG) signals, which combines vector and command control. In view of the possibility of random displacement of the position of the electrodes during operation, a mathematical model was built for vector analysis of EMG in spherical coordinates, which is invariant to the spatial arrangement of the electrodes on the forearm. Command control is based on gesture recognition by means of a pretrained artificial neural network (ANN). Vector control consists in solving the problem of calibrating the channels of EMG sensors according to the spatial arrangement of the electrodes and calculating the resulting vector of muscle forces used as an additional information channel to set the direction of movement of the operating point of the control object. The proposed method has been tested on actually recorded EMG signals. The influence of the duration of the processed signal fragments on the process of extracting information about the rotational movement of the hand was investigated. Since the change in the position of the electrodes between operating sessions is different, an algorithm for reassigning and calibrating the amplification of the EMG channels is presented, which makes it possible to use a once trained ANN for recognition and classification of gestures in the future. Practical application of the results of the work is possible in the development of algorithms for calibration, gesture recognition and control of technical objects based on electromyographic human-machine interfaces.

Human-machine interface; vector-command control; electromyogram.

Введение. Задача управления мехатронными и роботизированными системами особенно актуальна для разработки интерфейсов человек-машина для людей с ограниченными возможностями [1–7]. Численность инвалидов на ноябрь 2020 года составляет более 11 млн. человек, что соответствует 8 % от общего населения России. Более 200 тысяч нуждаются в протезировании рук или ног [8]. Одним из возможных примеров является управление роботом-ассистентом, имеющим в качестве исполнительного механизма, манипулятор, закрепленный на инвалидном кресле или другом устройстве.

Одной из главных проблем бионического управления является нестационарность сигналов ЭМГ и изменение положения электродов относительно мышц в процессе эксплуатации [9, 10]. Точность классификации значительно изменяется во времени, поскольку данные, записанные в один и тот же день, имеют характеристики, отличные от данных, записанных в другой день, из-за влияния реальных условий, таких как смещение датчиков, изменение состояния контакта кожа-электрод, изменение паттернов мышечной активности в процессе выработки стереотипа движения. То есть в настоящий момент одной из ключевых проблем явля-

ется не лабораторные, краткосрочные условия, а ежедневное использование [11, 12]. Ежедневное смещение электродов в процессе эксплуатации может привести к различиям в свойствах сигнала, что делает их нераспознаваемыми для предварительно обученного классификатора. Таким образом, возникает задача оценки пространственного положения электродов по данным ЭМГ для переназначения каналов, перед подачей информации на вход классификатора жестов. В данной работе предлагается производить процедуру калибровки на основе направления максимального вектора мышечных усилий во время кругового движения кисти, поскольку положение такого вектора обусловлено анатомическими особенностями строения предплечья и инвариантно к положению электродов.

Другой проблемой классического подхода к анализу ЭМГ, основанном на классификации и распознавании жестов, является ограничение по количеству степеней свободы для одновременного перемещения рабочей точки [13–16]. Например, система управления положением рабочей точки на основе классификатора жестов может быть основана на 4 жестах, кисть вверх, вниз, влево и вправо. Тогда, каждому жесту кисти будет соответствовать перемещение вдоль осей координат, таким образом, набор команд будет состоять из 4 направлений. Для наглядности рассмотрим пример на основе задачи перемещения рабочей точки (положения манипулятора или курсора) из точки А в точку В, рис. 1.

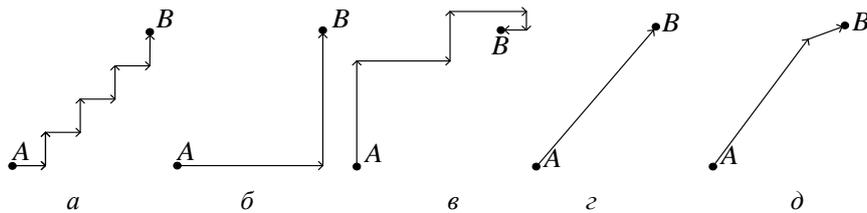


Рис. 1. а – е. Примеры решения задачи перемещения рабочей точки

В прямоугольной системе координат кратчайшим путем будет являться траектория, приведенная на рис. 1,а. Однако, такая многоступенчатая стратегия управления не свойственна человеку, и с большой вероятностью при ограниченном выборе направления перемещения рабочей точки он будет решать задачу перемещения по неоптимальной траектории, пример которых приведен на рис. 1,б,в. Это приводит к увеличению времени достижения результата и отрицательно влияет на его точность. Для человека более естественно перемещение к целевой точке по вектору с произвольным направлением, а не по последовательности движений, направленных вдоль осей прямоугольной системы, рис. 1,г,д. При этом появляется дополнительная возможность сокращения времени достижения цели за счет изменения скорости перемещения рабочей точки пропорционально мышечному усилию.

Таким образом, задача разработки системы управления на основе ЭМГ, позволяющей задавать произвольное направление вектора перемещения рабочего органа, является актуальной, поскольку позволяет повысить удобство пользования, уменьшить время выполнения конечного действия, и повысить точность позиционирования. Заметным исследованием в данном вопросе является двумерный указатель, управляемый ЭМГ, изобретенный Розенбергом (1998), который известен как указатель биологической обратной связи [17, 18]. Основная проблема метода, отмеченная авторами исследования, заключается в том, что движения пользователя могут не адекватно синхронизироваться с курсором [19]. Также следует отметить необходимость точного позиционирования электродов в данном методе.

В исследовании Таваколи и соавт. описана классификация жестов руки по двум каналам ЭМГ сигнала, для классификации предлагается метод опорных векторов (МОВ) [20]. Преимуществом данного метода является высокая скорость калибровки и точность классификации, а также устойчивость при работе в многомерном пространстве признаков. Также алгоритм МОВ показывает релевантные результаты при наличии маленькой выборки для обучения. Результатом является способность классификатора достигать точности от 95 % до 100 % для пяти выходных классов.

В данной работе представлены результаты анализа возможности осуществления векторного управления положением рабочей точки в произвольном двумерном пространстве сферических или декартовых координат на основе прямого расчета результирующего вектора мышечных усилий на основе пространственного положения электродов, а также приведены некоторые аспекты использования данной информации для калибровки системы датчиков.

Методология. Для съема ЭМГ мышц предплечья в работе используется браслет MYO Thalmic Labs, рис. 2,а. Браслет MYO состоит из 8 секций, который при одевании на руку располагаются равномерно по окружности предплечья. Рассмотрим схему расположения датчиков на предплечье, рис. 2,б.

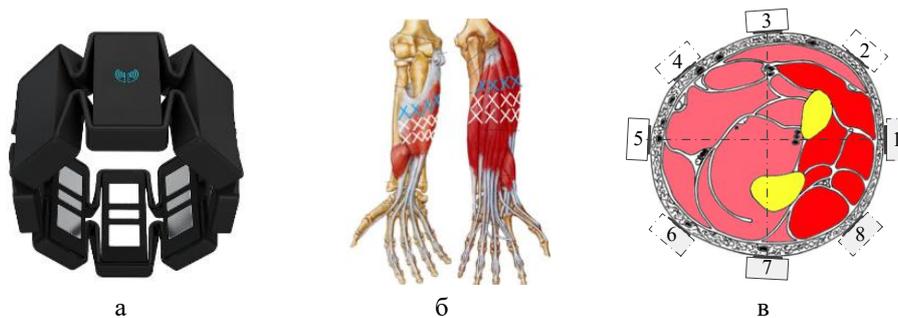


Рис. 2. Браслет MYO и схема расположения датчиков на предплечье: а – внешний вид браслета; б – схематичное изображение мышц предплечья и расположения датчиков; в – пример взаимоположения датчиков и мышц на срезе предплечья, красным цветом выделены условно активированные мышцы при поднятии кисти на себя

Как видно из рис. 2,в, точное расположение браслета над одними и теми же мышцами может быть затруднительно в силу сложного строения предплечья, поэтому представляется целесообразным ориентироваться на группы мышц, задействованные при направлении кисти в разные стороны. Например, при движении кисти вверх и удержании этого жеста активируются мышцы на внешней стороне предплечья. Это справедливо и для других направлений, таких как вниз («кисть на себя»), вправо, влево и промежуточных направлений.

Предварительная оценка данных. Получены записи сигналов для анализа исходных данных. Для анализа выбраны направления движения под углом 90° друг к другу. Этим позициям соответствуют следующие направления руки: от себя («Вверх»), к себе («Вниз»), влево и вправо. На рис. 3 представлены сигналы ЭМГ для исследованных жестов во временной области.

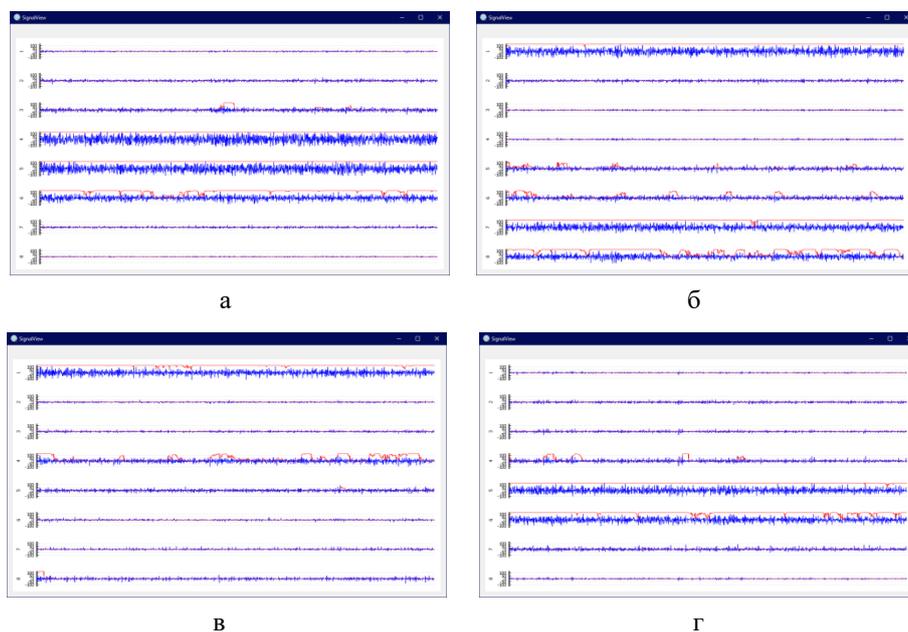


Рис. 3. Сигналы ЭМГ для различных жестов: а – кисть на себя («Вверх»), б – кисть от себя («Вниз»), в – кисть влево, г – кисть вправо

На рис. 4 приведены образы исследованных жестов в круговых диаграммах.

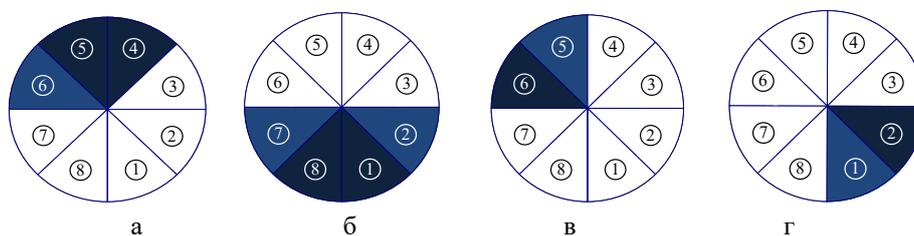


Рис. 4. Образы исследованных жестов в круговых диаграммах: а – кисть на себя («Вверх»), б – кисть от себя («Вниз»), в – кисть влево, г – кисть вправо

Насыщенность цвета сектора на рис. 4 определяется усредненной амплитудой сигнала соответствующего канала. Анализ данных показывает, что в целом направления результирующих векторов совпадают с направлением усилия оператора, однако имеют несимметричную картину.

Расчет результирующего вектора мышечных усилий. Поскольку браслет МУО имеет восемь датчиков ЭМГ, расположенных равномерно по окружности предплечья, можно представить выходные данные браслета в виде векторной диаграммы из 8 компланарных векторов, берущих начало в одной точке и расположенных под углом $360/8=45$ градусов друг к другу, как это показано на рис. 5,а. На рис. 5,б показан пример мгновенно регистрируемой векторной диаграммы при жесте «кисть вверх».

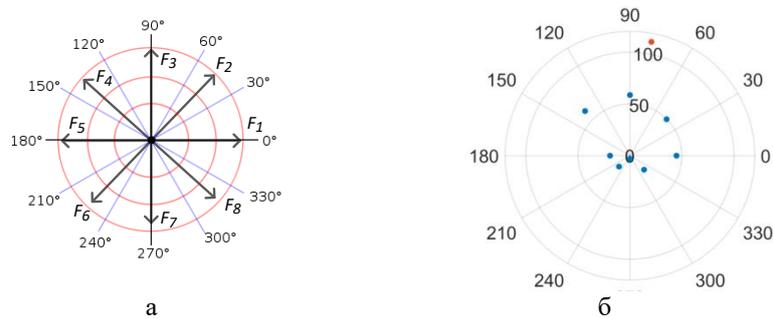


Рис. 5. Представление выходных сигналов в виде векторной диаграммы: а – схема построения векторов, б – пример расчета результирующего вектора для реально регистрируемых сигналов: результирующий вектор отмечен красным маркером

Для удобства выбрана полярная система координат. Начало координаты φ выбирается произвольной привязкой к любому каналу, поскольку в дальнейшем оно будет переопределено во время процедуры калибровки исходя из направления максимального вектора мышечных усилий для кругового движения. Тогда мгновенный вектор сигнала каждого для канала на полярной плоскости координат будет определяться его угловой координатой φ_i , являющейся константой для каждого канала, а также его величиной F_i , равной мгновенной амплитуде сигнала.

Рассмотрим процесс расчета результирующего вектора для всех 8 каналов. Для упрощения дальнейшего объяснения анализа векторной диаграммы проведем аналогию с системой сил, действующих на твердое тело, используемой, например, в теоретической механике. Тогда для сходящейся системы сил, представленной на рис. 5 результирующим вектором, будет являться вектор результирующей силы, воздействующей на тело в точке приложения, соответствующей началу радиальных координат векторов. Результирующая сила R будет равна сумме сил, определяемых векторами F_1 - F_8 (1):

$$\vec{R} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \vec{F}_4 + \vec{F}_5 + \vec{F}_6 + \vec{F}_7 + \vec{F}_8. \quad (1)$$

Для расчета R необходимо последовательно сложить все 8 векторов F_1 - F_8 . Результирующий вектор сложения двух компланарных векторов может быть вычислен тригонометрическим способом с помощью теоремы косинусов (2):

$$R_1 = \sqrt{F_1^2 + F_2^2 + 2 F_1 F_2 \cos(180 - \alpha)}, \quad (2)$$

где α = угол между исходными векторами.

Угол между результирующим вектором и одним из исходных векторов может быть вычислен по теореме синусов (3):

$$\varphi_{R_1} = \arcsin \left[F_2 \sin \left(\frac{180^\circ - \alpha}{R_1} \right) \right], \quad (3)$$

где α = угол между исходными векторами.

Для уравнения (3) на каждом шаге сложения векторов проверяется условие недопустимости деления на 0.

Анализ вращательного движения. Для оценки возможности управления положением рабочей точки на основе результирующего вектора мышечных усилий был произведен анализ ЭМГ при вращательном движении кисти. В процессе эксперимента записан сигнал при совершении кистью правой руки 5 вращательных движений по часовой стрелке с периодом около 2 секунд на один полный оборот (0,5 Гц). Длительность записи составила 13 секунд, что при частоте дис-

кретизации 200 Гц составляет 2600 отчетов. Рассмотрим поля расчетных координат результирующего вектора для полученной выборки. Поскольку мышцы предплечья, отвечающие за движение кисти в различных направлениях, сгруппированы и развиты неравномерно, то картина выборки без калибровки является несимметричной, как это показано на рис. 6,а. Для повышения удобства пользователя, на данном этапе необходимо произвести калибровку коэффициентов усиления каждого канала, имеющую своей целью получить максимально симметричную картину. Координаты вектора после процедуры калибровки усиления по каналам приведены на рис. 6,б.

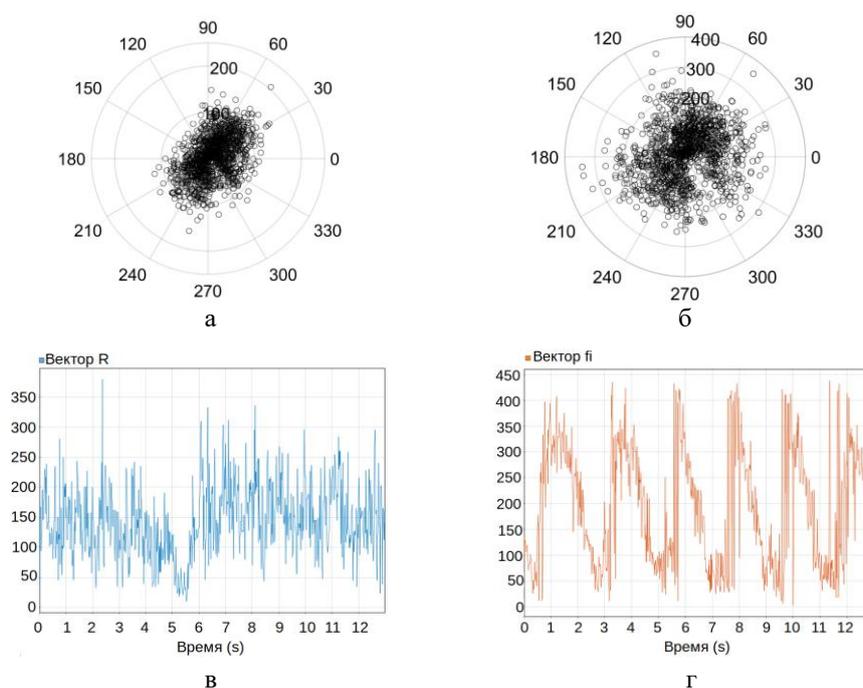


Рис. 6. Координаты расчетного вектора для полученной выборки: а – без калибровки усиления по каналам, б – с калибровкой усиления по каналам, в – график зависимости радиуса R от времени, г – график зависимости угла φ от времени

В данном исследовании коэффициенты усиления рассчитаны по усредненным значениям амплитуд всего вектора пропорционально по отношению к усредненному значению произвольно выбранного канала, принятого за опорный. Следует отметить, что задача настройки коэффициентов усиления сигнала по каналам для обеспечения максимальной симметричности векторной диаграммы, является одной из основных для получения качественного процесса управления.

Для выделения полезной части сигнала необходимо проведение предварительной фильтрации и обработки. После записи экспериментальной выборки необходимо произвести разбиение сигнала на фрагменты (фреймы) конечной длительности и выделение информативных признаков. Предварительная обработка для вычисления результирующего вектора мышечных усилий осуществляется на основе усреднения по модулю всех значений сигнала в пределах одного фрейма. Исследован диапазон длительностей фреймов от 10 мс до 1,5 с, что соответствует усреднению по 2–300 отчетам при частоте дискретизации 200 Гц. Шаг изменения

длительности фрейма составил 25 мс или 5 отчетов. На рис. 7 представлены примеры полученных расчетных координат результирующего вектора для различных длительностей фреймов.

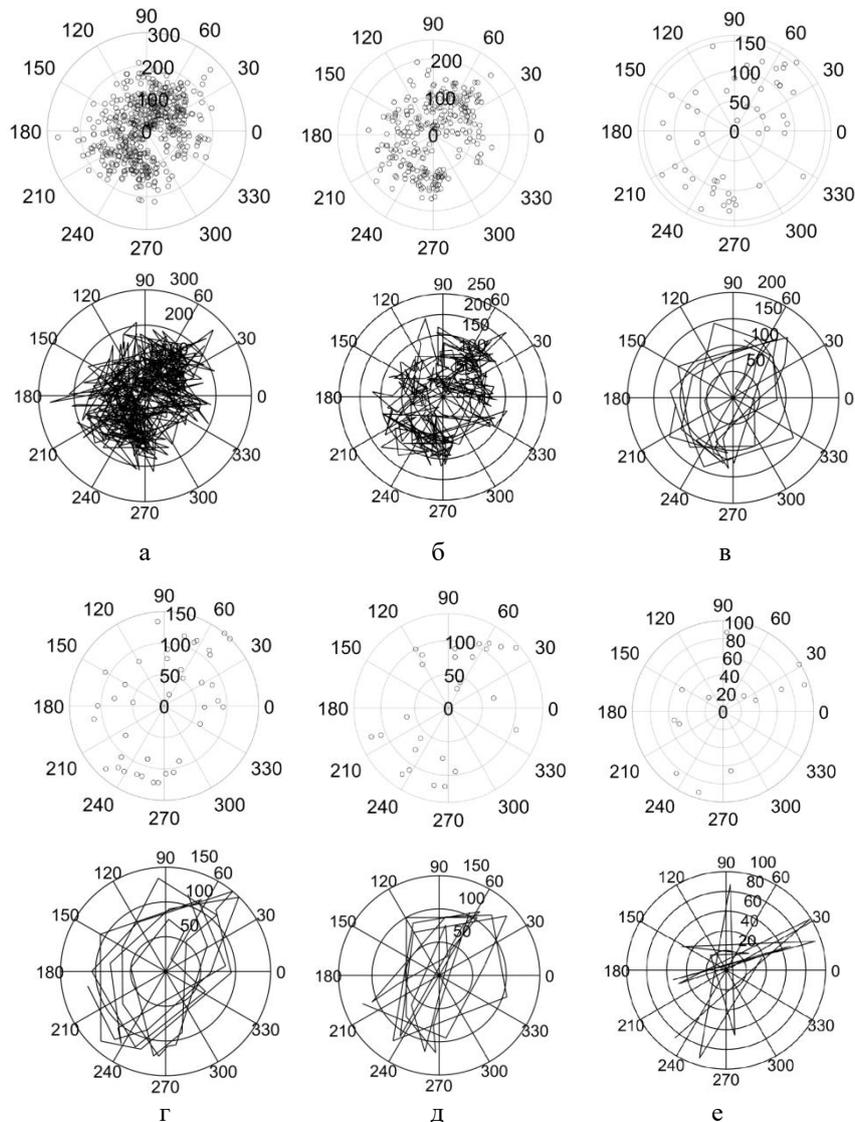


Рис. 7. Расчетные координаты результирующего вектора для пяти вращательных движений при различных длительностях фреймов: а – 25мс, б – 50мс, в – 250мс, г – 350, д – 1 сек, е – 1,5 сек

Анализ данных показал, что при использовании длительности фрейма более 400мс информация о характере действия (вращение кисти) теряется. При этом вращение кисти с большей частотой, чем в данном эксперименте, может дополнительно ограничить максимальную длину фрейма для усреднения. Для работы системы с большим быстродействием целесообразно использовать фреймы сигнала меньшей длины, что, однако ведет к зашумлению выходных координат. При построении подобной системы необходимо находить компромисс между быстродей-

ствием системы и плавностью изменения координат в зависимости от целей и особенностей объекта управления. В данной работе длительность фрейма выбрана 250 мс. Графики движения рабочей точки в полярной системе координат приведены на рис. 8.

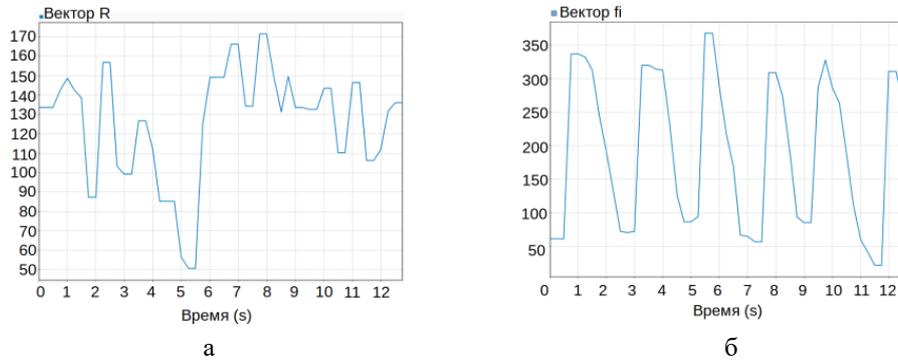


Рис. 8. Графики перемещения рабочей точки в полярной системе координат и значений координат во временной области: а – графики вектора R, б – угол φ в зависимости от времени в полярной системе координат

На рис. 8,б для наглядности траектории движения точки координата радиуса вектора посчитана с накоплением от начала к концу движения. Если это необходимо, то переход в декартову систему координат может быть осуществлен по формулам (5):

$$X_1 = F_1 \sin(\varphi_{12}), Y_1 = F_1 \cos(\varphi_{12}). \quad (4)$$

Для улучшения качества управления целесообразно произвести аппроксимацию и сглаживание выходных координат результирующего вектора. На рис. 9 показана траектория движения рабочей точки в декартовой системе координат, при этом координаты вектора были сглажены сплайнами.

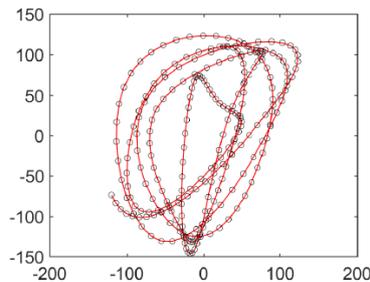


Рис. 9. Сглаженная траектория движения рабочей точки в декартовой системе координат

Анализ траектории движения рабочей точки на рис. 9 показывает, что, несмотря на введенные коэффициенты усиления, и процедуру сглаживания график остается несимметричным относительно начала отчета. Это указывает на анатомические особенности руки человека, в которой в определенном направлении может быть больше крупных и сильных мышц, чем в другом. Дальнейшее улучшение результата возможно с применением ИНС и систем с подкреплением для более

тонкой настройки коэффициентов усиления, применением алгоритмов обработки для векторов входных и выходных величин и т. д. При этом, информация об анатомических особенностях, получаемая на основе анализа подобного рода может быть использована для определения пространственного положения датчиков в процессе эксплуатации.

Алгоритм калибровки каналов на основе векторного анализа набора данных ЭМГ предплечья. На основе полученных данных предложен алгоритм калибровки при ежедневном использовании интерфейса, рис. 10.

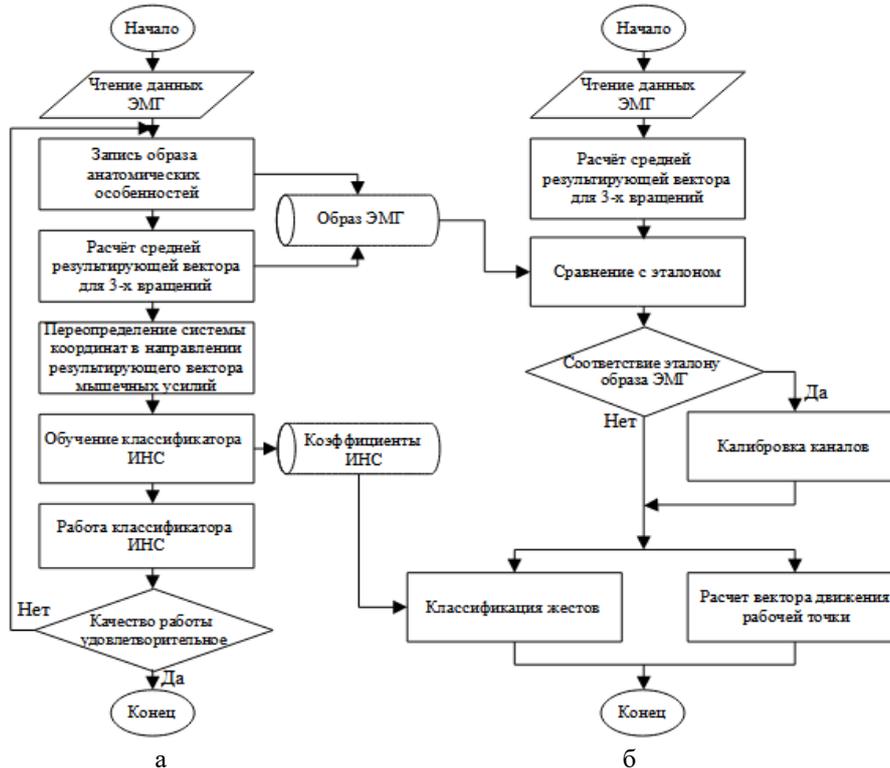


Рис. 10. а – алгоритм работы при первоначальной настройке, б – ежедневное использование

На рис. 10,а показан алгоритм работы при первоначальной настройке интерфейса. Производится чтение и запись сигналов ЭМГ, рассчитывается средний результирующий вектор для трёх вращений и сохраняется в памяти. Далее производится переопределение системы координат в направлении результирующего вектора мышечных усилий и осуществляется привязка каналов. Далее производится обучение классификатора жестов на основе ИНС и его тестирование, коэффициенты ИНС сохраняются в память. При ежедневном использовании, рис. 10,б, в начале работы в целях калибровки рассчитывается средний результирующий вектор для трёх вращений и сравнивается с эталоном, полученным при первичной настройке. При несовпадении данных сначала производится калибровка каналов, включающая переопределение номеров каналов и корректировку коэффициентов усиления. Процедура позволяет привести данные по каналам ЭМГ к эталонной конфигурации при обучении и использовать ранее полученные веса ИНС. Алгоритм калибровки более подробно показан на рис. 11.

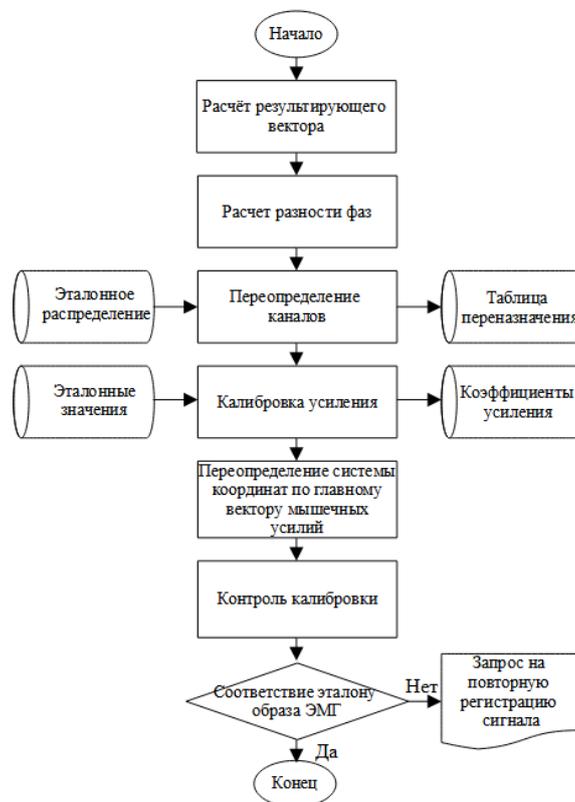


Рис. 3. Алгоритм калибровки

Для калибровки каналов на основе векторного анализа набора данных ЭМГ предплечья необходимо рассчитать результирующий вектор и разность фаз. Далее производится переопределение каналов и калибровка усиления с учётом эталонных значений. Полученные данные сохраняются в таблицах переназначения и коэффициентов усиления. Далее производится переопределение системы координат по вектору мышечных усилий и контроль калибровки. Если наблюдается соответствие эталону образа ЭМГ, то программа калибровки завершается, если нет, то осуществляется повторная регистрация сигнала и повторение процедуры калибровки.

Заключение. В работе предложен метод обработки ЭМГ на основе векторной диаграммы и усреднения по фреймам, позволяющий извлекать полезную информацию о направлении результирующего вектора мышечного усилия и анатомических особенностях предплечья. Представлены результаты экспериментального исследования применимости предложенного метода для случая вращательного движения кисти. Приведены качественные и количественные оценки положения кисти по данным ЭМГ. Предложен алгоритм калибровки датчиков, построенный на основе разработанного метода. Отличительными особенностями предложенного метода является инвариантность к положению электродов на предплечье и возможность извлечения дополнительной информации о положении кисти, которая может быть использована для управления манипулятором или курсором на экране компьютера при разработке перспективных человеко-машинных интерфейсов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Гурфинкель В.С.* Биоэлектрическое управление. – М.: Наука, 1972. – 243 с.
2. *Будко Н.А., Будко Р. Ю., Будко А.Ю.* Применение искусственных нейронных сетей в интерфейсах человек–машина // Моделирование, оптимизация и информационные технологии. – 2019. – Т. 7, № 1. – <http://moit.vivt.ru/>.
3. *Будко Р.Ю., Чернов Н.Н., Будко Н.А., Будко А.Ю.* Распознавание электромиограммы предплечья и выбор жестов для управления протезом // Моделирование, оптимизация и информационные технологии. – 2019. – Т. 7, № 1. – <http://moit.vivt.ru/>.
4. *Будко Р.Ю., Чернов Н.Н., Будко Н.А.* Исследование методов классификации ЭМГ в задаче управления жестами руки // Известия высших учебных заведений. Приборостроение. – 2019. – Doi 10.17586/0021-3454-2019-62-12-1098-1104.
5. *Будко Р.Ю., Чернов Н.Н., Будко Н.А.* Метод управления устройствами замещения утраченных функций на основе мио сигнала и его верификация в реальном масштабе времени // Вестник молодежной науки России. – 2019. – Вып. № 6.
6. *Славуцкий Я.Л.* Физиологические аспекты биоэлектрического управления протезами. – М.: Медицина, 1982. – 288 с.
7. *Буров Г.Н.* Современное состояние проблемы управления протезами верхних конечностей и достигаемый эффект протезирования // Вестник гильдии протезистов-ортопедов. – 2002. – № 4 (10). – С. 37-43.
8. Федеральная государственная информационная система. Федеральный реестр инвалидов. – Режим доступа: <https://sfri.ru/analitika/chislennost> (дата обращения: 01.12.20).
9. *Персон Р.С.* Электромиография в исследованиях человека. – М.: Наука, 1969. – 231 с.
10. *Resnik L.* Development and testing of new upper-limb prosthetic devices: research designs for usability testing // J. Rehabil. Res. Dev. – 2011. – Vol. 48, No. 6. – P. 697-706.
11. *Фарбер Б.С., Витензон А.С.* Теоретические основы построения протезов нижних конечностей и коррекции движения. – М.: ЦНИИПП, 1994. – 645 с.
12. *Vigreux B., Snockart J.C., Pertuzon E.* Factors influenced quantified surface EMG // Journal Applied Physiology. – 1979. – Vol. 41, No. 4. – P. 119-129.
13. *Ogiri Y.* Development of an upper limb neuroprosthesis to voluntarily control elbow and hand // Proc. 26th IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN). Lisbon, Portugal, 2017. – P. 298-303. – Doi: 10.1109/ROMAN.2017.8172317.
14. *Юсевич Ю.С.* Очерки по клинической электромиографии. – М.: Медицина, 1972. – 187 с.
15. *Коуэн Х., Брумлик Дж.* Руководство по электромиографии и электродиагностике. – М.: Медицина, 1975. – 358 с.
16. *Кондрашин Н.И., Санин В.Г.* Ампутация конечностей и первичное протезирование. – М.: Медицина, 1984. – 160 с.
17. *Ивановский Ю.В.* Принципы использования метода биологической обратной связи в системе медицинской реабилитации // Биологическая обратная связь. – 2000. – № 3. – С. 2-9.
18. *Кунельская Н.Л.* Метод биологической обратной связи в клинической практике // Журнал неврологии и психиатрии им. С.С. Корсакова. – 2014. – № 114 (8). – С. 46-50.
19. *Rosenberg, R.* The biofeedback pointer: EMG control of a two-dimensional pointer, Wearable Computers // Digest of Papers. Second International Symposium on 19-20 Oct. 1998. – P. 162-163.
20. *Tavakoli M.* Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier // Biomedical Signal Processing and Control. – 2018. – No. 46. – P. 121-130. – Doi:10.1016/j.bspc.2018.07.010.

REFERENCES

1. *Gurfinkel' V.S.* Bioelektricheskoe upravlenie [Bioelectrical control]. Moscow: Nauka, 1972, 243 p.
2. *Budko N.A., Budko R. Yu., Budko A.Yu.* Primenenie iskusstvennykh neyronnykh setey v interfeysakh chelovek–mashina [Application of artificial neural networks in human-machine interfaces], *Modelirovanie, optimizatsiya i informatsionnye tekhnologii* [Modeling, Optimization and Information Technologies], 2019, Vol. 7, No. 1. Available at: <http://moit.vivt.ru/>.
3. *Budko R.Yu., Chernov N.N., Budko N.A., Budko A.Yu.* Raspoznavanie elektromiogrammy predplech'ya i vybor zhestov dlya upravleniya protezom [Recognition of the forearm electromyogram and selection of gestures for controlling the prosthesis], *Modelirovanie, optimizatsiya i informatsionnye tekhnologii* [Modeling, optimization, and information technology], 2019, Vol. 7, No. 1. Available at: <http://moit.vivt.ru/>.

4. Budko R.Yu., Chernov N.N., Budko N.A. Issledovanie metodov klassifikatsii EMG v zadache upravleniya zhestami ruki [Research of methods of classification of EMG in the control of hand gestures], *Izvestiya vysshikh uchebnykh zavedeniy. Priborostroenie* [News of higher educational institutions. Instrumentation], 2019. Doi 10.17586/0021-3454-2019-62-12-1098-1104.
5. Budko R.Yu., Chernov N.N., Budko N.A. Metod upravleniya ustroystvami zameshcheniya utraennykh funktsiy na osnove miosignala i ego verifikatsiya v real'nom masshtabe vremeni [A method for controlling devices for replacing lost functions based on the miosignal and its verification in real time], *Vestnik molodezhnoy nauki Rossii* [Bulletin of Youth Science of Russia], 2019, Issue No. 6.
6. Slavutskiy Ya.L. Fiziologicheskie aspekty bioelektricheskogo upravleniya protezami [Physiological aspects of bioelectric control of prostheses]. Moscow: Meditsina, 1982, 288 p.
7. Burov G.N. Sovremennoe sostoyanie problemy upravleniya protezami verkhnikh konechnostey i dostigaemyy effekt protezirovaniya [Current state of the problem of management of upper limb prostheses and the achieved effect of prosthetics], *Vestnik gil'dii protezistov-ortopedov* [Bulletin of the Guild of Prosthetics and Orthopedists], 2002, No. 4 (10), pp. 37-43.
8. Federal'naya gosudarstvennaya informatsionnaya sistema. Federal'nyy reestr invalidov [Federal State Information System. Federal Register of Disabled Persons]. Available at: <https://sfri.ru/analitika/chislennost> (accessed 01 December 20).
9. Person R.S. Elektromiografiya v issledovaniyakh cheloveka [Electromyography in human research]. Moscow: Nauka, 1969, 231 p.
10. Resnik L. Development and testing of new upper-limb prosthetic devices: research designs for usability testing, *J. Rehabil. Res. Dev.*, 2011, Vol. 48, No. 6, pp. 697-706.
11. Farber B.S., Vitenzon A.S. Teoreticheskie osnovy postroeniya protezov nizhnikh konechnostey i korrektsii dvizheniya [Theoretical foundations of the construction of lower limb prostheses and motion correction]. Moscow: TSNIIPP, 1994, 645 p.
12. Vigreux B., Cnockart J.C., Pertuzon E. Factors influenced quantified surface EMG, *Journal Applied Physiology*, 1979, Vol. 41, No. 4, pp. 119-129.
13. Ogiri Y. Development of an upper limb neuroprosthesis to voluntarily control elbow and hand, *Proc. 26th IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN). Lisbon, Portugal, 2017*, pp. 298-303. Doi: 10.1109/ROMAN.2017.8172317.
14. Yusevich Yu.S. Ocherki po klinicheskoy elektromiografii [Essays on clinical electromyography]. Moscow: Meditsina, 1972, 187 p.
15. Kouen Kh., Brumlik Dzh. Rukovodstvo po elektromiografii i elektrodiagnostike [Manual of electromyography and electrodiagnostics]. Moscow: Meditsina, 1975, 358 p.
16. Kondrashin N.I., Sanin V.G. Amputatsiya konechnostey i pervichnoe protezirovanie [Amputation of limbs and primary prosthetics]. Moscow: Meditsina, 1984, 160 p.
17. Ivanovskiy Yu.V. Printsipy ispol'zovaniya metoda biologicheskoy obratnoy svyazi v sisteme meditsinskoj reabilitatsii [Principles of using the method of biological feedback in the system of medical rehabilitation], *Biologicheskaya obratnaya svyaz'* [Biological feedback], 2000, No. 3, pp. 2-9.
18. Kunel'skaya N.L. Metod biologicheskoy obratnoy svyazi v klinicheskoy praktike [Method of biological feedback in clinical practice], *Zhurnal nevrologii i psikiatrii im. S.S. Korsakova* [Journal of Neurology and Psychiatry named after S. S. Korsakov], 2014, No. 114 (8), pp. 46-50.
19. Rosenberg, R. The biofeedback pointer: EMG control of a two-dimensional pointer, *Wearable Computers, Digest of Papers. Second International Symposium on 19-20 Oct. 1998*, pp. 162-163.
20. Tavakoli M. Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier, *Biomedical Signal Processing and Control*, 2018, No. 46, pp. 121-130. Doi:10.1016/j.bspc.2018.07.010.

Статью рекомендовал к опубликованию д.т.н., профессор С.Г. Капустян.

Будко Наталья Александровна – НИИ РИПУ ЮФУ; e-mail: natalia.tb13@mail.ru; г. Таганрог, Россия; тел.: +79185670210; инженер; аспирант кафедры ЭИМ ИРТСУ ЮФУ.

Медведев Михаил Юрьевич – e-mail: medvmihal@gmail.com, тел.: +79198927349; в.н.с.; д.т.н.; профессор; зав. кафедрой ЭИМ.

Будко Артём Юрьевич – e-mail: aptem_budko@mail.com; тел.: +79281545833; с.н.с.; к.т.н.; доцент кафедры ЭИМ.

Budko Natalia Aleksandrovna – RDIRCS SFU; e-mail: natalia.tb13@mail.ru; Taganrog, Russia; phone: +79185670210; engineer; graduate student SFU.

Medvedev Mikhail Yurievich – e-mail: medvmihal@gmail.com; phone: +79198927349; leading researcher; dr. of eng. sc.; professor; head of department SFU.

Budko Artem Yurievich – e-mail: aptem_budko@mail.com; phone: +79281545833; senior research; cand. of eng. sc.; associate professor of the department SFU.

УДК 658.512

DOI 10.18522/2311-3103-2021-2-31-39

В.И. Данильченко, Е.В. Данильченко, В.М. Курейчик

МНОГОМЕРНЫЙ ПОИСК В ЗАДАЧЕ РАЗМЕЩЕНИИ ЭЛЕМЕНТОВ СБИС НА ОСНОВЕ ГЕНЕТИЧЕСКОГО АЛГОРИТМА

Все крупные производители идут к уменьшению габаритов современных микроэлектронных устройств. Это приводит к переходу на новые нормы проектирования и изготовления СБИС. Известные алгоритмы автоматизированного проектирования не в полной мере способны реализовывать новые требования при проектировании СБИС. В связи с этим, при решении задач конструкторского проектирования появляется необходимость в разработке новых методик решения данного класса задач. Одним из таких методик может стать гибридная многомерная система поиска, основанная на генетическом алгоритме (ГА). Описывается автоматизированный подход к проектированию СБИС на основе генетического алгоритма, позволяющий создать алгоритмическую среду в области многомерного генетического поиска для решения NP полных задач, в частности размещения элементов СБИС. Цель данной работы заключается в нахождении путей размещения элементов СБИС на основе генетического алгоритма. Научная новизна заключается в разработке модифицированного многомерного генетического алгоритма для автоматизированного проектирования сверхбольших интегральных схем. Постановка задачи в данной работе заключается в следующем: оптимизировать размещение элементов СБИС путем применения, многомерного модифицированного ГА. Практическая ценность работы заключается в создании подсистемы, позволяющей использовать разработанные многомерные архитектуры, методы и алгоритмы для эффективного решения задач конструкторского проектирования СБИС, а также проводить сравнительный анализ с существующими аналогами. Принципиальное отличие от известных подходов в применении новых многомерных генетических структур в автоматизированном проектировании СБИС, кроме того в работе приведен модифицированный генетический алгоритм. Приведенные результаты вычислительного эксперимента, показали преимущества предложенного в работе многомерного подхода к решению задач размещения элементов СБИС по сравнению с существующими аналогами. Таким образом, проблема создания методов, алгоритмов и программного обеспечения для автоматизированного размещения элементов СБИС в настоящее время имеет особую актуальность. Ее решение позволит улучшить качественные характеристики проектируемых устройств, сократит сроки и затраты на проектирование.

Генетические алгоритмы; графы и гиперграфы; эволюционные вычисления; размещение СБИС; САПР; многомерные вычисления; производства электронных средств.

V.I. Danilchenko, Y.V. Danilchenko, V.M. Kureichik

AUTOMATED STRUCTURAL-PARAMETRIC SYNTHESIS OF A STEPPED DIRECTIONAL RESPONDER ON CONNECTED LINES BASED ON A GENETIC ALGORITHM

All major manufacturers go to a decrease in the dimensions of modern microelectronic devices. This leads to the transition to new standards for designing and manufacturing SBSS. The well-known automated design algorithms are not fully able to implement new requirements when designing a SBI. In this regard, when solving the tasks of design design, there is a need to

develop new methods for solving this class task. One of these techniques can be a hybrid multidimensional search system based on a genetic algorithm (GA). An automated approach to the design of the SBI based on a genetic algorithm is described, which makes it possible to create an algorithmic medium in the field of multidimensional genetic search to solve the NP full tasks, in particular the placement of the VSA elements. The purpose of this work is to find ways to place the elements of the SBI based on the genetic algorithm. The scientific novelty is to develop a modified multidimensional genetic algorithm for automated design of super-high integrated circuits. The formulation of the problem in this paper is as follows: optimize the placement of the ELEMENTS of the SBI by using, multidimensional modified heuristics. The practical value of the work is to create a subsystem that allows you to use the developed multidimensional architecture, methods and algorithms to effectively solve the tasks of the design design of the SDI, as well as conduct a comparative analysis with existing analogues. The fundamental difference from the well-known approaches in the application of new multidimensional genetic structures in the automated design of the SBI, in addition, the modified genetic algorithm was righteous. The results of the computational experiment showed the advantages of a multidimensional approach to solving the tasks of placing the Elements of the SBI compared to existing analogues. Thus, the problem of creating methods, algorithms and software for the automated placement of the SBS elements is currently of particular relevance. Its solution will improve the qualitative characteristics of the designable devices, will reduce the timing and costs of design.

Genetic algorithms; graphs and hypergraphs; evolutionary calculations; SBI; CAD; Multidimensional calculations; electronic means.

Введение. Основным требованием при размещении элементов является создание условий, обеспечивающих полное размещение элементов. Для выполнения данного требования необходимо учитывать характерные свойства конструкции кристалла, оказывающие существенное влияние на результаты размещения. Согласно работе [1], задача размещения элементов в блоках на модели кристалла относится к NP полным задачам [2], это означает, что в общем случае ее решение в принципе не может быть найдено за конечное время ни одним алгоритмом.

Генетический алгоритм. Базовая структура ГА: инициализация первой популяции, генетический оператор и завершение поиска. На рис. 1 приведена обобщенная схема работы алгоритма с оценкой по обратной связи [6].

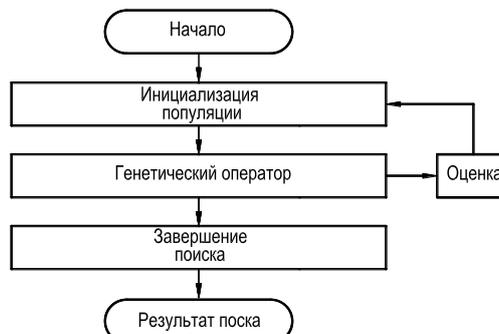


Рис. 1. Базовая структура работы ГА

На рис. 2 приведена модификация базовой структуры генетического поиска на основе использования моделей эволюции Дарвина, Ламарка, де Фриза и Поппера. Здесь, в отличие от базисной структуры, шкала эволюции, взаимодействуя только с внешней средой, вырабатывает сигналы на выбор эволюции Дарвина (0), Ламарка (1), де Фриза (0,5). После этого выполняется модель эволюции Поппера, реализующая один из видов эвристического поиска в виде метода проб и ошибок.

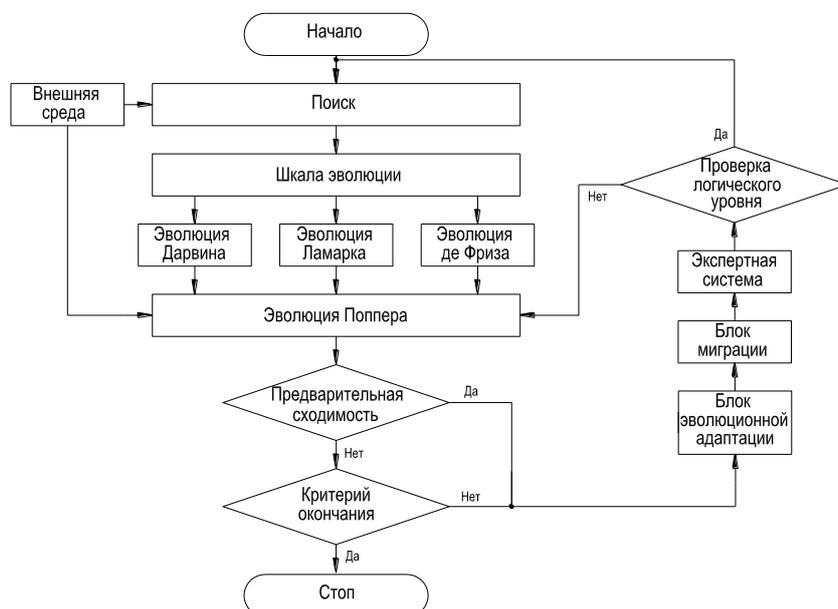


Рис. 2. Модификация базовой структуры

В цепи обратной связи добавлены блоки адаптации и миграции. Они позволяют производить построение порядка из хаоса, установление баланса в системе, выбор параметров для управления эволюционным поиском с целью получения оптимальных и квазиоптимальных решений.

В рассматриваемой задаче методы локального поиска на основе производных обычно не используются, поскольку выражение для производных в аналитическом виде для исследуемых функций получить трудно, а их численное определение с помощью разностных схем может приводить к ощутимой ошибке, в окрестности экстремума. Поэтому предпочтительнее применять методы поиска, основанные на вычислении только самой целевой функции. Информация, полученная на каждом новом шаге поиска, сравнивается с предыдущей, что позволяет сделать процесс поиска экстремума целенаправленным. Например, в одном из таких методов, называемом прямым поиском, сначала дискретно изменяется параметр x_1 при постоянных значениях всех остальных x_k до тех пор, пока целевая функция не станет минимальной. Получение значения $x'_{1\text{опт}}$ фиксируется, и в новом цикле начинается изменение параметра x_2 и т.д. Подобный цикл можно повторять несколько раз. Такой простой алгоритм успешно действует только при слабой зависимости между собой всех параметров x_k в выражении для целевой функции $F_{ц}(x)$. При сильной зависимости между x_k , например при наличии членов типа $x_1 x_2$, стратегии простого поиска может не привести к желаемому результату. Поэтому обычно используют более сложные алгоритмы многомерного поиска минимума целевой функции $F_{ц}(x)$.

В самом общем виде такой алгоритм сводится к следующей итерационной процедуре. В исходной базисной точке $x_0^{(1)}$, за которую можно принять точку с координатами, каждая из которых равна середине зоны изменения соответствующего параметра $x_{ок}^{(1)} = 0,5(x_{k\text{max}} - x_{k\text{min}})$, проводится одномерный поиск: каждая переменная x_i последовательно изменяется на Δx_i при постоянных значениях остальных параметров и вычисляется целевая функция $F_{ц}(x)$. Следовательно, на

первом этапе направления поиска параллельны осям x_1, x_2, \dots, x_n . Фиксируются к какому результату приводит возмущение Δx_i : к оптимальному результату, т.е. к уменьшению целевой функции при минимизации, или попадания в локальный оптимум. После того как пройдены все N направлений (N – число направлений переменных параметров), по каждому параметру вводится новое возмещение: $\alpha \Delta x_i$ в случае успеха и $-\beta \Delta x_i$ в случае неудачи (принимается $\alpha = 3$ и $\beta = 0,5$). Возмущения по всем направлениям задаются до тех пор, пока по каждому из них после предварительного результата не попадает в локальный оптимум. На этом первый этап поиска, состоящий из нескольких циклов последовательного опроса целевой функции по N направлениям, заканчивается.

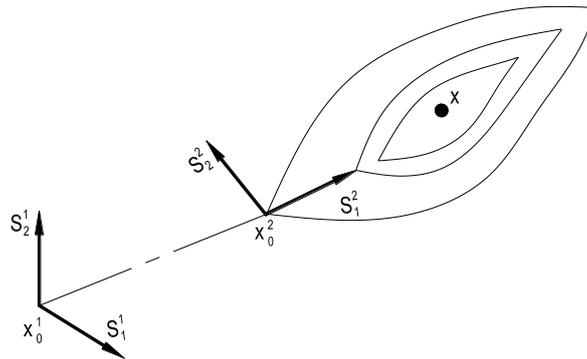


Рис. 3. Построение новых направлений поиска

Полученная в конце 1-го этапа точка становится начальной, базисной для 2-го этапа: $x_0^{(2)} = x_n^{(1)}$. Новое нормированное 1-е направление поиска $S_1^{(2)}$ берется параллельным $(x_0^{(2)} - x_0^{(1)})$, а остальные направления с помощью определенной процедуры выбираются ортонормированными друг к другу и к $S_1^{(2)}$. Построение новых направлений поиска в случае двух переменных показано на рис. 3. Таким образом, новая система координат формируется соответствующим вращением направлений поиска по отношению к предыдущим так, что они оказываются вытянутыми вдоль главных осей квадратурной аппроксимации целевой функции (проход вдоль оврага). Поскольку направления поиска на каждом этапе линейно независимы и ортогональны, то это исключает взаимодействие переменных.

На 2-м этапе поиска по аналогии с 1-м последовательно вводятся возмещения по всем N направлениям. В результате в конце 2-го этапа определяются: новое уменьшенное значение целевой функции, конечная точка 2-го этапа $x_n^{(2)}$, условия для построения новой системы координат для 3-го этапа, поиск в котором начинается с точки $x_0^{(3)} = x_n^{(2)}$. Описанная итерационная процедура, состоящая из K этапов, заканчивается нахождением локального минимума целевой функции или точки, близко к ней расположенной. Для исключения явления «зацикливания» и ограничения времени поиска устанавливаются ограничения как числа вычислений цели на каждом этапе, так и числа K этапов *оптимизации*.

Архитектура гибридного поиска. Одним из основных способов минимизировать сложность задач размещения является сокращение их размерности. На рисунке 4 показана архитектура предлагаемого гибридного подхода на основе двух методов – Розенброка и генетическом поиске [9, 10].

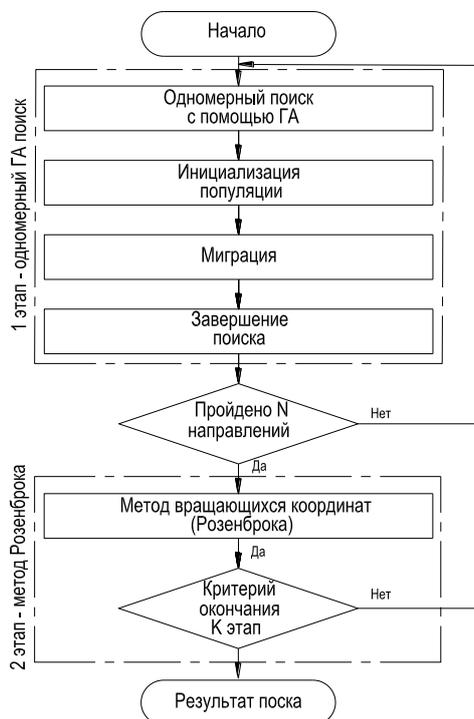


Рис. 4. Архитектура предлагаемого гибридного подхода

Таким образом, отличительной чертой предложенного метода является сведение направлений поиска к новой ортогональной системе в начале каждого этапа минимизации $F_{ц}(x)$ на основании данных, полученных в конце предыдущего этапа. Такой прием вращения направлений поиска на каждой итерации существенно улучшает процесс сходимости, сравнительно быстро приводящий к локальному минимуму или окружающей его окрестности, что подтверждается экспериментально.

Пример. В результате анализа выходных данных, авторами отмечается, что временная сложность разработанного модифицированного алгоритма не выходит за пределы полиномиальной зависимости, и может быть выражена формулой: $O(I\alpha N^2) - O(I\beta N^3)$, где N – число элементов схемы, I – количество итераций (шагов) выполняемых алгоритмом (размер решаемой задачи). Усредненные результаты экспериментов отражены в табл. 1.

Таблица 1

Сравнительная характеристика работы алгоритмов

Алгоритмы	Число элементов схемы									
	100	200	300	400	500	600	700	800	900	1000
Модификация базовой структуры ГА (рис. 2)	0,2	0,71	0,94	1,31	1,53	2,04	2,32	3,33	4,71	10,36
Предлагаемый гибридный подход многомерного ГА (рис. 4)	0,23	0,9	1,13	1,43	1,81	2,33	2,58	3,81	5,28	7,28

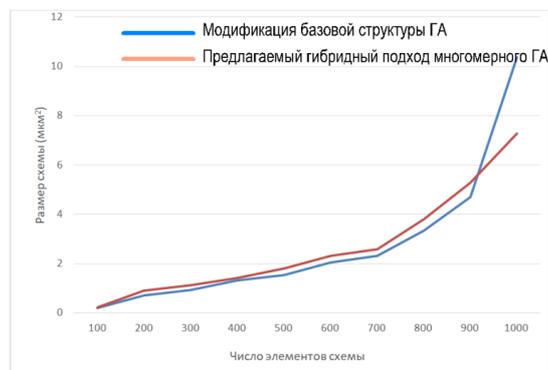


Рис. 5. Сравнительная характеристика работы алгоритмов

Макет из 1000 блоков при проектировании с использованием модифицированной базовой структурой ГА составляет $10,36 \text{ мкм}^2$. Тот же вход при подаче в наш предложенный гибридный подход многомерного ГА возвращается результат с площадью $7,28 \text{ мкм}^2$, что позволяет разместить элементы с минимальным уровнем незадействованного пространства площадки.

При размещении элементов на модели кристалла с помощью гибридного подхода многомерного ГА расположение размещенных элементов имеют оптимальную форму, что дает преимущество на этапе корректировки топологии схемы.

Пример оптимального размещения модифицированным генетическим алгоритмом показан рис. 6.

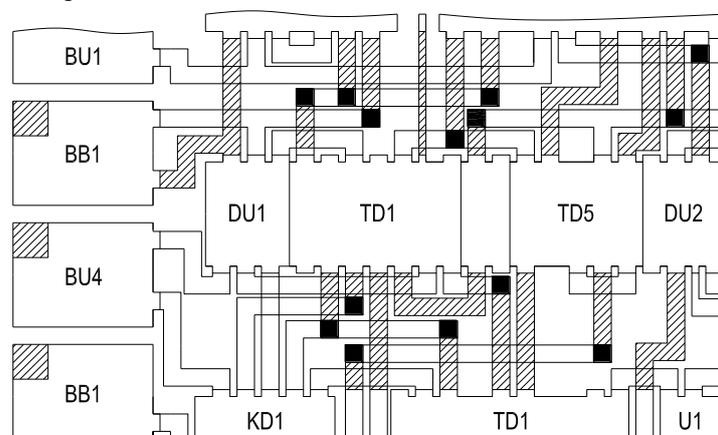


Рис. 6. Оптимальное размещение модифицированным генетическим алгоритмом

При размещении элементов на модели кристалла с помощью модифицированного многомерного генетического алгоритма расположение размещенных элементов имеют оптимальную форму, что дает преимущество на этапе корректировки топологии схемы.

Заключение. В статье описывается автоматизированный подход к проектированию СБИС на основе генетического алгоритма, позволяющий создать алгоритмическую среду в области многомерного генетического поиска для решения NP полных задач, в частности размещения элементов СБИС. Представлена новая архитектура многомерного гибридного подхода на основе ГА, которая дает возможность получить одновременно практически реализуемое схемотехническое

решение с учетом особенностей технологии изготовления и первоначальный вариант размещения элементов СБИС. Описана часть программной реализации методики многомерного поиска. Представленный сравнительный пример работы алгоритмов, подтверждает эффективность предлагаемой архитектуры многомерного гибридного подхода на основе ГА. Проведенные серии экспериментов показали преимущество разработанного алгоритма в среднем на 7 % по сравнению с существующими аналогами.

Поддержка. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-37-90151 и при поддержке Фонда содействия инновациям по договору № 437ГУЦЭС8-Д3/62058 от 05 октября 2020 г.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Тарасов В.Б.* От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. – М.: Эдиториал УРСС, 2002. – 352 с.
2. *Капра Ф.* Паутина жизни. Новое научное понимание живых систем: пер. с англ. – М.: ИД “Гелиос”, 2002.
3. *Курейчик В.М., Данильченко В.И.* Генетический алгоритм планирования размещения СБИС // Известия ЮФУ. Технические науки. – 2019. – № 2 (204). – С. 26-34.
4. *Данильченко В.И., Курейчик В.М.* Классификация и анализ методов решения задачи размещения СБИС // Информатика, вычислительная техника и инженерное образование. – 2018. – Вып. 1.
5. *Danilchenko V.I., Danilchenko Y.V., Kureichik V.M.* Bio-inspired Approach to Microwave Circuit Design // IEEE EAST-WEST DESIGN & TEST SYMPOSIUM. – 2020. – P. 362-366. – DOI: 10.1109/EWDTS 50664.2020.9224737.
6. *Калентьев А.А., Гарайс Д.В., Добуш И.М., Бабак Л.И.* Структурно-параметрический синтез СВЧ транзисторных усилителей на основе генетического алгоритма с использованием моделей монолитных элементов // Доклады ТУСУРа. – Декабрь 2012. – № 2 (26). – Ч. 2. – С. 104-112.
7. *Tang, Maolin and Yao, Xin.* A memetic algorithm for VLSI floorplanning // IEEE Transactions on Systems, Man, And Cybernetics–Part B: Cybernetics. – 2007. – No. 37 (1).
8. *Горяинов А.Е., Добуш И.М., Бабак Л.И.* Построение параметрических моделей пассивных компонентов СВЧ монолитных интегральных схем с использованием программы Extraction-P.
9. *Kokolov A.A., Salmikov A.S., Sheyerman F.I. and Babak L.I.* Broadband Double-Balanced SiGe BiCMOS Mixer With Integrated Asymmetric MBaluns // Int. Conf. “Dynamics of Systems, Mechanisms and Machines” (Dynamics-2017), Omsk, Russia, 2017 (accepted for publication).
10. *Wenyuan L. and Qian Z.* A 0.7–1.9GHz Broadband Pseudo-differential Power Amplifier Using 0.13-um SiGe HBT Technology // 2012 Int. Conf. on Microwave and Millimeter Wave Technology (ICMMT), July 2012. – P. 1-4.
11. *Коколов А.А., Добуш И.М., Шеерман Ф.И., Бабак Л.И. и др.* Сложно-функциональные блоки широкополосных усилителей радиочастоты для однокристалльных приемников L- и S-диапазонов на основе технологии SiGe // 3-я Междунар. науч. конф. «ЭКБ и электронные модули» (Международный форум «Микроэлектроника-2017»), г. Алушта, октябрь 2017. – М.: Техносфера, 2017. – С. 395-401.
12. *Bockleemann D.E. and Eisenstadt W.R.* Combined Differential and Common-Mode Scattering Parameters: Theory and Simulation // IEEE Trans. on Microwave Theory and Techniques. – July 1995. – Vol. MTT-43, No. 7. – P. 520-523.
13. *Курейчик В.М.* Гибридные генетические алгоритмы // Известия ЮФУ. Технические науки. – 2007. – № 2 (77). – С. 5-12.
14. *Zhabin D.A., Garays D.V., Kalentyev A.A., Dobush I.M. and Babak L.I.* Automated Synthesis of Low Noise Amplifiers Using S-parameter Sets of Passive Elements, Asia-Pacific Microwave Conference (APMC 2017), Kuala Lumpur, Malaysia, 2017 (accepted for publication).
15. *Kalentyev A.A., Garays D.V. and Babak L.I.* Genetic-Algorithm-Based Synthesis of Low-Noise Amplifiers with Automatic Selection of Active Elements and DC Biases, European Microwave Week 2014, Rome, Italy. October 2014. – P. 520-523.

16. Babak L.I., Kokolov A.A. and Kalentyev A.A. A New Genetic-Algorithm-Based Technique for Low Noise Amplifier Synthesis, European Microwave Week 2012, Amsterdam, The Netherlands, November 2012. – P. 520-523.
17. Mann G.K.I., Gosine R.G. Three-dimensional min–max-gravity based fuzzy PID inference analysis and tuning // Fuzzy Sets and Systems. – 2005. – Vol. 156. – P. 300-323.
18. Голицын Г.А. Петров В.М. Информация и биологические принципы оптимальности: Гармония и алгебра живого. – М.: КомКнига, 2005.
19. Колесов Ю.Б., Сениченков Ю.Б. Моделирование систем. Динамические и гибридные системы: учеб. - метод. пособие. – СПб.: БХВ-Петербург, 2006.
20. Abraham A., Grosan G., Ramos V. Swarm Intelligence in Data Mining. – Berlin. Heidelberg: SpringerVerlag, 2007. – 267 p.

REFERENCES

1. Tarasov V.B. Ot mnogoagentnykh sistem k intellektual'nym organizatsiyam: filosofiya, psikhologiya, informatika [From multi-agent systems to intellectual organizations: philosophy, psychology, informatics]. Moscow: Editorial URSS, 2002, 352 p.
2. Kapra F. Pautina zhizni. Novoe nauchnoe ponimanie zhivyykh system [The web of life. New scientific understanding of living systems]: trans. from engl. Moscow: ID “Gelios”, 2002.
3. Kureychik V.M., Danil'chenko V.I. Geneticheskii algoritm planirovaniya razmeshcheniya SBIS [Genetic algorithm for planning the placement of VLSI], *Izvestiya YuFU. Tekhnicheskoe nauki* [Izvestiya SFedU. Engineering Sciences], 2019, No. 2 (204), pp. 26-34.
4. Danil'chenko V.I., Kureychik V.M. Klassifikatsiya i analiz metodov resheniya zadachi razmeshcheniya SBIS [Classification and analysis of methods for solving the problem of VLSI placement], *Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie* [Computer science, computer engineering and engineering education], 2018, Issue 1.
5. Danilchenko V.I., Danilchenko Y.V., Kureichik V.M. Bio-inspired Approach to Microwave Circuit Design, *IEEE EAST-WEST DESIGN & TEST SYMPOSIUM*, 2020, pp. 362-366. DIO: 10.1109/EWDTS 50664.2020.9224737.
6. Kalent'ev A.A., Garays D.V., Dobush I.M., Babak L.I. Strukturno-parametricheskii sintez SVCh tranzistornykh usiliteley na osnove geneticheskogo algoritma s ispol'zovaniem modeley monolitnykh elementov [Structural-parametric synthesis of microwave transistor amplifiers based on a genetic algorithm using models of monolithic elements], *Doklady TUSURA* [Reports of TUSUR], December 2012, No. 2 (26), Part 2, pp. 104-112.
7. Tang, Maolin and Yao, Xin. A memetic algorithm for VLSI floorplanning, *IEEE Transactions on Systems, Man, And Cybernetics–Part B: Cybernetics*, 2007, No. 37 (1).
8. Goryainov A.E., Dobush I.M., Babak L.I. Postroenie parametricheskikh modeley passivnykh komponentov SVCh monolitnykh integral'nykh skhem s ispol'zovaniem programmy Extraction-P [Construction of parametric models of passive components of microwave monolithic integrated circuits using the Extraction-P program].
9. Kokolov A.A., Salnikov A.S., Sheyerman F.I. and Babak L.I. Broadband Double-Balanced SiGe BiCMOS Mixer With Integrated Asymmetric MBaluns, *Int. Conf. “Dynamics of Systems, Mechanisms and Machines” (Dynamics-2017), Omsk, Russia, 2017* (accepted for publication).
10. Wenyuan L. and Qian Z. A 0.7–1.9GHz Broadband Pseudo-differential Power Amplifier Using 0.13-um SiGe HBT Technology, *2012 Int. Conf. on Microwave and Millimeter Wave Technology (ICMMT)*, July 2012, pp. 1-4.
11. Kokolov A.A., Dobush I.M., Sheerman F.I., Babak L.I. i dr. Slozhno-funktsional'nye bloki shirokopolosnykh usiliteley radiochastoty dlya odnokristal'nykh priemnikov L- i S-diapazonov na osnove tekhnologii SiGe [Complex-functional blocks of broadband radio frequency amplifiers for single-chip L- and S-band receivers based on SiGe technology], *3-ya Mezhdunar. nauch. konf. «EKB i elektronnye moduli» (Mezhdunarodnyy forum «Mikroelektronika-2017»)*, g. Alushta, oktyabr' 2017 [3rd International Scientific Conference " ECB and Electronic modules "(International Forum "Microelectronics-2017")], Alushta, October 2017]. Moscow: Tekhnosfera, 2017, pp. 395-401.
12. Bocklemann D.E. and Eisenstadt W.R. Combined Differential and Common-Mode Scattering Parameters: Theory and Simulation, *IEEE Trans. on Microwave Theory and Techniques*, July 1995, Vol. MTT-43, No. 7, pp. 520-523.

13. *Kureychik V.M.* Gibridnye geneticheskie algoritmy [Hybrid genetic algorithms], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2007, No. 2 (77), pp. 5-12.
14. *Zhabin D.A., Garays D.V., Kalentyev A.A., Dobush I.M. and Babak L.I.* Automated Synthesis of Low Noise Amplifiers Using S-parameter Sets of Passive Elements, Asia-Pacific Microwave Conference (APMC 2017), Kuala Lumpur, Malaysia, 2017 (accepted for publication).
15. *Kalentyev A.A., Garays D.V. and Babak L.I.* Genetic-Algorithm-Based Synthesis of Low-Noise Amplifiers with Automatic Selection of Active Elements and DC Biases, European Microwave Week 2014, Rome, Italy. October 2014, pp. 520-523,
16. *Babak L.I., Kokolov A.A. and Kalentyev A.A.* A New Genetic-Algorithm-Based Technique for Low Noise Amplifier Synthesis, European Microwave Week 2012, Amsterdam, The Netherlands, November 2012, pp. 520-523.
17. *Mann G.K.I., Gosine R.G.* Three-dimensional min-max-gravity based fuzzy PID inference analysis and tuning, *Fuzzy Sets and Systems*, 2005, Vol. 156, pp. 300-323.
18. *Golitsyn G.A., Petrov V.M.* Informatsiya i biologicheskie printsipy optimal'nosti: Garmoniya i algebra zhivogo [Information and biological principles of optimality: Harmony and algebra of the living]. Moscow: KomKniga, 2005.
19. *Kolesov Yu.B., Senichenkov Yu.B.* Modelirovanie sistem. Dinamicheskie i gibridnye sistemy: ucheb.-metod, posobie [System modeling. Dynamic and hybrid systems: a training manual]. Saint Petersburg: BKhV-Peterburg, 2006.
20. *Abraham A., Grosan G., Ramos V.* Swarm Intelligence in Data Mining. Berlin. Heidelberg: SpringerVerlag, 2007, 267 p.

Статью рекомендовал к опубликованию к.т.н., доцент Ю.П. Волощенко.

Данильченко Евгения Владимировна – Южный Федеральный университет; e-mail: lipkina@sfedu.ru; г. Таганрог, Россия; тел.: +79525691761; кафедра САПР.

Данильченко Владислав Иванович – e-mail: vdanielchenko@sfedu.ru; тел.: +79526088561; кафедра САПР.

Курейчик Виктор Михайлович – e-mail: vmkureychik@sfedu.ru; тел.: +79282132730; кафедра САПР; д.т.н.; профессор.

Danilchenko Yevgenia Vladimirovna – Southern Federal University; e-mail: lipkina@sfedu.ru; Taganrog, Russia; phone: +79525691761; the department of computer-aided design.

Danilchenko Vladislav Ivanovich – e-mail: vdanielchenko@sfedu.ru; phone: +79526088561; the department of computer-aided design.

Kureichik Viktor Mikhailovich – e-mail: vmkureychik@sfedu.ru; phone: +78634311487; the department of computer-aided design; dr. of eng. sc.; professor.

УДК 004.896

DOI 10.18522/2311-3103-2021-2-39-50

В.В. Курейчик, Вл.Вл. Курейчик

ПРОГРАММНАЯ ПОДСИСТЕМА ДЛЯ РЕШЕНИЯ NP-СЛОЖНЫХ КОМБИНАТОРНО-ЛОГИЧЕСКИХ ЗАДАЧ НА ГРАФАХ*

Работа посвящена созданию программной подсистемы для решения NP-трудных и NP-сложных комбинаторно-логических задач на графах. В статье приведено описание комбинаторно-логических задач на графах. Для эффективного их решения предлагаются новые многоуровневые архитектуры поиска, такие как простая комбинированная, параллельная комбинированная, двухуровневая, интегрированная и гибридная. Данные архитектуры основаны на методах, инспирированных природными системами. Ключевым отлич-

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-01-00059.

ем данных архитектур является разделение поиска на два или три уровня и применение на них различных алгоритмов эволюционного моделирования и биоинспирированного поиска. Это позволяет получать наборы квазиоптимальных решений выполнять параллельную обработку и частично устранять проблему преждевременной сходимости. В статье приведено подробное описание разработанной программной подсистемы и ее модулей. В качестве модулей в подсистеме имеется пять разработанных архитектур и набор разработанных алгоритмов эволюционного моделирования и биоинспирированного поиска, таких как эволюционный, генетический, пчелиный, муравьиный, светлячковый и обезьяний. Благодаря модульной структуре в подсистеме имеется возможность конструировать более 50 различных вариантов комбинаций поиска. Это позволяет использовать все достоинства методов биоинспирированной оптимизации для эффективного решения NP-сложных комбинаторно-логических задач на графах. Для подтверждения эффективности разработанной программной подсистемы был проведен вычислительный эксперимент на тестовых примерах. Проведенные серии тестов и экспериментов показали преимущество использования программного продукта для решения комбинаторно-логических задач на графах большой размерности, по сравнению с известными алгоритмами, что говорит о перспективности применения такого подхода. Временная сложность разработанных алгоритмов в лучшем случае $\approx O(n \log n)$, в худшем случае – $O(n^3)$.

Программная подсистема; комбинаторно-логические задачи на графах; многоуровневые архитектуры; эволюционное моделирование; биоинспирированный поиск.

V.V. Kureichik, VI.VI. Kureichik

SOFTWARE SUBSYSTEM FOR SOLVING NP-COMPLEX COMBINATORIAL LOGIC PROBLEMS ON GRAPHS

The paper is devoted to the development of the software for solving NP-hard and NP-hard combinatorial-logical problems on graphs. The paper contains a description of graphs combinatorial-logical problems. New multilevel search architectures such as simple combo, parallel combo, two-levels, integrated, and hybrid are proposed to effectively address them. These architectures are based on methods inspired by natural systems. The key difference between these architectures is the division of search into two or three levels and the use of various algorithms for evolutionary modeling and bioinspired search on them. This allows obtaining sets of quasi-optimal solutions to perform parallel processing and partially eliminate the problem of premature convergence. The article provides a detailed description of the developed software subsystem and its modules. As modules in the subsystem, there are five developed architectures and a set of developed algorithms for evolutionary modeling and bioinspired search, such as evolutionary, genetic, bee, ant, firefly and monkey. Thanks to its modular structure, the subsystem has the ability to design more than 50 different search combinations. This makes it possible to use all the advantages of bioinspired optimization methods for efficiently solving NP-complex combinatorial-logical problems on graphs. To confirm the effectiveness of the developed software subsystem, a computational experiment was carried out on test examples. The series of tests and experiments carried out have shown the advantage of using a software product for solving combinatorial-logical problems on graphs of large dimension, in comparison with known algorithms, which indicates the prospects of using this approach. The time complexity of the developed algorithms is $\approx O(n \log n)$ at best, and $O(n^3)$ at worst.

Program system; graph combinatorial and logic problems; multilevel architecture; evolutionary modeling; bioinspiring search.

Введение. В настоящее время теория графов является одним из самых востребованных инструментов при создании искусственного интеллекта. При решении прикладных задач искусственного интеллекта особое значение приобретают эффективные методы решения комбинаторно-логических задач на графах. Их решения с успехом используются в различных областях науки и техники. Разработка эффективных методов и алгоритмов для решения задач на графах является актуальной проблемой в связи с тем, что эти задачи являются NP-полными и NP-трудными [1]. Появление новых, более совершенных средств электронно-

вычислительной техники является причиной для модернизации и разработки новых эффективных подходов и методов при решении комбинаторно –логических задач на графах. Одним из таких подходов является многоуровневая оптимизация, основанная на методах эволюционного моделирования и биоинспирированного поиска [2–9]. Данный подход позволяет получать наборы квазиоптимальных решений за приемлемое время.

1. Описание задачи. Под комбинаторно-логическими задачами на графах понимается задача, в которой необходимо найти наилучшее решение, или, как говорят, оптимальное. Данное решение должно приниматься оптимальным на основе критерия (меры оценки исследуемого явления) или целевой функции (ЦФ) [10]. В настоящее время существует большое количество комбинаторно-логических задач на графах, и они могут иметь различную природу. Однако их постановка имеет много аналогий. Для эффективного их решения предлагаются новые многоуровневые архитектуры поиска, основанные на методах, инспирированных природными системами. Ключевым отличием данного подхода является разделение поиска на два уровня и применение на них различных биоинспирированных алгоритмов [9, 11–13]. Это позволяет выполнять параллельную обработку и частично устранять проблему их преждевременной сходимости. Для реализации предложенного подхода и разработанных алгоритмов создана программная подсистема.

2. Программная подсистема. Главной целью разработки программной подсистемы стала реализация в вычислительном комплексе построенных многоуровневых архитектур и алгоритмов, позволяющих эффективно решать NP-сложные комбинаторно-логические задачи на графах. На рис. 1. представлена укрупненная архитектура разработанной программной подсистемы, предназначенная для эффективного решения NP-сложных комбинаторно-логических задач на графах, а также для проведения вычислительных экспериментов. Здесь сначала выполняется ввод исходных данных о решаемой задаче на графах. Далее вводятся параметры разработанных биоинспирированных алгоритмов. Затем переход на следующий блок, где выбирается тот или иной вычислительный метод. Отметим, что в разработанной программной подсистеме предусмотрены две базы данных, в которых находятся построенные архитектуры и набор разработанных алгоритмов биоинспирированной оптимизации. Данные базы подключаются к блоку «Вычислительный метод». После реализации поиска в блоке статистики можно ознакомиться с результатами процесса оптимизации.

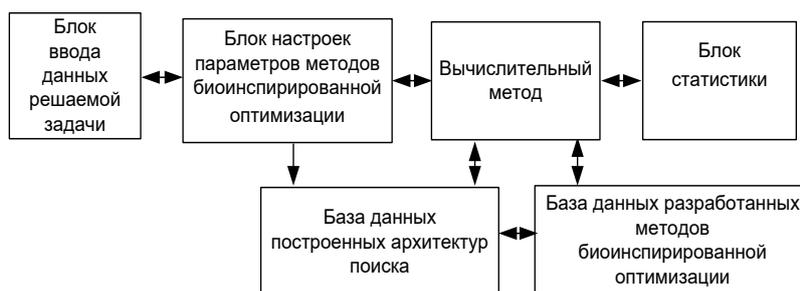


Рис. 1. Укрупненная архитектура программной подсистемы

Интерфейс главного окна программной подсистемы приведен на рис. 2.

Разработка выполнялась в среде программирования Borland C++ Builder, Visual C++. Отладка и тестирование проводилось на ЭВМ типа IBM PC с процессором core i7 с ОЗУ-8Гб. Входными параметрами для работы подсистемы являются

ся случайные графы (количество вершин и ребер), матрица смежности графа выбор схемы поиска; выбор реализуемого алгоритма, размер популяции, количество итераций, порядок и вероятности применения операторов. При этом выходными параметрами являются: время работы алгоритмов и лучшее решение, достигнутое в процессе работы.

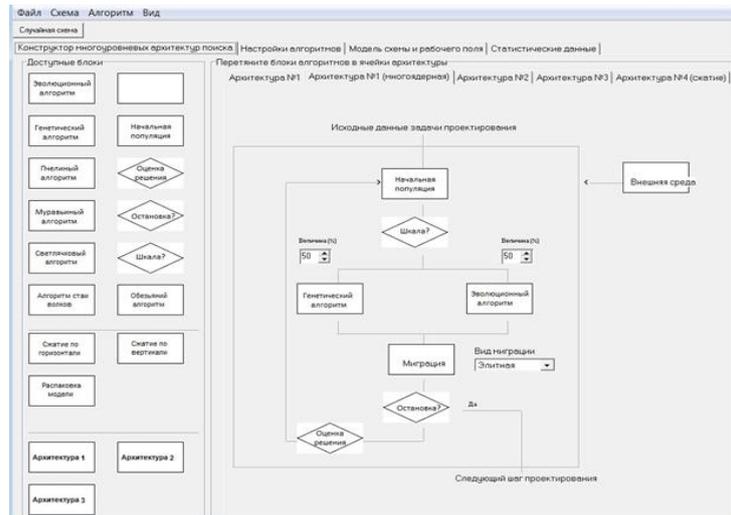


Рис. 2. Интерфейс главного окна программной подсистемы

Главной особенностью программной подсистемы является наличие конструктора архитектур, реализующие многоуровневые подходы. Многоуровневые подходы легко адаптируются к внешней среде, так как позволяют достаточно легко встраивать различные эвристики, направленные на улучшение получаемого решения, в свою структуру. Конструктор архитектур состоит из области доступных блоков и пяти разработанных типовых шаблонов. В качестве архитектуры №1 в программной подсистеме используется модифицированная комбинированная архитектура поиска, состоящая из двух вычислительных алгоритмов эволюционного и генетического [14] рис. 3., хотя можно использовать и другие методы биоинспирированного поиска.

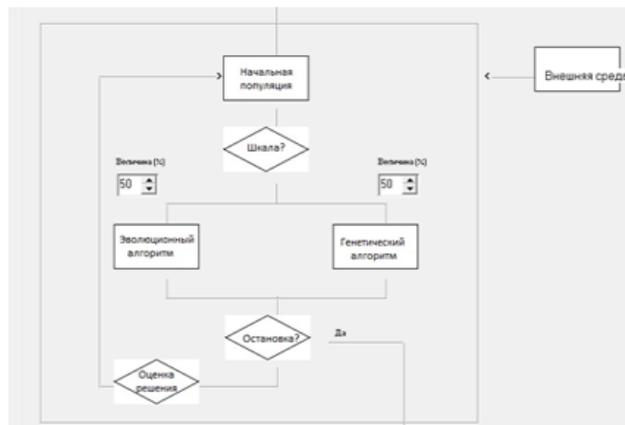


Рис. 3. Простая комбинированная архитектура

На основе данной архитектуры можно реализовать два вида поиска: последовательный и параллельный [15]. При реализации последовательного поиска реализуется только эволюционный алгоритм. Данный алгоритм основан на выполнении только одного оператора - мутации и его различных модификаций. Данный метод позволяет достаточно быстро получать наборы квазиоптимальных решений за приемлемое время. При получении неудовлетворительных результатов подключается блок выполнения генетического алгоритма. Генетический алгоритм реализует весь набор операторов и их модификаций. Этот процесс реализуется через блок эволюционной адаптации.

Параллельный поиск, используя многопроцессорную или многоядерную систему, позволяет одновременно и независимо друг от друга реализовать генетический и эволюционный алгоритмы. Процесс поиска ведется в течение нескольких заданных генераций. В архитектуру данного поиска вводится блок миграции рис. 4. Блок миграции позволяет производить обмен решениями, что ускоряет процесс получения эффективных решений и помогает избежать попадания в локальные оптимумы.

В качестве архитектуры №2 в программной подсистеме используется двухуровневый подход, сочетающий в себе методы биоинспирированного поиска и эволюционного моделирования. В данной архитектуре, приведенной на рис. 5. в качестве оптимизационных методов на первом уровне применяются методы биоинспирированного поиска, а на втором уровне применены - эволюционный и генетический алгоритмы, хотя можно использовать и другие методы биоинспирированной оптимизации.

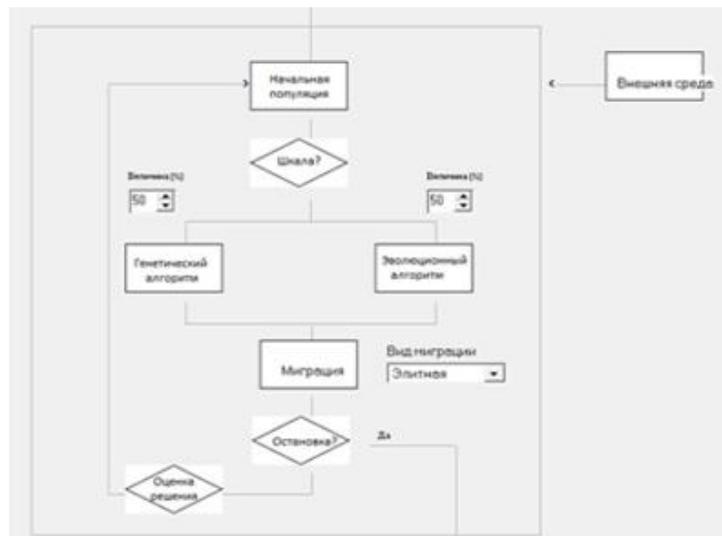


Рис. 4. Комбинированная архитектура для выполнения параллельных процедур

В качестве архитектуры №3 в программной подсистеме используется интегрированная схема поиска на основе многоуровневого подхода [16]. Такая трехуровневая архитектура, в которой реализованы четыре алгоритма биоинспирированного поиска, приведена на рис. 6.

Здесь на первом уровне используется пчелиный алгоритм [17], который быстро разбивает всю область поиска на подобласти с высоким значением ЦФ. На втором уровне реализуется муравьиный алгоритм [18], который позволяет за кратчайшее время получить список критических связей, а также определять и запоминать длину кратчайших маршрутов.

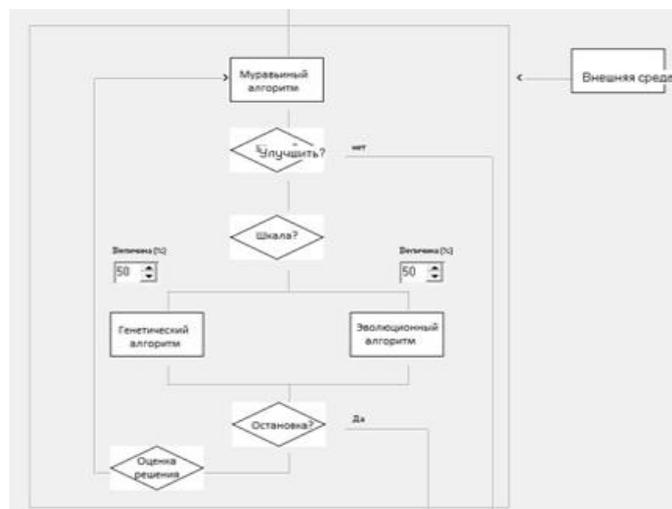


Рис. 5. Двухуровневая архитектура поиска

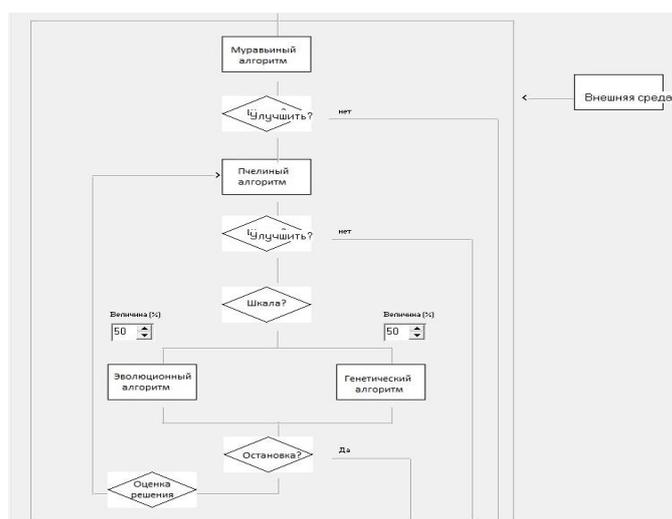


Рис. 6. Интегрированная архитектура поиска

И на третьем уровне если это необходимо на основе шкалы производится выбор того или иного метода эволюционного моделирования. При наличии многопроцессорной или многоядерной системы данные методы могут выполняться параллельно.

В качестве архитектуры №4 в программной подсистеме используется гибридная схема поиска рис. 7. Его основная идея состоит в том, что на первом уровне уменьшается размерность задачи за счет сжатия графовой модели [19], а на втором уровне выполняется оптимизация на основе методов эволюционного моделирования или биоинспирированного

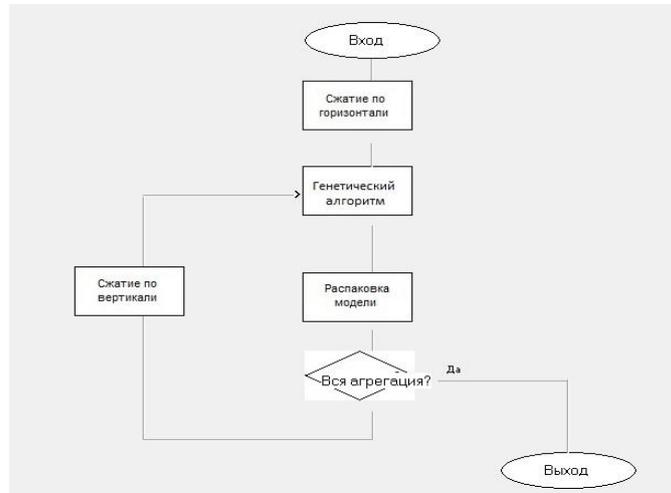


Рис. 7. Гибридная трехуровневая архитектура поиска

Заметим, что данный подход позволяет уменьшать размерность решаемой задачи, получать наборы квазиоптимальных решений за полиномиальное время.

Отметим, что разработанные в программной подсистеме архитектуры могут быть модифицированы на любое число уровней в зависимости от используемых средств вычислительной техники и временных ограничений на получение результата.

Данная программная подсистема, кроме разработанных архитектур, также поддерживает различные разработанные алгоритмы биоинспирированной оптимизации, из которых как из строительных блоков, показанных на рис. 8, собираются архитектуры различной сложности. Это эволюционный, генетический, пчелиный, муравьиный, светлячковый и обезьяний алгоритмы [3].



Рис. 8. Набор алгоритмов и строительных блоков

Далее для работы программной подсистемы проводится настройка параметров разработанных алгоритмов биоинспирированной оптимизации. Данное окно показано на рис. 9.

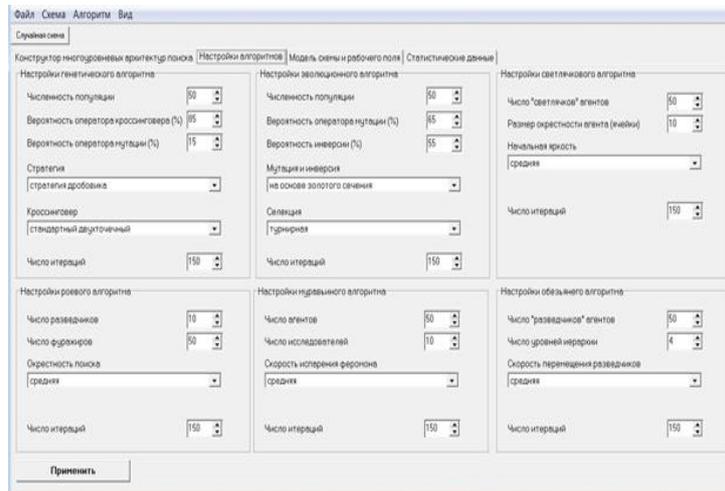


Рис. 9. Окно настройки параметров алгоритмов

Здесь производится ввод параметров алгоритмов биоинспирированной оптимизации и выполняется запуск программной системы на решение поставленной задачи. После окончания работы программной подсистемы выводится окно статистики и график изменения значения целевой функции (ЦФ), что отражает рис. 10. В данном окне показываются входные данные графовой модели и выходные результаты моделирования, такие как: значения начальной и конечной ЦФ; используемая в процессе поиска архитектура; а также время работы подсистемы в целом.

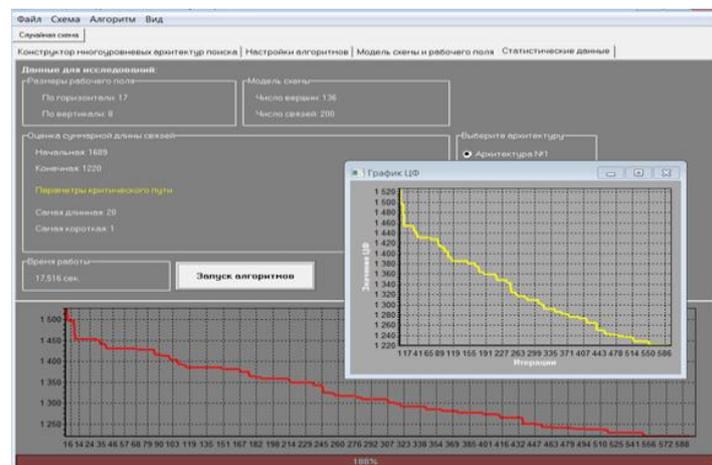


Рис. 10. Окно статистики

Заметим, что в разработанной программной подсистеме имеется возможность в строительных блоках изменять некоторые значения управляющих параметров поиска решений. Это блоки шкала выбора алгоритма, миграции и критерия остановки поиска.

Для некоторых предустановленных блоков архитектуры в типовых шаблонах предусмотрена возможность изменения значений управляющих параметров, например, таких как процент миграции начальной популяции, шкала выбора пути следования (по вероятностному критерию), настройка сходимости для определения критерия остановки или необходимости применения этапа выхода из локального оптимума.

3. Вычислительный эксперимент. Для подтверждения эффективности разработанной программной подсистемы был проведен вычислительный эксперимент на тестовых примерах [20]. Сущность тестирования заключается в запуске предложенных архитектур на одной конкретной тестовой задаче и определении значений ЦФ и времени решения. Здесь в качестве методов, реализующих тестовые примеры были выбраны известные методы QPlace [21], mPG [22] и разработанные и программно реализованные авторами четыре многоуровневые архитектуры A1, A2, A3, A4. Приведем зависимость времени и качества работы данных архитектур от тестовых схем. Полученные результаты вычислительного эксперимента представлены в виде гистограмм зависимостей, приведенных на рис. 11 и 12.

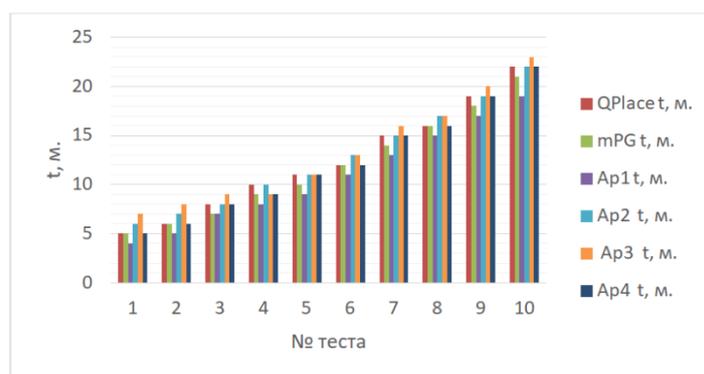


Рис. 11. Гистограмма сравнения времени работы архитектур поиска при реализации тестового примера

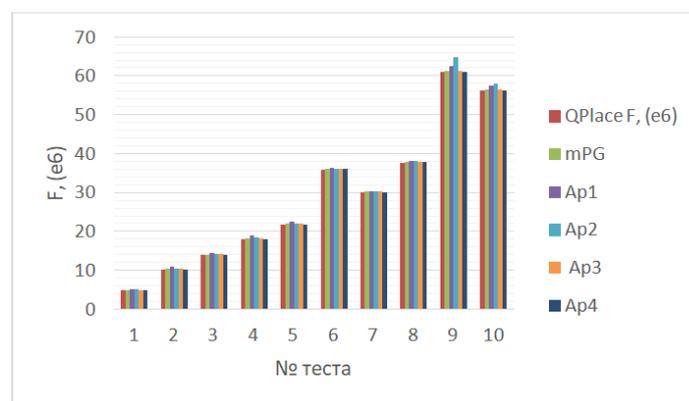


Рис. 12. Гистограмма сравнения значений целевой функции архитектур поиска при реализации тестового примера

В результате анализа представленных графиков зависимостей можно сделать вывод, что разработанные архитектуры биоинспирированного поиска (Ар1, Ар2, Ар3, Ар4) при реализации тестового примера [20] показали сопоставимые результаты по времени решения с известными программными средствами Qplace и mPG. Причем разработанная архитектура Ар4 по качеству решения при реализации тестового примера в среднем на 10 % превосходит другие разработанные архитектуры поиска и на 8 % превосходит известные программные средства Qplace и mPG.

Заключение. Отметим, что предложенный и разработанный функционал конструирования многоуровневых модифицированных архитектур биоинспирированного поиска работает на основе пяти типовых шаблонов и ограниченного числа алгоритмов (блоков), но даже этот ограниченный набор управляющих параметров предоставляет возможность комбинирования более 50 вариантов конкретных архитектур поиска. Это позволяет использовать все достоинства методов биоинспирированной оптимизации для эффективного решения NP-сложных комбинаторно-логических задач на графах. Такой подход дает возможность распараллеливать процесс оптимизации, получать оптимальные и квазиоптимальные решения за полиномиальное время, а также частично решать проблему преждевременной сходимости.

Проведенные экспериментальные исследования, показали преимущество использования программного продукта для решения комбинаторно-логических задач на графах большой размерности, по сравнению с известными методами, так как предложенные архитектуры многоуровневого поиска позволяют получать решения, сопоставимые с известными методами по времени решения, и превосходят последние по качеству получаемых решений в среднем на 8 %.

Временная сложность алгоритмов, реализованных на таких архитектурах, лежит в пределах $O(n \log n) - O(n^3)$, где n – число входов алгоритмов.

Благодарности. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-01-00059.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Кормен Т., Лейзерсон И., Ривест Р.* Алгоритмы: построения и анализ. – М.: МЦМО, 2000.
2. *De Jong K.* Evolutionary Computation: Recent Development and Open Issues // Proceedings 1st International conf., Evolutionary Computation and Its Application. – М., 1996. – P. 7-18.
3. *Карпенко А.П.* Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой: учеб. пособие. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2014. – 446 с.
4. *Abraham A., Ramos V., Grosan G.* Swarm Intelligence in Data Mining. – Berlin. Heidelberg: Springer Verlag, 2007.
5. *Hassaniien E. Emary E.* Swarm Intelligence. Principles Advances, and Applications. – CRC Press, 2015.
6. *Mourelle M., Nedjah L. De.* Swarm intelligent systems. – Berlin: Heidelberg: Springer Verlag, 2006.
7. *Родзин С.И., Курейчик В.В.* Состояние, проблемы и перспективы развития биоэвристик // Программные системы и вычислительные методы. – 2016. – № 2. – С. 158-172.
8. *Курейчик В.В., Курейчик Вл.Вл.* Биоинспирированный поиск при проектировании и управлении // Известия ЮФУ. Технические науки. – 2012. – № 11 (136). – С. 178-183.
9. *Kureichik V., Kureichik Vl., Bova V.* Placement of VLSI fragments based on a multilayered approach // Advances in Intelligent Systems and Computing. – 2016. – Vol. 464. – P. 181-190.
10. *Курицкий Б.Я.* Оптимизация вокруг нас: учебник. – СПб.: Машиностроение, 1989.
11. *Hendrickson B., Leland R.* A Multilevel Algorithm for Partitioning Graphs // Proceedings of the 1995 ACM/IEEE conference on Super computing. – P. 626-657.
12. *Schloegel K., Karypis G., Kumar V.* Multilevel diffusion schemes for repartitioning of adaptive meshes. – University of Minnesota, Department of Computer Science, 1997. – P. 109-124.

13. *Barnard S.T., Simon H.D.* A fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems // *Proceedings 6th SIAM Conf. Parallel Processing for Scientific Computing*. – 1993. – P. 711-718.
14. *Holland John H.* *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Application to Biology, Control, and Artificial Intelligence*. – USA: University of Michigan, 1975. – 183 p.
15. *Курейчик В.М., Курейчик В.В., Родзин С.И.* Модели параллелизма эволюционных вычислений // *Вестник Ростовского государственного университета путей сообщения*. – 2011. – № 3 (43). – С. 93-97.
16. *Бова В.В., Курейчик В.В.* Интегрированная подсистема гибридного и комбинированного поиска в задачах проектирования и управления // *Известия ЮФУ. Технические науки*. – 2010. – № 12 (113). – С. 37-42.
17. *Karaboga D.* An idea based on honey bee swarm for numerical optimization. – Erciyes University, Engineering Faculty, Computer Engineering Department, 2005. – 110 p.
18. *Dorigo M., Maniezzo V., Colomi A.* The Ant System: Optimization by a colony of cooperating objects // *IEEE Trans. on Systems, Man, and Cybernetics*. – 1996. – Part B. – No. 26 (1). – P. 29-41.
19. *Курейчик В.В., Курейчик Вл.Вл.* Размещения фрагментов СБИС на основе механизма агрегации фракталов // *Известия ЮФУ. Технические науки*. – 2015. – № 2 (163). – С. 196-205.
20. *Adya S.N., Markov I.L.* ISPD02 IBM-MS Mixed-size Placement Benchmarks. – <http://vlsicad.eecs.umich.edu/BK/ISPD02bench/>.
21. *Adya S.N., Markov I.L.* Consistent placement of macro-blocks using floor planning and standard-cell placement // *In Proc. Intl. Symp. on Physical Design*. – 2002. – P. 12-17.
22. *Wang M., Yang X., Sarrafzadeh M.* Dragon 2000: Standard-cell Placement Tool for Large Industry Circuits // *ICCAD – 2000*. – P. 260-263.

REFERENCES

1. *Kormen T., Leyzerson I., Rivest R.* *Algorithms: constructions and analysis*. Moscow: MTSMO, 2000.
2. *De Jong K.* Evolutionary Computation: Recent Development and Open Issues, *Proceedings 1st International conf., Evolutionary Computation and Its Application*. Moscow, 1996, pp. 7-18.
3. *Karpenko A.P.* *Sovremennyye algoritmy poiskovoy optimizatsii. Algoritmy, vdokhnovlennyye prirodoy: ucheb. posobie* [Modern search engine optimization algorithms. Algorithms inspired by nature: a textbook]. Moscow: Izd-vo MGTU im. N.E. Bauman, 2014, 446 p.
4. *Abraham A., Ramos V., Grosan G.* *Swarm Intelligence in Data Mining*. Berlin. Heidelberg: Springer Verlag, 2007.
5. *Hassanien E., Emary E.* *Swarm Intelligence. Principles Advances, and Applications*. CRC Press, 2015.
6. *Mourelle M., Nedjah L. De.* *Swarm intelligent systems*. Berlin: Heidelberg: Springer Verlag, 2006.
7. *Rodzin S.I., Kureychik V.V.* Sostoyanie, problemy i perspektivy razvitiya bioevristik [State, problems and prospects of development of bio-heuristics], *Programmnye sistemy i vychislitel'nye metody* [Software systems and computational methods], 2016, No. 2, pp. 158-172.
8. *Kureychik V.V., Kureychik V.I.* Bioinspirirovannyy poisk pri proektirovanii i upravlenii [Bioinspired search in design and management], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 11 (136), pp. 178-183.
9. *Kureichik V., Kureichik V.I., Bova V.* Placement of VLSI fragments based on a multilayered approach, *Advances in Intelligent Systems and Computing*, 2016, Vol. 464, pp. 181-190.
10. *Kuritskiy B.Ya.* *Optimizatsiya vokrug nas: uchebnyk* [Optimization around us: a tutorial]. Saint Petersburg: Mashinostroenie, 1989.
11. *Hendrickson B., Leland R.* A Multilevel Algorithm for Partitioning Graphs, *Proceedings of the 1995 ACM/IEEE conference on Super computing*, pp. 626-657.
12. *Schloegel K., Karypis G., Kumar V.* Multilevel diffusion schemes for repartitioning of adaptive meshes. University of Minnesota, Department of Computer Science, 1997, pp. 109-124.
13. *Barnard S.T., Simon H.D.* A fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems, *Proceedings 6th SIAM Conf. Parallel Processing for Scientific Computing*, 1993, pp. 711-718.

14. *Holland John H.* Adaptation in Natural and Artificial Systems: An Introductory Analysis with Application to Biology, Control, and Artificial Intelligence. USA: University of Michigan, 1975, 183 p.
15. *Kureychik, V.M., Kureychik V.V., Rodzin S.I.* Modeli parallelizma evolyutsionnykh vychisleniy [Models of parallelism of evolutionary calculations], *Vestnik Rostovskogo gosudarstvennogo universiteta putey soobshcheniya* [Bulletin of the Rostov State University of Railway Engineering], 2011, No. 3 (43), pp. 93-97.
16. *Bova V.V., Kureychik V.V.* Integrirovannaya podsistema gibridnogo i kombinirovannogo poiska v zadachakh proektirovaniya i upravleniya [Integrated subsystem of hybrid and combined search in design and management tasks], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2010, No. 12 (113), pp. 37-42.
17. *Karaboga D.* An idea based on honey bee swarm for numerical optimization. Erciyes University, Engineering Faculty, Computer Engineering Department, 2005, 110 p.
18. *Dorigo M., Maniezzo V., Colnani A.* The Ant System: Optimization by a colony of cooperating objects, *IEEE Trans. on Systems, Man, and Cybernetics*, 1996, Part B, No. 26 (1), pp. 29-41.
19. *Kureychik V.V., Kureychik, V.V.* Razmeshcheniya fragmentov SBIS na osnove mekhanizma agregatsii fraktalov [Placement of VLSI fragments based on the fractal aggregation mechanism], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2015, No. 2 (163), pp. 196-205.
20. *Adya S.N., Markov I.L.* ISPD02 IBM-MS Mixed-size Placement Benchmarks. Available at: <http://vlsicad.eecs.umich.edu/BK/ISPD02bench/>.
21. *Adya S.N., Markov I.L.* Consistent placement of macro-blocks using floor planning and standard-cell placement, *In Proc. Intl. Symp. on Physical Design*, 2002, pp. 12-17.
22. *Wang M., Yang X., Sarrafzadeh M.* Dragon 2000: Standard-cell Placement Tool for Large Industry Circuits, *ICCAD – 2000*, pp. 260-263.

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; зав. кафедрой САПР; д.т.н.; профессор.

Курейчик Владимир Владимирович – ООО «Газпром подземремонт Уренгой»; e-mail: Kureichik@yandex.ru; г. Санкт-Петербург, Россия; тел.: +78126096700; отдел автоматизации систем управления; к.т.н.; ведущий инженер.

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

Kureichik Vladimir Vladimirovich – «Gazprom podzemremont Urengoi» company; e-mail: Kureichik@yandex.ru; St. Peterburg, Russia; phone: +78126096700; Control systems automation department; cand. of eng. sc.; lead engineer.

УДК 004.896

DOI 10.18522/2311-3103-2021-2-50-59

В.В. Курейчик, Вл.Вл. Курейчик, А.Э. Саак

ЭВОЛЮЦИОННЫЙ АЛГОРИТМ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ДИСПЕТЧЕРИЗАЦИИ*

Рассмотрена одна из важных задач оптимизации – задача диспетчеризации. Она относится к классу NP- сложных оптимизационных задач. В работе приведена и описана постановка задачи диспетчеризации. Здесь массив заявок пользователей на компьютерное

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-01-00148.

обслуживание в Grid- системах моделируется протяжённой линейной полиэдралью координатных ресурсных прямоугольников. При этом диспетчирование представляется локализацией линейной полиэдры в оболочку области вычислительно-временных ресурсов системы согласно многоцелевому критерию качества применяемого назначения заявок на обслуживание. В связи со сложностью данной задачи для ее эффективного решения предлагаются методы эволюционного моделирования. В статье предложена и описана модифицированная архитектура эволюционного поиска. В качестве модификации введены дополнительно три блока. Это блок «внешней среды», блок эволюционной адаптации и блок «неперспективных решений». Для ее реализации авторами разработан модифицированный эволюционный алгоритм, использующий в качестве отбора решений модели эволюций Ч. Дарвина и Ж. Б. Ламарка. Это позволяет значительно сократить время получения результата, частично решить проблему преждевременной сходимости алгоритма и получить наборы квазиоптимальных решений за полиномиальное время. Разработан программный модуль на языке C#. Проведен вычислительный эксперимент на тестовых примерах. Проведенные экспериментальные исследования, показали, что качество решений, полученных на основе разработанного эволюционного алгоритма, в среднем на 5 процентов превосходит результаты решений, полученные с использованием известных алгоритмов последовательного, начально-кольцевого и уровневого при сопоставимом времени, что говорит об эффективности предложенного подхода.

Grid-система; диспетчирование; эволюционное моделирование; эволюционный алгоритм.

V.V. Kureichik, VI.VI. Kureichik, A.E. Saak

AN EVOLUTIONARY ALGORITHM FOR SOLVING THE DISPATCHING PROBLEM

The paper considers one of the most important optimization tasks – the dispatching task that belongs to the class of NP-complex optimization problems. The paper presents the formulation of this problem. In Grid systems the array of users' requests for computer services is modelled by an extended linear polyhedral of coordinate resource rectangles. In this case, dispatching is represented by the localization of a linear polyhedron in the envelope of the area of computational and time resources of the system according to the multipurpose criterion of the quality of the applied assignment. Due to the complexity of this problem, the authors propose methods of evolutionary modelling for its effective solution and describe a modified evolutionary search architecture. Three additional blocks are introduced as a modification. This is a block of "external environment", a block of evolutionary adaptation and a block of "unpromising solutions." The authors have developed a modified evolutionary algorithm that uses the Darwin's and Lamarck's evolution models. This makes it possible to significantly reduce the time for obtaining the result, partially solve the problem of premature convergence of the algorithm, and obtain sets of quasi-optimal solutions in polynomial time. A software module has been developed in the C # language. A computational experiment has carried out on test examples and shown that the quality of solutions obtained on the basis of the developed evolutionary algorithm is, on average, 5 percent higher than the results of solutions obtained using the known algorithms of sequential, initial-ring and level at comparable time, which indicates the effectiveness of the proposed approach.

Grid-system; dispatching; evolutionary modelling; evolutionary algorithm.

Введение. Экспоненциальный рост вычислительной мощности, описываемый законом Мура, пропускной способности сетей связи, описываемый законом Баттера и ёмкости хранения, описываемый законом Крайдера, показывают неудовлетворённость пользователей существующей информационно-коммуникационной инфраструктурой. Всё возрастающие потребности пользователей, задачи, требующие больших объёмов вычислений, привели к распараллеливанию вычислений, созданию многопроцессорных вычислительных систем, многоядерных процессоров, развитию технологий облачных вычислений и Grid-систем. В Grid- системах актуальной задачей, во многом определяющей эффективность функционирования, является управление, распределение вычислительно-временных ресурсов при обслуживании набора заявок [1–6].

Принцип оптимизации с машинным поиском наилучшего распределения массива планарных ресурсных элементов по многоцелевому критерию решает задачу диспетчирования множественным компьютерным обслуживанием с неприемлемой трудоёмкостью составления расписания. Таким образом диспетчеризация является NP-сложной задачей, не имеющей точных алгоритмов ее решения за полиномиальное время [7–11].

В связи с этим, разработка новых эффективных методов и алгоритмов решения задачи диспетчеризации является актуальной и важной задачей.

1. Постановка задачи. Возникшая в конце 70-х годов прошлого века, задача диспетчирования процессорным и временным ресурсами остаётся актуальной и для современных параллельных систем, в том числе, Grid-систем. Геометрической интерпретацией первоначальной постановки задачи предлагалась упаковка прямоугольников в полубесконечную полосу с минимизацией занятой части. В этой постановке моделью задачи служил ориентированный прямоугольник, с мерой сторон, равной числу единиц процессорного и временного ресурса. Модель параллельной системы представлялась горизонтальной полосой с процессорным ограниченным ресурсом. Экспоненциальное развитие цифровых технологий позволяет считать процессорный и временной ресурсы – равноценными, паритетными. Соответственно, моделью Grid-системы предлагается принять первый квадрант и сформулировать задачу диспетчирования как упаковку ориентированных прямоугольников в этом квадранте (рис. 1) [7, 11].

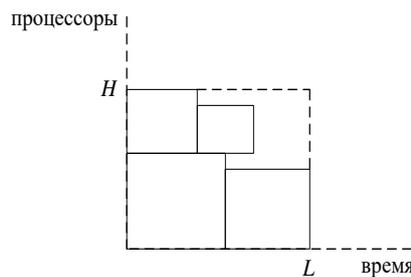


Рис. 1. Ресурсная оболочка заявок пользователей

Здесь эффективность диспетчирования определяется минимизацией занятой части квадранта, оцениваемой по неэвклидовой эвристической мере, учитывающей площадь и форму занятой ресурсной области:

$$M = \frac{1}{2} \left(\frac{LH + (L-H)^2}{\sum_{j=0}^{k-1} a(j)b(j)} \right),$$

где L – протяжённость, H – уровень по вертикали ресурсной оболочки, $a(j)$ – число единиц времени, $b(j)$ – число единиц процессоров, требующихся j -й заявке.

В работе предлагаются эвристические алгоритмы диспетчирования и оценивается их эффективность. При этом, предложенная эвристическая мера принимает минимальное значение, равное 0,5 при безпустотной укладке в квадрат.

2. Методы эволюционного моделирования. Методы эволюционного моделирования – это направление исследований в теории и практике искусственного интеллекта. Ключевой идеей этих методов является использование в качестве основных операций некоторых формализованных принципов естественного эволюцион-

ного отбора [12]. В настоящее время число разновидностей алгоритмов данного класса достаточно велико. Но все данные алгоритмы различаются между собой реализацией тех или иных процессов и механизмов, происходящих в живой природе.

Теоретической основой данных методов является моделирование эволюционного процесса, основанного на принципе «выживания сильнейших».

Методы эволюционного моделирования являются приближенными (эвристическими) методами решения задач оптимизации и структурного синтеза. Они в основном основаны на статистическом подходе к исследованию ситуаций, а также при поиске решений используют итерационное приближение. Одним из основных методов эволюционного моделирования являются эволюционные алгоритмы.

Эволюционные алгоритмы (ЭА) как новое направление оптимизации начали развиваться в начале 80-х гг. [12–17]. В настоящее время их начали широко использовать для решения различных практических задач науки и техники, в том числе для решения задач проектирования. Данные методы позволяют обрабатывать большие области поиска с учётом множества критериев.

В данных методах любое альтернативное решение задачи представляется в виде хромосомы. В свою очередь хромосома состоит из дискретных элементов, называемых генами. Они размещаются по позициям – локусам и могут иметь различные функциональные значения.

ЭА нацелены на оптимизацию целевой функции. Их эффективное использование зависит от способа представления и отбора решений, выбора операторов мутации и принципов формирования начальной популяции альтернативных решений.

Процесс реализации данных методов заключается в последовательном выполнении следующих операций. Предварительно на основе известных принципов [15] создается популяция решений рассматриваемой задачи. Далее процесс представляет собой последовательность циклов выполнения операторов мутации с определенной вероятностью. При чем в каждом цикле происходит оценка всех альтернативных решений решаемой задачи. Далее альтернативные решения, в зависимости от полученной оценки, либо участвуют в создании новой популяции, либо устраняются из дальнейшего поиска. Процесс поиска считается завершённым по выполнению критерия окончания работы алгоритма. Заметим, что данные методы эффективно и одновременно способны анализировать различные области поискового пространства, а также позволяют быстро адаптироваться к нахождению новых, с точки зрения значения ЦФ, лучших областей.

В связи с этим обстоятельством и сложностью решаемой задачи, авторы предлагают использовать методы эволюционного моделирования. Применение этих методов позволит эффективно справиться с трудностями, возникающими при использовании классических подходов.

Базовая структура эволюционного поиска представлена на рис. 2 [15, 17].

В отличие от базовой структуры эволюционного поиска авторы предлагают следующую модифицированную архитектуру и технологию эволюционного поиска, представленную на рис. 3. В качестве модификации в известную архитектуру эволюционного поиска в диссертационной работе предлагается ввести три новых блока. Это блок «внешней среды», блок эволюционной адаптации и блок «неперспективных решений», а также использовать комбинацию двух моделей эволюции. Это модели эволюции Ч. Дарвина и Ж.Б. Ламарка. Здесь будут использованы следующие механизмы отбора: наследование «благоприобретенных признаков» и «выживание сильнейших» [18].

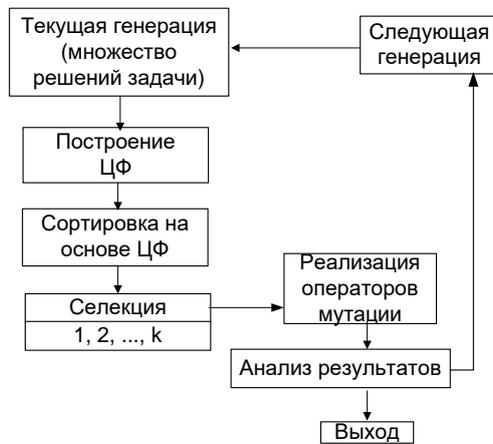


Рис. 2. Базовая структура эволюционного поиска

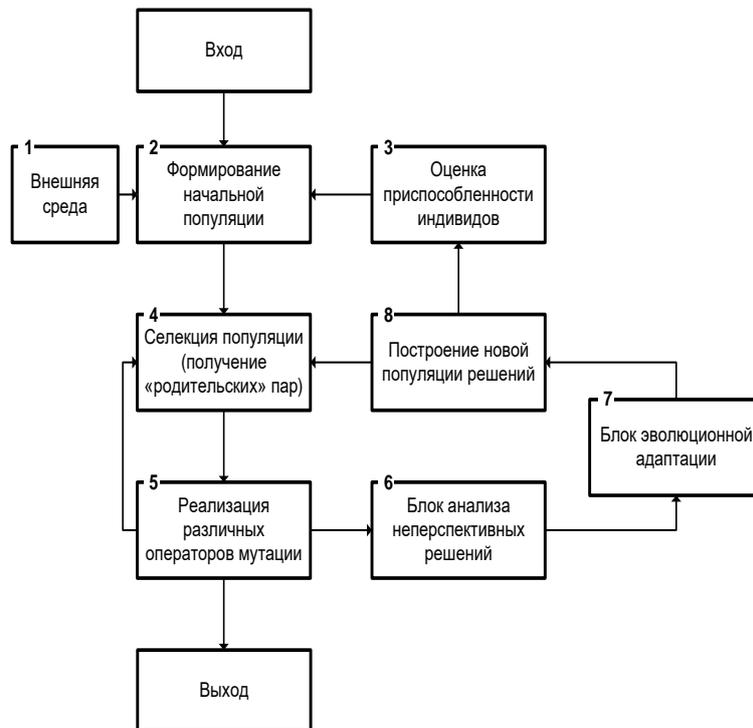


Рис. 3. Модифицированная архитектура эволюционного поиска

Рассмотрим более детально модифицированную архитектуру эволюционного поиска и опишем работу ее блоков. Она представляется следующим образом. Предварительно задается область поиска допустимых решений с учетом ограничений. Таких как число единиц процессоров, число единиц времени, площадь и форма занятой ресурсной области. Затем на основе известных принципов [15, 17] создается начальная популяция альтернативных решений решаемой задачи диспетчеризации.

Данная популяция эволюционирует с учетом моделей эволюции Ч. Дарвина и Ж.Б. Ламарка [18]. Далее производится выбор ЦФ и затем на ее основе производится оценка начальной популяции. Следующим этапом реализуется оператор репродукции (селекции). После этого формируются новые решения за счет выполнения различных операторов мутации (мутация, инверсия) [15, 17]. Данные, полученные после реализации операторов, передаются в блок анализа неперспективных решений. Здесь каждому решению присваивается определенная метка (перспективное, неперспективное, тривиальное и др.) с учетом ограничений и знаний ЛППР. Заметим, что такое ранжирование решений за счет структуризации позволяет динамически управлять поиском, что повышает эффективность работы эволюционного алгоритма в целом. Таким образом, в данной архитектуре появляется дополнительный инструмент для самоадаптации и настройки параметров поиска.

Далее результаты передаются в блок эволюционной адаптации. Данный блок оказывает непосредственное влияние на процесс перестройки текущей популяции и создания на ее основе новой популяции альтернативных решений. Блок внешней среды (ЛППР) позволяет производить выбор используемой на данном этапе модели эволюции и управлять всеми изменяемыми параметрами поиска. Процесс работы заканчивается при достижении определенного критерия окончания работы алгоритма. Таким критерием может быть время, число заданных генераций или получение набора квазиоптимальных решений.

3. Вычислительный эксперимент. Разработана программная среда для решения задачи диспетчеризации. При построении комплекса программ использовался пакет C#. Отладка и тестирование разработанных алгоритмов выполнялось на компьютере типа IBM PC с процессором ryzen 5 3600x с ОЗУ-16Гб. Проведен вычислительный эксперимент. Для определения эффективности разработанного эволюционного алгоритма были проведены исследования времени и качества решения (ресурсной меры M) для разного набора тестовых примеров, различающихся количеством блоков [19, 20]. В качестве алгоритмов, реализующих тестовые примеры, были выбраны известные алгоритмы последовательный (ПА) [9], начально-кольцевой (НК) [9], уровневый (УА) [21] и разработанные, и программно-реализованные авторами модифицированный эволюционный (ЭА). Результаты проведенных исследований представлены в табл. 1, а также на рис. 4 и 5.

Таблица 1

Сравнение результатов размещения тестовых примеров алгоритмами последовательным (ПА), начально-кольцевым (НК), уровневым (УА) и разработанным авторами модифицированным эволюционным (ЭА) по значению эвристической меры M в у. е. и времени t в сек

№ теста	Число блоков (k)	ПА		НК		УА		ЭА	
		M	t, с.						
1	4	0,7	7,3	0,7	7,5	0,7	7,5	0,65	7,6
2	8	0,67	7,7	0,68	7,8	0,69	7,8	0,62	7,9
3	12	0,63	8,1	0,65	8,1	0,65	8,2	0,6	8,2
4	16	0,61	8,4	0,62	8,5	0,62	8,7	0,59	8,5
5	20	0,61	8,7	0,61	8,7	0,61	8,9	0,58	8,9
6	24	0,59	8,8	0,58	8,9	0,57	9,1	0,57	9,1
7	28	0,58	9,1	0,57	9,2	0,57	9,3	0,55	9,4
8	32	0,57	9,3	0,56	9,4	0,57	9,5	0,54	9,6

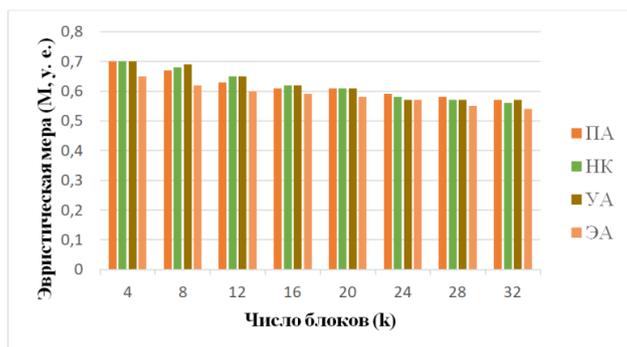


Рис. 4. Гистограмма сравнения значений эвристической меры алгоритмов размещения от числа размещаемых блоков

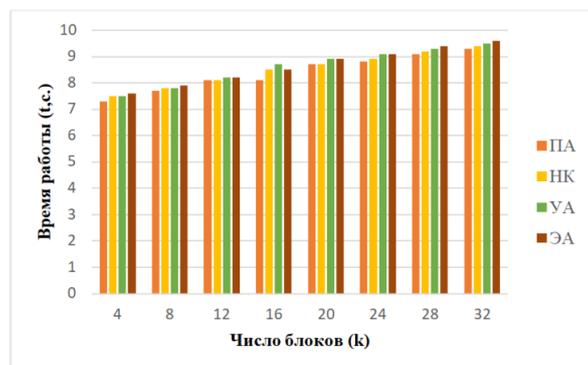


Рис. 5. Гистограмма сравнения времени работы алгоритмов размещения от числа размещаемых блоков вершин

В результате анализа представленной таблице и графиков зависимостей можно сделать вывод, что разработанный модифицированный эволюционный алгоритм при размещении блоков позволяет лучше производить размещение блоков в среднем на 5 %, чем известные алгоритмы при сопоставимом времени, что говорит об эффективности предложенного подхода.

Заключение. В работе рассмотрен подход к решению задач диспетчеризации массивами заявок. Для их эффективного решения предложена новая архитектура поиска, основанная на методах эволюционного моделирования. Для ее реализации авторами разработан модифицированный эволюционный алгоритм, позволяющий сократить время получения результата, частично решить проблему преждевременной сходимости алгоритма и получать наборы квазиоптимальных решений за полиномиальное время. Разработан программный модуль на языке C#. Проведен вычислительный эксперимент на тестовых примерах. Проведенные экспериментальные исследования, показали, что качество решений, полученных на основе разработанного эволюционного алгоритма, в среднем на 5 процентов превосходит результаты решений, полученные с использованием известных алгоритмов последовательного, начально-кольцевого и уровнявого при сопоставимом времени, что говорит об эффективности предложенного подхода.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Коваленко В.Н., Корягин Д.А. Грид: истоки, принципы и перспективы развития // Информационные технологии и вычислительные системы. – 2008. – № 4. – С. 38-50.
2. Васенин В.А., Шундеев А.С. Эволюция технологии Grid // Информационные технологии. – 2012. – № 1. – С. 2-9.
3. Барский А.Б. Алгоритмические, архитектурные и структурные методы организации управляющих процессов в виртуальном пространстве средств Grid-системы // Информационные технологии. – 2012. – № 5. – С. 2-6.
4. Коваленко В.Н., Семьякин Д.А. Методы и алгоритмы управления параллельными заданиями в гриде с ресурсами в форме кластеров // Вестник Южного научного центра РАН. – 2008. – Т. 4, № 3. – С. 23-34.
5. Каляев И.А., Левин И.И., Семерников Е.А., Шмойлов В.И. Реконфигурируемые мультиточечные вычислительные структуры / под общ. ред. И.А. Каляева. – Ростов-на-Дону: Изд-во ЮНЦ РАН, 2009.
6. Liu C., Baskiyar S. A general distributed scalable grid scheduler for independent tasks // J. Parallel Distrib. Comput. – 2009. – Vol. 69. – P. 307-314.
7. Саак А.Э. Локально-оптимальные ресурсные распределения // Информационные технологии. – 2011. – № 2. – С. 28-34.
8. Саак А.Э. Диспетчеризация в Grid-системах на основе однородной квадратичной типизации массивов заявок пользователей // Информационные технологии. – 2012. – № 4. – С. 32-36.
9. Саак А.Э. Полиномиальные алгоритмы распределения ресурсов в Grid-системах на основе квадратичной типизации массивов заявок // Информационные технологии. – 2013. – № 7. Приложение. – 32 с.
10. Korf R. Huang E. New Improvements in Optimal Rectangle Packing // In Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009). Pasadena, California, USA, July 11-17, 2009. – P. 511-516.
11. Саак А.Э. Диспетчеризация массивов заявок суммарной ресурсной меры, равной квадрату целого числа // Информационные технологии. – 2015. – Т. 21, № 9. – С. 675-679.
12. Емельянов В.В., Курейчик В.В., Курейчик В.М. Теория и практика эволюционного моделирования: монография. – М.: Физматлит, 2003.
13. Курейчик В.В., Курейчик В.М., Родзин С.И. Концепция эволюционных вычислений, инспирированных природными системами // Известия ЮФУ. Технические науки. – 2009. – № 4 (93). – С. 16-24.
14. Курейчик В.В., Курейчик В.М., Родзин С.И. Теория эволюционных вычислений: монография. – М.: Физматлит, 2012.
15. Курейчик В.В., Родзин С.И. О правилах представления решений в эволюционных алгоритмах // Известия ЮФУ. Технические науки. – 2010. – № 7 (108). – С. 13-21.
16. Карпенко А.П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой: учеб. пособие. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2014. – 446 с.
17. Курейчик В.М., Курейчик В.В. Об управлении на основе генетического поиска // Автоматика и телемеханика РАН. – 2001. – № 10. – С. 174-187.
18. Курейчик В.В., Курейчик В.М., Сороколетов П.В. Анализ и обзор моделей эволюции // Известия Российской академии наук. Теория и системы управления. – 2007. – № 5. – С. 114-126.
19. Korf R. Huang E. Optimal Rectangle Packing on Non-Square Benchmarks // In Proceedings of the twenty-fourth AAAI Conference on Artificial Intelligence (AAAI-10). Atlanta, Georgia, USA, July 11–15, 2010. – P. 83-88.
20. Саак А.Э. Сравнительный анализ полиномиальных алгоритмов диспетчеризации в Grid-системах // Информационные технологии. – 2012. – № 9. – С. 28-32.
21. Саак А.Э. Уровневые алгоритмы диспетчеризации массивами заявок кругового типа в Grid-системах // Известия ЮФУ. Технические науки. – 2015. – № 6 (167). – С. 223-231.

REFERENCES

1. Kovalenko V.N., Koryagin D.A. Grid: istoki, printsipy i perspektivy razvitiya [Grid: sources, principles and prospects of development], *Informatsionnye tekhnologii i vychislitel'nye sistemy* [Information technologies and computing systems], 2008, No. 4, pp. 38-50.

2. Vasenin V.A., Shundeev A.S. Evolyutsiya tekhnologii Grid [Evolution of Grid technology], *Informatsionnye tekhnologii* [Information Technologies], 2012, No. 1, pp. 2-9.
3. Barskiy A.B. Algoritmicheskie, arkhitekturnye i strukturnye metody organizatsii upravlyayushchikh protsessov v virtual'nom prostranstve sredstv Grid-sistemy [Algorithmic, architectural and structural methods of organizing control processes in the virtual space of Grid-system tools], *Informatsionnye tekhnologii* [Information Technologies], 2012, No. 5, pp. 2-6.
4. Kovalenko V.N., Semyachkin D.A. Metody i algoritmy upravleniya parallel'nymi zadaniyami v gride s resursami v forme klasterov [Methods and algorithms for managing parallel tasks in a grid with resources in the form of clusters], *Vestnik Yuzhnogo nauchnogo tsentra RAN* [Bulletin of the Southern Scientific Center of the Russian Academy of Sciences], 2008, Vol. 4, No. 3, pp. 23-34.
5. Kalyaev I.A., Levin I.I., Semernikov E.A., Shmoylov V.I. Rekonfiguriruemye mul'tikonveyernye vychislitel'nye struktury [Reconfigurable multiconveyor computing structures], under the general ed. of I.A. Kalyaeva. Rostov-on-Donu: Izd-vo YuNTS RAN, 2009.
6. Liu C., Baskiyar S. A general distributed scalable grid scheduler for independent tasks, *J. Parallel Distrib. Comput.*, 2009, Vol. 69, pp. 307-314.
7. Saak A.E. Lokal'no-optimal'nye resursnye raspredeleniya [Locally optimal resource distributions], *Informatsionnye tekhnologii* [Information technologies], 2011, No. 2, pp. 28-34.
8. Saak A.E. Dispetcherizatsiya v Grid-sistemakh na osnove odnorodnoy kvadrachnoy tipizatsii massivov zayavok pol'zovateley [Dispatching in Grid-systems based on homogeneous quadratic typing of user requests], *Informatsionnye tekhnologii* [Information Technologies], 2012, No. 4, pp. 32-36.
9. Saak A.E. Polinomial'nye algoritmy raspredeleniya resursov v Grid-sistemakh na osnove kvadrachnoy tipizatsii massivov zayavok [Polynomial algorithms for resource allocation in Grid systems based on quadratic typing of request arrays], *Informatsionnye tekhnologii* [Information Technologies], 2013, No. 7. Appendix, 32 p.
10. Korf R. Huang E. New Improvements in Optimal Rectangle Packing. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009). Pasadena, California, USA, July 11-17, 2009*, pp. 511-516.
11. Saak A.E. Dispetchepizatsiya massivov zayavok summapnoy pesupsnoy mery, pavnoy kvadpatu tselogo chisla [Dispatching of request arrays with a total resource measure equal to the square of an integer], *Informatsionnye tekhnologii* [Information Technologies], 2015, Vol. 21, No. 9, pp. 675-679.
12. Emel'yanov V.V., Kureychik V.V., Kureychik V.M. Teoriya i praktika evolyutsionnogo modelirovaniya: monografiya [Theory and practice of evolutionary modeling: monograph]. Moscow: Fizmatlit, 2003.
13. Kureychik V.V., Kureychik V.M., Rodzin S.I. Kontseptsiya evolyutsionnykh vychisleniy, inspirirovannykh prirodnyimi sistemami [The concept of evolutionary calculations inspired by natural systems], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2009, No. 4 (93), pp. 16-24.
14. Kureychik V.V., Kureychik V.M., Rodzin S.I. Teoriya evolyutsionnykh vychisleniy: monografiya [Theory of evolutionary calculations: monograph]. Moscow: Fizmatlit, 2012.
15. Kureychik V.V., Rodzin S.I. O pravilakh predstavleniya resheniy v evolyutsionnykh algoritmakh [On the rules for the representation of solutions in evolutionary algorithms], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2010, No. 7 (108), pp. 13-21.
16. Karpenko A.P. Sovremennyye algoritmy poiskovoy optimizatsii. Algoritmy, vdokhnovlennyye prirodoy: ucheb. posobie [Modern search engine optimization algorithms. Algorithms inspired by nature: a textbook]. Moscow: Izd-vo MGTU im. N.E. Bauman, 2014, 446 p.
17. Kureychik V.M., Kureychik V.V. Ob upravlenii na osnove geneticheskogo poiska [About management on the basis of genetic search], *Avtomatika i telemekhanika RAN* [Avtomatika i telemekhanika RAS], 2001, No. 10, pp. 174-187.
18. Kureychik V.V. Kureychik V.M., Sorokoletov P.V. Analiz i obzor modeley evolyutsii [Analysis and review of models of evolution], *Izvestiya Rossiyskoy akademii nauk. Teoriya i sistemy upravleniya* [Izvestiya Rossiyskoy akademii nauk. Theory and control systems], 2007, No. 5, pp. 114-126.

19. Korf R. Huang E. Optimal Rectangle Packing on Non- Square Benchmarks. *In Proceedings of the twenty-fourth AAAI Conference on Artificial Intelligence (AAAI-10). Atlanta, Georgia, USA, July 11–15, 2010*, pp. 83-88.
20. Saak A.E. Sravnitel'nyy analiz polinomial'nykh algoritmov dispetcherizatsii v Grid-sistemakh [Comparative analysis of polynomial dispatching algorithms in Grid systems], *Informatsionnye tekhnologii* [Information Technologies], 2012, No. 9, pp. 28-32.
21. Saak A.E. Urovnevye algoritmy dispetcherizatsii massivami zayavok krugovogo tipa v Grid-sistemakh [Level algorithms for dispatching circular-type request arrays in Grid-systems], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2015, No. 6 (167), pp. 223-231.

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

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

Саак Андрей Эрнестович – e-mail: saak@sfned.ru; тел.: 88634393140; кафедра государственного и муниципального управления; д.т.н.; доцент; зав. кафедрой ГиМУ.

Курейчик Владимир Владимирович – ООО «Газпром подземремонт Уренгой»; e-mail: Kureichik@yandex.ru; г. Санкт-Петербург, Россия; тел.: +78126096700; отдел автоматизации систем управления; к.т.н.; ведущий инженер.

Kureichik Vladimir Victorovich – Southern Federal University; e-mail: vkur@sfned.ru; Taganrog, Russia; phone: +78634371651; the department of computer aided design; head of CAD department; dr. of eng. sc.; professor.

Saak Andrey Ernestovich – e-mail: aesaak@sfned.ru; Taganrog, Russia; phone: +78634393373; the department of state and municipal management, head of department.

Kureichik Vladimir Vladimirovich – «Gazprom podzemremont Urengoi» company; e-mail: Kureichik@yandex.ru; St. Peterburg, Russia; phone: +78126096700; Control systems automation department; cand. of eng. sc.; lead engineer.

УДК 621.315.611

DOI 10.18522/2311-3103-2021-2-59-73

Н.К. Полуянович, М.Н. Дубяго

АЛГОРИТМ ОБУЧЕНИЯ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ФАКТОРНОГО ПРОГНОЗИРОВАНИЯ РЕСУРСА ИЗОЛЯЦИОННЫХ МАТЕРИАЛОВ СИЛОВЫХ КАБЕЛЬНЫХ ЛИНИЙ

Статья посвящена исследованиям термофлуктуационных процессов в соответствии с теорией теплопроводности для решения задач факторного прогнозирования остаточного ресурса изоляционных материалов на основе неразрушающего температурного метода. Обоснована актуальность задачи разработки алгоритма для прогнозирования температуры жил СКЛ в режиме реального времени на основе данных системы температурного мониторинга, с учетом изменения токовой нагрузки линии и внешних условий теплоотвода. Экспериментальным методом выявлены типы искусственных нейронных сетей, их архитектура и состав, которые обеспечивают максимальную точность прогнозирования при минимальном наборе значимых факторов. Разработана нейросеть для определения температурного режима токоведущей жилы силового кабеля. Определен минимальный набор значимых факторов и размерность входного обучающего вектора, который обеспечивает универсальность нейросетевого метода прогнозирования. Разработана нейросеть для определения температурного режима токоведущей жилы заключается в диагностике и прогнозировании ресурса электроизоляции (ЭИ) силового кабеля. Модель позволяет оценивать

текущее состояние изоляции и прогнозировать остаточный ресурс СКЛ. Проведен сравнительный анализ экспериментальных и расчетных характеристик алгоритмов обучения искусственной нейронной. Установлено, что предлагаемый алгоритм искусственной нейронной сети может быть использована для составления прогноза температурного режима токоведущей жилы, на три часа вперед с точностью до 2,5 % от фактического значения температуры жилы. Основная область применения разработанной нейросети для определения температурного режима токоведущей жилы заключается в диагностике и прогнозировании ресурса электроизоляции (ЭИ) силового кабеля. Разработка интеллектуальной системы прогнозирования (ИНС) температуры жилы СКЛ способствует планированию режимов работы электросети с целью повышения надежности и энергоэффективности их взаимодействия с объединенной энергосистемой.

Факторное прогнозирование; изоляционные материалы; термофлуктуационные процессы; алгоритм Левенберга-Марквардта; архитектура нейронной сети; надежность систем энергоснабжения.

N.K. Poluyanovich, M. N. Dubyago

ALGORITHM FOR TRAINING THE ARTIFICIAL NEURAL NETWORK OF FACTOR PREDICTING THE POWER CABLE LINES INSULATING MATERIALS LIFE

The article is devoted to the research of thermofluctual processes in accordance with the theory of thermal conductivity for solving the problems of factor prediction of the residual life of insulating materials based on the non-destructive temperature method. The relevance of the task of developing algorithm for predicting the temperature of SCL cores in real time based on the data of the temperature monitoring system, taking into account the change in the current load of the line and external heat removal conditions, is justified. The experimental method revealed the types of artificial neural networks, their architecture and composition, which provide maximum prediction accuracy with a minimum set of significant factors. A neural network has been developed to determine the temperature regime of the current-carrying core of the power cable. The minimum set of significant factors and the dimension of the input training vector is determined, which provides the versatility of the neural network prediction method. A neural network for determining the temperature mode of the current-carrying core is designed to diagnose and predict the electrical insulation (EI) life of a power cable. The model allows assessing the current insulation state and predicting the residual resource of the SCL. Comparative analysis of experimental and calculated characteristics of learning algorithms of isostic neural is carried out. It has been found that the proposed algorithm of artificial neural network can be used for prediction of current-carrying core temperature mode, three hours in advance with accuracy up to 2.5 % of actual value of core temperature. The main field of application of the developed neural network for determining the temperature mode of the current-carrying core is in diagnostics and predicting the electrical insulation (EI) life of the power cable. The development of an intelligent system for predicting the temperature of the LCS core contributes to the planning of the operation modes of the electric network in order to increase the reliability and energy efficiency of their interaction with the integrated energy system.

Factor prediction; insulating materials; thermoflux processes; Levenberg-Marquardt algorithm rhythm; neural network architecture; reliability of power supply systems.

Введение. Силовые кабели напряжением 6 (10) кВ, в системе электроснабжения потребителей городских электрических сетей являются наиболее протяженными элементами. Устойчивое функционирование электросетевого комплекса невозможно без надежной и качественной работы распределительных электрических сетей 6–10/0,4 кВ, которые являются завершающим звеном в системе обеспечения потребителей электрической энергией. Исследования причин отказов в кабельных линиях (КЛ) вызвана результатами анализа аварийности в местах с населением более 100 тыс. жителей, где повреждения КЛ 6 (10) кВ являются причиной 80–90 % всех отключений. Во время эксплуатации на изоляцию КЛ 6 (10) кВ воз-

действует целая группа факторов: электрические, тепловые, климатические, механические, коррозия [1]. Значительная часть работающих КЛ выработала свой нормативный ресурс, а результаты эксплуатации КЛ свидетельствуют о том, что основной причиной его отказов является снижение электрической прочности изоляции.

Важной характеристикой КЛ систем электроснабжения является нагрузочная способность, определяемая тепловым режимом, по которой рассчитывается площадь поперечного сечения и их пропускная способность [2, 3]. Расчёт пропускной способности необходим и осуществляется при проектировании КЛ. Допустимая токовая нагрузка КЛ может быть получена из формулы превышения температуры жилы над температурой окружающей среды [4, 5]. Однако использование даже самых точных методик не позволяет учесть на практике все нюансы, которые весьма существенно влияют на температурный режим работы КЛ [6]. Одним из решений указанной проблемы является мониторинг температуры кабеля во время работы. Таким образом, пропускная способность линий электропередачи зависит от температуры проводника.

Прогнозирование пропускной способности является важной областью исследования в электроэнергетике. Оно необходимо для решения практически всего спектра задач текущего планирования и оперативного управления режимами функционирования электроэнергетической системы (ЭЭС) [7, 8]. На его основе рассчитываются исходные и оптимальные электрические нагрузки ЭЭС, оценивается их надежность и безопасность электроснабжения. Точность прогноза нагрузок влияет на экономичность загрузки генерирующего оборудования, и, следовательно, на стоимость электроэнергии [9].

Обзор литературных источников показал, прогнозирование проводится в различных временных диапазонах, так как не существует точной классификации временных периодов прогнозирования [10]. В работе [7], например, временные диапазоны прогнозирования классифицируются следующим образом: в пределах текущих суток – оперативный прогноз; вперед на сутки-неделю-месяц – краткосрочный; на месяц-квартал-год – долгосрочный. К настоящему времени разработано большое число методов и моделей прогнозирования электрических нагрузок как традиционных, так и нетрадиционных. Традиционные статистические модели могут быть условно разделены на регрессионные модели и модели на основе временных рядов. С развитием теории новых информационных технологий в последнее десятилетие было предложено решение задачи прогнозирования нагрузки нетрадиционными методами, а именно, с использованием моделей на основе экспертных систем и искусственных нейронных сетей [4, 11–14]. Предпочтение таких моделей традиционным обусловлено тем, что не требуется построения модели объекта, не теряется работоспособность при неполной входной информации. Они обладают устойчивостью к помехам, имеют высокое быстродействие. Текущий анализ диагностических систем подтверждает, что имеется объективная научно-техническая проблема создания комплексных систем диагностирования, построенных на разносторонних принципах, дающих высокий уровень достоверности постановки диагноза и прогнозирования технического состояния КЛ.

В работе представлены результаты исследований прогнозирования пропускной способности кабельных электрических сетей с применением искусственных нейронных сетей как одних из наиболее совершенных и перспективных направлений решения данной задачи. Теоретический подход к рассмотренным вопросам сочетается с результатами экспериментальных исследований, проведенных с использованием примерного графика нагрузки энергосистемы.

Цели и задачи исследования. Температура высоковольтной кабельной линии является важным эксплуатационным и диагностическим параметром [2, 5, 15].

С одной стороны, она характеризует рабочую нагрузку КЛ, а с другой является признаком наличия и развития дефектов в изоляции. С повышением температуры существенно возрастает диффузия воды в полимеры, что снижает их механическую и электрическую прочность [16]. В процессе эксплуатации изоляция СКЛ подвергаются тепловым, электрическим, химическим, механическим и другим видам воздействий, в результате чего происходит изменение её электрических свойств и, как следствие, некоторые изменения претерпевают и технические характеристики изоляционных материалов [9]. Происходит старение изоляции, которое выражается в ухудшении её электрофизических характеристик. Возникает проблема определения механизма воздействия на изоляцию факторов, обусловленных действием тока нагрузки КЛ.

Перспективным направлением является создание инфраструктуры диагностирования неисправностей СКЛ, основанной на использовании искусственных нейронных сетей. Искусственные нейронные сети (ИНС) в задачах диагностирования и прогнозирования остаточного ресурса изоляционных материалов могут служить в роли подсистемы выборки и принятия решений, а затем отдавать полученные данные другим подсистемам управления.

Применение интеллектуальной системы управления приведет к повышению общей надежности энергосети и повышению энергоэффективности функционирования энергетического сектора, уменьшению влияния человеческого фактора и снижению аварийных ситуаций.

Задачей исследования являются разработка алгоритма интеллектуальной системы прогнозирования температуры жилы СКЛ для планирования режимов работы электросети с целью повышения надежности и энергоэффективности их взаимодействия с объединенной энергосистемой;

Целью исследования является повышение надежности и энергоэффективности распределенных энергосистем за счёт применения новых неразрушающих методов и алгоритмов мониторинга и диагностики ресурса СКЛ на базе искусственных нейронных сетей для краткосрочного и среднесрочного прогнозирования температурных режимов токоведущей жилы силовых кабельных линий.

Синтез конфигурации ИНС. Существует несколько десятков различных нейросетевых архитектур, причем эффективность многих из них доказана математически. Для решения задач прогнозирования применяют следующие типы ИНС: линейную сеть, многослойный перцептрон, сеть на радиальных базисных функциях, обобщенно-регрессионную сеть. Достоинством ИНС является возможность факторного прогнозирования. При этом сеть имеет X входов, каждый из которых соответствует какому-либо фактору, оказывающему влияние на прогнозный параметр [5, 15, 16]. Подача на вход сети вместе предысторией температуры токоведущей жилы значимых факторов (величина тока жилы, температуры кабеля и др.) способствует повышению точности прогноза. При этом сеть сама определяет зависимости между факторами, корректируя свои параметры при обучении. Таким образом могут быть смоделированы достаточно сложные регрессионные зависимости.

Для эффективного решения задачи факторного прогнозирования необходимо производить выбор подходящей структуры для искусственной нейронной сети. Анализ методов искусственного интеллекта, предназначенных для решения задач факторного прогнозирования, показал, что многослойный перцептрон может формировать на выходе произвольную многомерную функцию при соответствующем выборе количества слоев, диапазона изменения сигналов и параметров нейронов [17]. Для решения задачи с помощью ИНС необходимо выбрать конфигурацию нейронной сети и сформировать окружение, в котором будет работать данная нейронная сеть. Под формированием окружения понимается реализация следующих этапов:

- ◆ создание обучающего задачника;
- ◆ обучение нейронной сети;
- ◆ тестирование ИНС.

Для определения входных переменных нейронной сети при решении задач прогнозирования для прогнозирования температуры жилы СКЛ была использована модель, описывающая изменения во времени фактических значений температур, которая в общем виде представляется нелинейной функцией:

$$Y_t = f(X_t - n, T_t - n, N_t) + \varepsilon_t,$$

где X_t – фактическое значение температуры жилы в момент времени t ; t – текущее время; $X_t - n$ – предшествующие наблюдения температуры жилы, $T_t - n$ – предшествующие наблюдения температуры окружающей среды; n – индекс ретроспективы данных; I_t – величина действующего тока ($I_{\min} < I < I_{\max}$); ε_t – случайная составляющая, представляющая ненаблюдаемые факторы, влияющие на температуры жилы.

Интеллектуальная система факторного прогнозирования. Многие нелинейные задачи не поддаются строгой формализации традиционными математическими методами [16]. В таких условиях, когда решение задачи аналитически в общем виде невозможно, оправдан нейросетевой подход, позволяющий обеспечить достаточно высокое качество выполнения задачи. В ряде работ, касающихся аппроксимации нелинейных функций, заложен математический базис нейросетевой теории, определяющий универсальные аппроксимирующие свойства нейронных сетей. В большинстве случаев аналитические модели диагностирования это нелинейные соотношения, затрудняющие формирование модели объекта диагностирования по моделям составляющих компонент. Формально высказывание об универсальных аппроксимационных свойствах нелинейности представляется в виде:

$$F(x_1, x_2, \dots, x_n) = \sum_q^{2n+1} h_p \left[\sum_p^n \Phi_q^p(x_p) \right], \quad (1)$$

Оно утверждает, что с помощью линейных операций и каскадного соединения можно из произвольных нелинейных элементов получить любой требуемый результат с заранее заданной точностью. Здесь h_p – непрерывная функция; $\Phi_q^p(x_p)$ – функция, зависящая от F .

Выбор алгоритма обучения сети. Для решения задачи факторного прогнозирования ресурса кабельных линий необходимо выбрать алгоритм. Для обучения ИНС были рассмотрены различные алгоритмы [18]. Проводится сравнение алгоритмов обучения ИНС таких, как алгоритм обучения ИНС основанный на регуляризации Байеса. Суть данного алгоритма заключается в том, что изменение весов синаптических функций [19] прекращается при достижении наименьшей среднеквадратической ошибки. Обучение по данному алгоритму занимает больше времени по сравнению с алгоритмом Левенберга-Марквардта, но при этом достигается минимальная среднеквадратическая ошибка (порядка $2,02E-05$). Также при прогнозировании временных рядов, ошибка на тестовой выборке становится меньше, чем на обучающей.

Метод Ньютона-Гаусса – это итерационный численный метод нахождения решения задачи наименьших квадратов. Является разновидностью метода Ньютона. В общих чертах, этот метод использует матрицу Якобиана J производных первого порядка функции F для нахождения вектора x значений параметра, который минимизирует остаточные суммы квадратов (сумму квадратных отклонений предсказанных значений от наблюдаемых). В стандартном итерационном методе Ньютона на каждой итерации требуется вычисление и обращение матрицы Гесса, что

зачастую является достаточно сложным процессом. В методе Ньютона-Гаусса подобная необходимость отпадает, причем скорость сходимости также может достигать квадратичной, хотя вторые производные и не учитываются. Также данный метод прост в реализации и присутствует в большинстве программных пакетов по прикладной математике. Тем не менее, в методе Ньютона-Гаусса часто встречается ряд проблем в ситуации, когда член второго порядка значителен по своей величине, что приводит к некорректной работе и медленной сходимости.

Классический алгоритм обратного распространения ошибки использует метод градиентного спуска для подстройки весов после каждого обучающего примера и, таким образом, обеспечивается «движение» в многомерном пространстве весов [20]. Градиентный спуск – метод нахождения локального экстремума (минимума или максимума) функции с помощью движения вдоль градиента. Наиболее простой в реализации из всех методов локальной оптимизации. Имеет довольно слабые условия сходимости, но при этом скорость сходимости достаточно мала (линейна). Другими словами, данный алгоритм при движении по поверхности ошибки данный алгоритм учитывает лишь наклон поверхности ошибки (в точке нахождения). Если уклон сильный, то алгоритм делает большой шаг по поверхности ошибки. При слабом уклоне алгоритм движется более малыми шагами. По этой причине алгоритм может как бесконечно долго сходиться, так и застрять в локальном минимуме.

Кроме функции потерь градиентный спуск также требует градиент, который является dJ/dw (производная функции потерь относительно одного веса, выполненная для всех весов). dJ/dw зависит от вашего выбора функции потерь. Наиболее распространена функция потерь среднеквадратичной ошибки.

$$J(\theta) = \frac{1}{m} \sum_{i=1}^M (h_{\theta}(x^{(i)}) - y^{(i)})^2. \quad (2)$$

Производная этой функции относительно любого веса (эта формула показывает вычисление градиента для линейной регрессии):

$$\frac{\partial}{\partial \theta_j} MSE(\theta) = \frac{2}{m} \sum_{i=1}^m (\theta^T x^{(i)} - y^{(i)}) x_j^{(i)}. \quad (3)$$

Выбор функции повлияет на вычисление градиента каждого веса.

Блок-схема алгоритма обучения искусственной нейронной сети представлена на рис. 1

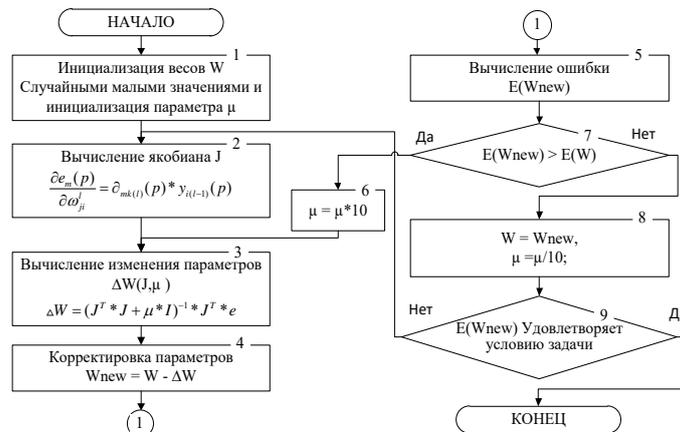


Рис. 1. Блок-схема алгоритма обучения нейронной сети

Алгоритм обратного распространения ошибки учитывает структуру сети и позволяет отказаться от многократного вычисления откликов нейронов на каждом шаге градиентного метода оптимизации.

Использование метода обратного распространения ошибки для корректировки весовых коэффициентов позволяет отказаться от больших объемов вычислений, необходимых при прямом вычислении градиента, благодаря учету структуры нейронной сети и отказу от многократного вычисления откликов нейронов на каждом шаге градиентного метода оптимизации.

Таким образом, за конечное число шагов веса всех нейронов сети подстраиваются так, чтобы обеспечить минимальную ошибку обучения. Недостатком данного алгоритма является большая вероятность попадания в локальный минимум, медленная работа, а также заранее неизвестное число циклов обучения [21].

Алгоритм Левенберга-Марквардта. Существуют ускоренные алгоритмы обучения многослойных нейронных сетей, базирующиеся на вычислении матрицы Гессе вторых частных производных функции ошибки по настраиваемым параметрам. К таким алгоритмам относятся квазиньютоновские алгоритмы, алгоритм Левенберга-Марквардта и другие. Алгоритм Левенберга – Марквардта – метод оптимизации, направленный на решение задач о наименьших квадратах. Является альтернативой методу Гаусса-Ньютона [22]. Предполагается, что в качестве критерия оптимизации используется среднеквадратичная ошибка модели на обучающей выборке. Обучение сети прерывается в тот момент, когда перестает происходить ее уменьшение. Алгоритм заключается в последовательном приближении заданных начальных значений параметров к искомому локальному оптимуму [23]. Алгоритм отличается от метода сопряженных градиентов тем, что использует матрицу Якоби модели, а не градиент вектора параметров. От алгоритма Гаусса-Ньютона этот алгоритм отличается тем, что использует параметр регуляризации [24]. К достоинствам данного алгоритма обучения ИНС можно отнести быстроту обучения и достаточно низкую среднеквадратическую ошибку, которая составляет порядка $3,35E-05$.

В работе [25] описан алгоритм и программная реализация адаптивного метода обучения искусственных нейронных сетей, а в [26] алгоритм и программная реализация гибридного метода обучения искусственных нейронных сетей.

Таким образом, для решения поставленной задачи прогнозирования термических процессов в СКЛ выбор сделан в пользу алгоритма Левенберга-Марквардта, т. к. данный алгоритм обеспечивает необходимую точность прогноза, а также обладает меньшим количеством необходимых вычислительных операций по сравнению с традиционными методами.

Описание алгоритма и подстройки весовых коэффициентов. Для решения задачи факторного прогнозирования термических процессов кабельных линий выбран алгоритм обратного распространения ошибки, рис 2. На первой стадии происходит инициализация весов малыми случайными значениями (шаг 1). Обучение предполагается управляемым, поскольку с каждым входным образцом из обучающей выборки связывается целевой выходной образец обучающей выборки (шаг 2). Обучение происходит до тех пор, пока изменение усредненной квадратичной ошибки не окажется меньше некоторого допустимого значения при переходе от одной эпохи к следующей (шаги 3, 4). Если в процессе обучения наступает момент, когда ошибка попадает в рамки допустимых значений (шаги 5-8), говорят, что наблюдается сходимость.

$$\Delta w_{ij}(n+1) = \eta \delta_j o_i + a \Delta w_{ij}(n).$$

Другим критерием окончания обучения можно считать наступление момента, когда выход для каждого обучающего образца оказывается в рамках допустимого отклонения от соответствующего целевого выходного образца. Чтобы уменьшить

вероятность того, что изменения весов приобретут осциллирующий характер, вводится инерционный член (а), добавляемый в пропорции, соответствующей предыдущему изменению веса (шаг 8):

Таким образом изменение веса на шаге $n + 1$ оказывается зависящим от веса на шаге n .

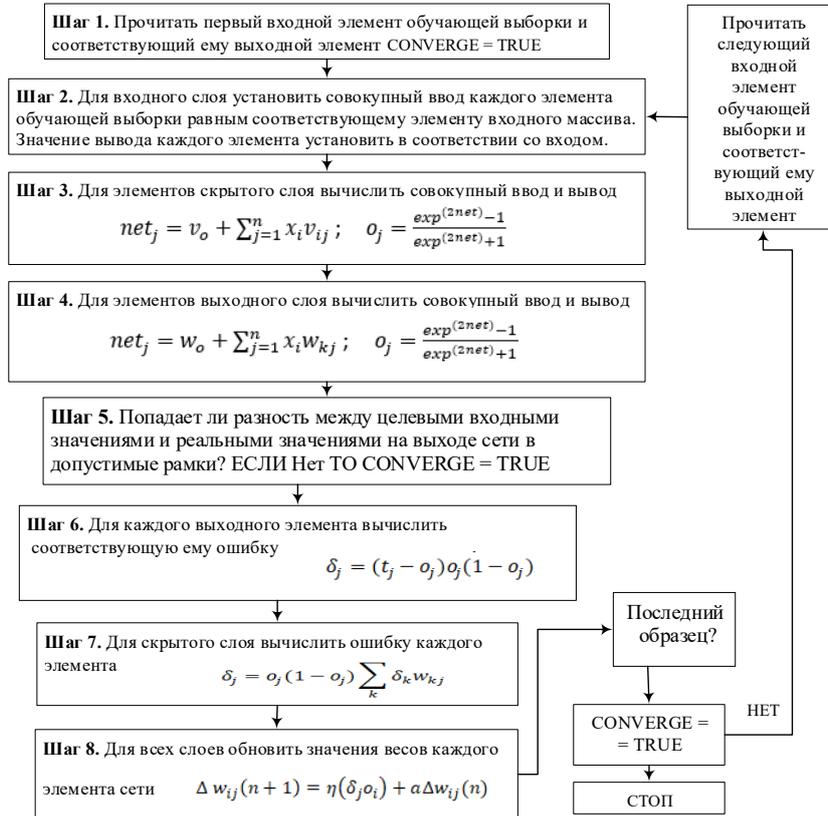


Рис. 2. Алгоритм определения и подстройки весовых коэффициентов $V(i,j)$, $W(j,k)$ в нейронных связях

Метод обратного распространения ошибки. Нейронная сеть имеет n входов и n нейронов распределительного слоя, L нейронов скрытого слоя и m выходных нейронов [3]. Используются две матрицы весов: скрытого слоя v размером $i \times j$ и выходного слоя w размером $j \times k$. Кроме этого, с каждым слоем нейронов связан массив порогов: Q – для скрытого слоя, T – для выходного. Эти данные представляют собой знания сети, настраиваемые в процессе обучения и определяющие ее поведение [2].

Прямой ход алгоритма обратного распространения ошибки в процессе обучения описывается в виде:

$$g_j = f\left(\sum_{i=1}^n v_{ij} x_i + Q_j\right); \quad y_k = f\left(\sum_{j=1}^h w_{kj} g_j + T_k\right), \quad (4)$$

где g_j – выходной сигнал нейронов скрытого слоя, y_k – выходной сигнал нейронов выходного слоя.

Обратный ход алгоритма обратного распространения ошибки запишется в следующем виде:

Изменение весовых коэффициентов и порогов нейронной сети происходит по следующим формулам:

$$w_{jk}(t+1) = w_{jk}(t) + a \frac{\partial E}{\partial w_{jk}(t)}; \quad T_k(t+1) = T_k(t) + a \frac{\partial E}{\partial T_k(t)}. \quad (5)$$

Мерой ошибки в обучающих алгоритмах выбирают среднеквадратичную ошибку E . Она определяется как сумма квадратов между экспериментальным значением температур $-y_k^r$ и значением полученным нейросетью $-y_k$, для каждого примера $X(n,m)$.

$$E = \frac{1}{2} \sum_{k=1}^m (y_k^r - y_k)^2, \quad (6)$$

где y_k^r – экспериментальные значения температур; y_k – значения, полученные на сети.

Ошибка k -го нейрона выходного слоя определяется как:

$$d_k = \frac{\partial E}{\partial y_k} = y_k^r - y_k. \quad (7)$$

Выразим производные из формул (5) через легко вычисляемые величины. Определим взвешенную сумму, аргумент функции активации как:

$$S_k = \sum_{j=1}^h w_{jk} g_j + T_k. \quad (8)$$

Из первого соотношения (5), $\partial E / \partial w_{jk}$ можно представить как:

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial S_k} \frac{\partial S_k}{\partial w_{jk}}, \quad (9)$$

где $\partial E / \partial y_k = d_k = y_k^r - y_k$, – ошибка k -го нейрона; $\partial y_k / \partial S_k = f(S_k)$, – производная функция активации; $\partial S_k / \partial w_{jk} = g_j$, – значение j -го нейрона предыдущего слоя.

Получаем

$$\frac{\partial E}{\partial w_{jk}} = d_k f(S_k) g_j. \quad (10)$$

Аналогично (9), с учетом того, что $\partial S_k / \partial T_k = 1$, получаем

$$\frac{\partial S_k}{\partial T_k} = \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial S_k} \frac{\partial S_k}{\partial T_k} = d_k f(S_k). \quad (11)$$

Корректировка весов и пороги скрытого слоя производится также по выражениям (5), с учетом (10). Определение ошибки нейрона скрытого слоя является трудной задачей.

Явное вычисление ошибки по формуле, аналогичной (7), невозможно. Ее расчет возможен через ошибки нейронов выходного слоя, что и обусловило название алгоритма обратного распространения ошибки:

$$e_j = \frac{\partial E}{\partial g_j} = \sum_{k=1}^m \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial S_k} \frac{\partial S_k}{\partial g_j} = \sum_{k=1}^m d_k f(S_k) w_{jk}. \quad (12)$$

В результате работы программы выдается значение ошибки (рис. 3). Анализ результатов обработки средней ошибки обучения, одного из образцов кабеля, показал уменьшение среднего значения ошибки на 72 %, до уровня 0,4.

Матрица весовых коэффициентов ($E\delta.in$) представляет собой значения весов между входными нейронами и скрытым слоем нейронов, и матрица весовых коэффициентов ($E\delta.out$), которая представляет из себя значения весов между нейронами скрытого слоя и выходным нейроном, представлены ниже.

$$E\delta.in = \begin{bmatrix} V_{11} & V_{12} & \dots & V_{1j} \\ V_{21} & V_{22} & \dots & V_{2j} \\ V_{31} & V_{32} & \dots & V_{ij} \end{bmatrix} \quad E\delta.out = [W_{11} \quad W_{12} \quad \dots \quad W_{jk}]$$

Для корректировки и минимизации среднеквадратичной ошибки необходимо произвести подбор значений в матрице весовых значений, которая представляет из себя значения весов между входными нейронами и скрытым слоем нейронов. Количество строк определяет число входных переменных, а количество столбцов определяет число нейронов в скрытом слое.

Для корректировки и минимизации среднеквадратичной ошибки необходимо произвести подстройку матрицы весов $\xi\delta$ с помощью минимизации среднеквадратичной ошибки E_δ выполняется с помощью итерирования по эпохам обучения в соответствии с алгоритмом обратного распространения ошибки по формуле [3]:

$$\xi_{\delta+1} = \xi_\delta - \mu \left(\frac{\partial E_\delta}{\partial \xi_\delta} \right) + \nu (\xi_\delta + \xi_{\delta-1}), \quad (13)$$

где μ , ν – параметры алгоритма, определяющие скорость и устойчивость итерационного процесса обучения. Величина Градиента – $\partial E_\delta / \partial \xi_\delta$, при нахождении глобального минимума гиперповерхности ошибки E_δ определяет направление спуска (метод градиентного спуска) на пути к глобальному минимуму, результатом которого будут новые очередные значения коэффициенты матрицы $\xi_{\delta+1}$. График ошибки прогнозируемой температуры жилы представлен на рис. 3.

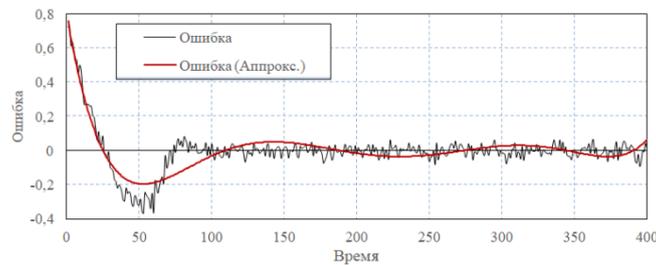


Рис. 3. График ошибки прогнозируемой температуры жилы

В результате работы программы выдается значение ошибки. Анализ показал уменьшение ошибки уже на 27 шаге в 6,7 раз (на 85 %).

Таким образом, алгоритм обратного распространения ошибки учитывает структуру сети и позволяет отказаться от многократного вычисления откликов нейронов на каждом шаге градиентного метода оптимизации. Использование метода обратного распространения ошибки для корректировки весовых коэффициентов позволяет отказаться от больших объемов вычислений, необходимых при прямом вычислении градиента, благодаря учету структуры нейронной сети и отказу от многократного вычисления откликов нейронов на каждом шаге градиентного метода оптимизации.

Экспериментальные исследования. Цель исследования заключается в нахождении таких значений весовых коэффициентов, при которых ошибка функционирования сети была бы минимальной. На каждой эпохе на вход сети поочередно подается все обучающие наблюдения $X(n, m)$, выходные значения сети сравниваются с целевыми значениям $(t_{j,i})$, и вычисляется ошибка. Значение ошибки, а также градиента поверхности ошибок используются для корректировки весов $V(i,j)$, $W(j,k)$, после чего все действия повторяются. После расчета градиентных спусков в пространстве весов W_{ji} и W_{kj} и их корректировки вычисляется суммарная ошибка по всем образцам и производится проверка выполнения условия точности расчетов. В случае выполнения условия точности расчетов процесс обучения заканчивается, в противном случае – процесс обучения повторяется. Опасность метода кроется в том, что таким образом мы минимизируем ошибку на обучающем множестве, и это не то же самое что минимизировать «настоящую» ошибку на поверхности ошибок. Сильнее всего это различие сказывается в проблеме переобучения, или излишней подгонки. Важно отметить, что вся информация, которую нейронная сеть имеет о задаче прогнозирования, содержится в наборе примеров. Поэтому качество обучения нейронной сети напрямую зависит от количества примеров в обучающей выборке, а также от того, насколько полно эти примеры описывают данную задачу. В работе для полноценной тренировки нейронной сети использовалось по каждому исследуемому образцу двадцать примеров в каждом по четыреста значений.

Обучение происходит следующим образом. Имеется некоторая база данных (задачник), содержащая примеры (величины тока жилы СКЛ ($I_{ж}$), температуры внешней изоляции $\theta_{з.о}$ и температуры окружающей среды $\theta_{о.с}$, рис. 4). Предъявляя их на вход нейронной сети, получаем от нее некоторый ответ, не обязательно верный. Известен и верный (желаемый) ответ. Вычисляя разность между желаемым и реальным ответами сети, получаем вектор ошибки.

Проведен сравнительный анализ точности обучения при различных алгоритмах обучения, табл. 1.

Таблица 1

Обучение НС

Алгоритм обучения нейронной сети	Средняя ошибка прогноза, при 80% выборки	
	ε , °C	ε , %
Ливенберга-Марквардта	0.87	2.2%
Байесова регуляризация	1.6	4.21%
Градиентного спуска	2.13	5.6%

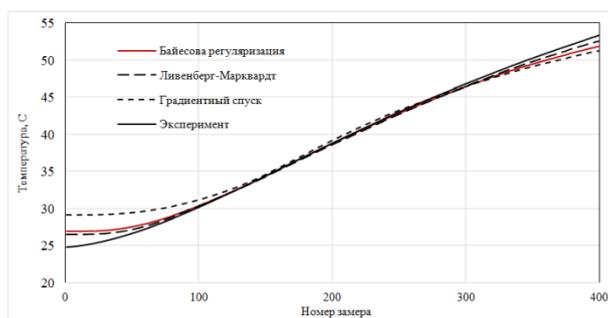


Рис. 4. Сравнение графиков прогнозирования температуры жилы в зависимости от алгоритма обучения

Заключение. Показано, что из рассмотренных алгоритмов обучения НС для решения задачи прогнозирования ресурса СКЛ, самую высокую точность имеет Ливенберга-Марквардта. Оптимизация по минимальной ошибке. Установлено, что минимум ошибки прогнозирования термических процессов в силовых кабельных сетях является НС с функцией активации: logsig – в скрытом слое и purelin – в выходном слое. Для решения задачи прогнозирования ресурса СКЛ была выбрана сеть с прямым распространением данных и обратного распространения ошибки, т.к. сети такого типа в совокупности с активационной функцией являются в некоторой степени универсальной структурой для многих задач аппроксимации, приближения и прогнозирования. Сеть обучается каждый раз «с нуля», весовые коэффициенты полученные при первичном обучении сети не используются для повторного обучения. Использование метода обратного распространения ошибки для корректировки весовых коэффициентов позволяет отказаться от больших объемов вычислений, необходимых при прямом вычислении градиента, благодаря учету структуры нейронной сети и отказу от многократного вычисления откликов нейронов на каждом шаге градиентного метода оптимизации.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Anders G.J. Rating of Cables on Riser Poles, in Trays, in Tunnels and Shafts - a Review // IEEE Transactions on Power Delivery. – 1996. – Vol. 11, No. 1.
2. Полуянович Н.К., Азаров Н.В., Огреничев А.В., Дубяго М.Н., Шурыкин А.А. Синтез нейросетевой модели для прогнозирования термических процессов электроизоляционных материалов силовых кабелей // Инженерный вестник Дона. – 2020. – № 4 (64). – С. 12.
3. Полуянович Н.К., Дубяго М.Н. Прогнозирование ресурса кабельных линий с использованием метода искусственных нейронных сетей // Известия ЮФУ. Технические науки. – 2019. – № 3. – С. 51-62.
4. Poluyanovich N.K., Dubyago M.N., Shurykin A.A., Burkov D.V. Estimation of partial discharge energy in a mathematical model of thermal fluctuation processes of a power cable. – Ufa: USATU, 2019. – 642 p.
5. Дубяго М.Н., Полуянович Н.К. Термодинамический способ выявления деструкции изоляции в задачах диагностики и прогнозирования ресурса кабельных систем // Инженерный вестник Дона. – 2017. – № 3 (46). – С. 25.
6. Prime J.B., Valdes J.G. Systems to Monitor the Conductor Temperature of Underground Cable // IEEE Transactions on Power Apparatus and Systems. – 1981. – Vol. PAS-100. – No. 1. – P. 211-219.
7. Меламед М.А. Современные методы анализа и прогнозирования режимов электропотребления в электроэнергетических системах // Итоги науки и техники. Серия Энергетические системы и их автоматизация. – 1988. – Т. 4. – С. 4-111.

8. *Anders G.J., Napieralski A., Orlikowski M., Zubert M.* Advanced Modeling Techniques for Dynamic Feeder Rating Systems // IEEE Transactions on Industry Applications. – 2003. – Vol. 39, No. 3. – P. 619-626.
9. *Chen S.T., David C.Y., Moghaddamjo A.R.* Weather sensitive short-term load forecasting using non fully connected artificial neural network // IEEE Trans. on Power Systems. – 1992. – Vol. 7, No. 3. – P. 1098-1105.
10. *Торопов А.С., Туликов А.Н.* Прогнозирование почасового электропотребления региональной энергосистемы с использованием искусственных нейронных сетей // Вестник ИрГТУ. – 2017. – Т. 21, № 5.
11. *Dash P.K., Ramakrishna G., Liew A.C., Rahman S.* Fuzzy neural networks for time-series forecasting of electric load // IEE Proc. - Gener. Transm. Distrib. – 1995. – Vol. 142, No. 5. – P. 535- 544
12. *Hsy Y., Ho K.* Fuzzy expert systems: An application to short term load forecasting // IEE Proceedings – C. – 1992. – Vol. 139, No. 6. – P. 471-477.
13. *Lee K.Y., Park J.H.* Short-term load forecasting using an artificial neural network // IEEE Trans. on Power Systems. – 1992. – Vol. 7, No. 1. – P.124-130.
14. *Peng T.M., Hubele N.F., Karady G.G.* An adaptive neural network approach to one - week ahead load forecasting // IEEE Trans. on Power Systems. – 1993. – Vol. 8, No. 3. – P. 1195-1201.
15. *Дубяго М.Н., Полуянович Н.К., Пишихонов В.Х.* Метод исследования термофлуктуационных процессов в задачах диагностики и прогнозирования изоляционных материалов // Вестник Донского государственного технического университета. – 2017. – Т. 17, № 3 (90). – С. 117-127.
16. *Андреев А.М., Канискин В.А., Полонский Ю.А.* Исследование строения электроизоляционных материалов силовых кабелей и конденсаторов // Электричество. – 1999. – № 1. – С. 39-44.
17. *Пучков Е.В.* Сравнительный анализ алгоритмов обучения искусственной нейронной сети // Инженерный вестник Дона. – 2013. – № 4. – URL: ivdon.ru/ru/magazine/archive/n4y2013/2135.
18. *Poluyanovich, N.K., Dubyago, M.N., Azarov, N.V., Ogrenichev, A.V., Shurykin, A.A.* Multifactor Model for Forecasting Thermal Processes in the Insulating Materials of Cable Lines // International Multi-Conference on Industrial Engineering and Modern Technologies, FarEastCon 2020. – Doi: 10.1109/FarEastCon50210.2020.9271367.
19. *Лавров Ю.А.* Кабели 6–35 кВ с пластмассовой изоляцией // Новости электротехники. – 2007. – № 1.
20. *Грешняков Г.В.* Цепно-полевой подход к оценке нагрузочной способности кабельных систем // Cabex 2014: Международная выставка кабельно-проводниковой продукции. Новые разработки в области кабельных изделий и материалов: семинар. – М., 2014. – 17 с. – Режим доступа: <http://www.cabex.ru/ru-RU/home/press/news/18513.aspx> (дата обращения: 12.03.2020).
21. *Поляк Б.Т.* Метод Ньютона и его роль в оптимизации и вычислительной математике // Тр. Института Системного Анализа Российской Академии Наук. – 2006. – Т. 28. – С. 44-62.
22. Алгоритм Левенберга-Марквардта. – URL <http://www.machinelearning.ru> (дата обращения: 21.09.2019).
23. *Борисов Е.С.* О методах обучения многослойных нейронных сетей прямого распространения. Ч. 3. Градиентные методы второго порядка. – <http://mechanooid.kiev.ua> (дата обращения: 18.09.2019).
24. *Лиля В.Б.* Алгоритм и программная реализация адаптивного метода обучения искусственных нейронных сетей // Инженерный вестник Дона. – 2012. – № 1. – Режим доступа: <http://ivdon.ru/magazine/archive/n1y2012/626> (доступ свободный).
25. *Белявский Г.И., Пучков Е.В., Лиля В.Б.* Алгоритм и программная реализация гибридного метода обучения искусственных нейронных сетей // Программные продукты и системы. – 2012. – № 4. – С. 96-100.
26. *Галушка В.В., Фахти В.А.* Формирование обучающей выборки при использовании искусственных нейронных сетей в задачах поиска ошибок баз данных // Инженерный вестник Дона. – 2013. – № 2. – URL: ivdon.ru/ru/magazine/archive/n2y2013/1597/.

REFERENCES

1. *Anders G.J.* Rating of Cables on Riser Poles, in Trays, in Tunnels and Shafts - a Review, *IEEE Transactions on Power Delivery*, 1996, Vol. 11, No. 1.
2. *Poluyanovich N.K., Azarov N.V., Ogrenichev A.V., Dubyago M.N., Shurykin A.A.* Sintez neyrosetevoy modeli dlya prognozirovaniya termicheskikh protsessov elektroizolyatsionnykh materialov silovykh kabeley [Synthesis of a neural network model for predicting thermal processes of electrical insulation materials of power cables], *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don], 2020, No. 4 (64), pp. 12.
3. *Poluyanovich N.K., Dubyago M.N.* Prognozirovanie resursa kabel'nykh liniy s ispol'zovaniem metoda iskusstvennykh neyronnykh setey [Predicting the resource of cable lines using the method of artificial neural networks], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2019, No. 3, pp. 51-62.
4. *Poluyanovich N.K., Dubyago M.N., Shurykin A.A., Burkov D.V.* Estimation of partial discharge energy in a mathematical model of thermal fluctuation processes of a power cable. Ufa: USATU, 2019, 642 p.
5. *Dubyago M.N., Poluyanovich N.K.* Termodinamicheskiy sposob vyyavleniya destruktssii izolyatsii v zadachakh diagnostiki i prognozirovaniya resursa kabel'nykh sistem [Thermodynamic method for detecting insulation degradation in problems of diagnostics and forecasting of cable systems resource], *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don], 2017, No. 3 (46), pp. 25.
6. *Prime J.B., Valdes J.G.* Systems to Monitor the Conductor Temperature of Underground Cable, *IEEE Transactions on Power Apparatus and Systems*, 1981, Vol. PAS-100, No. 1, pp. 211-219.
7. *Melamed M.A.* Sovremennye metody analiza i prognozirovaniya rezhimov elektropotrebleniya v elektroenergeticheskikh sistemakh [Modern methods of analysis and forecasting of power consumption modes in electric power systems], *Itogi nauki i tekhniki. Seriya Energeticheskoye sistemy i ikh avtomatizatsiya* [Results of science and technology. Energy Systems and their Automation series], 1988, Vol. 4, pp. 4-111.
8. *Anders G.J., Napieralski A., Orlikowski M., Zubert M.* Advanced Modeling Techniques for Dynamic Feeder Rating Systems, *IEEE Transactions on Industry Applications*, 2003, Vol. 39, No. 3, pp. 619-626.
9. *Chen S.T., David C.Y., Moghaddamjo A.R.* Weather sensitive short-term load forecasting using non fully connected artificial neural network, *IEEE Trans. on Power Systems*, 1992, Vol. 7, No. 3, pp. 1098-1105.
10. *Toropov A.S., Tulikov A.N.* Prognozirovanie pochasovogo elektropotrebleniya regional'noy energosistemy s ispol'zovaniem iskusstvennykh neyronnykh setey [Forecasting the hourly power consumption of a regional power system using artificial neural networks], *Vestnik IrGTU* [Vestnik IrSTU], 2017, Vol. 21, No. 5.
11. *Dash P.K., Ramakrishna G., Liew A.C., Rahman S.* Fuzzy neural networks for time-series forecasting of electric load, *IEE Proc. - Gener. Transm. Distrib.*, 1995, Vol. 142, No. 5, pp. 535- 544
12. *Hsy Y., Ho K.* Fuzzy expert systems: An application to short term load forecasting, *IEE Proceedings – C*, 1992, Vol. 139, No. 6, pp. 471-477.
13. *Lee K.Y., Park J.H.* Short-term load forecasting using an artificial neural network, *IEEE Trans. on Power Systems*, 1992, Vol. 7, No. 1, pp.124-130.
14. *Peng T.M., Hubele N.F., Karady G.G.* An adaptive neural network approach to one - week ahead load forecasting, *IEEE Trans. on Power Systems*, 1993, Vol. 8, No. 3, pp. 1195-1201.
15. *Dubyago M.N., Poluyanovich N.K., Pshikhopov V.Kh.* Metod issledovaniya termoflukuatsionnykh protsessov v zadachakh diagnostiki i prognozirovaniya izolyatsionnykh materialov [Method of research of thermofluctuation processes in problems of diagnostics and forecasting of insulating materials], *Vestnik Donskogo gosudarstvennogo tekhnicheskogo universiteta* [Vestnik Donskogo gosudarstvennogo tekhnicheskogo universiteta], 2017, Vol. 17, No. 3 (90), pp. 117-127.
16. *Andreev A.M., Kaniskina V.A., Polonskiy Yu.A.* Issledovanie stroeniya elektroizolyatsionnykh materialov silovykh kabeley i kondensatorov [Investigation of the structure of electrical insulation materials of power cables and capacitors], *Elektrichestvo* [Electricity], 1999, No. 1, pp. 39-44.

17. Puchkov E.V. Sravnitel'nyy analiz algoritmov obucheniya iskusstvennoy neyronnoy seti [Comparative analysis of artificial neural network training algorithms], *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don], 2013, No. 4. Available at: ivdon.ru/ru/magazine/archive/n4y2013/2135.
18. Poluyanovich, N.K., Dubyago, M.N., Azarov, N.V., Ogrenichev, A.V., Shurykin, A.A. Multifactor Model for Forecasting Thermal Processes in the Insulating Materials of Cable Lines, *International Multi-Conference on Industrial Engineering and Modern Technologies, FarEastCon 2020*. Doi: 10.1109/FarEastCon50210.2020.9271367.
19. Lavrov Yu.A. Kabeli 6–35 kV s plastmassovoy izolyatsiey [Cables of 6-35 kV with plastic insulation], *Novosti elektrotehniki* [News of electrical engineering], 2007, No. 1.
20. Greshnyakov G.V. Tsepno-polevoy podkhod k otsenke nagruzochnoy sposobnosti kabel'nykh sistem [Chain-field approach to assessing the load capacity of cable systems], *Cabex 2014: Mezhdunarodnaya vystavka kabel'no-provodnikovoy produktsii. Novyye razrabotki v oblasti kabel'nykh izdeliy i materialov: seminar* [Cabex 2014: International Exhibition of Cable and Wire Products. New developments in the field of cable products and materials: seminar]. Moscow, 2014, 17 p. Available at: <http://www.cabex.ru/ru-RU/home/press/news/18513.aspx> (accessed 12 March 2020).
21. Polyak B.T. Metod N'yutona i ego rol' v optimizatsii i vychislitel'noy matematike [The Newton method and its role in optimization and computational mathematics], *Tr. Instituta Sistemnoy Analiza Rossiyskoy Akademii Nauk* [Proceedings of the Institute of System Analysis of the Russian Academy of Sciences], 2006, Vol. 28, pp. 44-62.
22. Algoritm Levenberga-Markvardta [The Levenberg-Marquardt algorithm]. Available at: <http://www.machinelearning.ru> (accessed 21 September 2019).
23. Borisov E.S. O metodakh obucheniya mnogosloynnykh neyronnykh setey pryamogo rasprostraneniya. Ch. 3. Gradientnye metody vtorogo poryadka [On methods of training multi-layer neural networks of direct propagation. Part 3. Gradient methods of the second order]. Available at: <http://mechanoid.kiev.ua> (accessed 21 September 2019).
24. Lila V.B. Algoritm i programmaya realizatsiya adaptivnogo metoda obucheniya iskusstvennykh neyronnykh setey [Algorithm and software implementation of adaptive training method for artificial neural networks], *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don], 2012, No. 1. Available at: <http://ivdon.ru/magazine/archive/n1y2012/626> (access is free).
25. Belyavskiy G.I., Puchkov E.V., Lila V.B. Algoritm i programmaya realizatsiya gibridnogo metoda obucheniya iskusstvennykh neyronnykh setey [Algorithm and software implementation of the hybrid method of training artificial neural networks], *Programmnye produkty i sistemy* [Software products and systems], 2012, No. 4, pp. 96-100.
26. Galushka V.V., Fakhti V.A. Formirovanie obuchayushchey vyborki pri ispol'zovanii iskusstvennykh neyronnykh setey v zadachakh poiska oshibok baz dannykh [Formation of a training sample when using artificial neural networks in problems of searching for database errors], *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don], 2013, No. 2. Available at: ivdon.ru/ru/magazine/archive/n2y2013/1597/.

Статью рекомендовала к опубликованию к.т.н., доцент М.Г. Левина.

Полюянович Николай Константинович – Южный федеральный университет; e-mail: nik1-58@mail.ru; г. Таганрог, Россия; тел.: 89185693365; кафедра электротехники и мехатроники.

Дубяго Марина Николаевна – e-mail: w_m88@mail.ru; тел.: 89281758225; кафедра электротехники и мехатроники; аспирант.

Poluyanovich Nikolay Konstantinovich – Southern Federal University; e-mail: nik1-58@mail.ru; Taganrog, Russia; phone: +79185693365; the department of electric technics and mechatronics.

Dubyago Marina Nikolaevna – e-mail: w_m88@mail.ru; phone: +79281758225; the department of electrical engineering and mechatronics; graduate student.

Раздел II. Связь, навигация и наведение

УДК 621.396.67

DOI 10.18522/2311-3103-2021-2-74-83

А.А. Ваганова, Н.Н. Кисель, А.И. Паньчев

НАПРАВЛЕННЫЕ И ПОЛЯРИЗАЦИОННЫЕ СВОЙСТВА МИКРОПОЛОСКОВОЙ РЕКОНФИГУРИРУЕМОЙ АНТЕННЫ, ПЕРЕСТРАИВАЕМОЙ ПО ЧАСТОТЕ И ПОЛЯРИЗАЦИИ

Реконфигурируемой антенной называется антенна, параметры которой могут изменяться в соответствии с требованиями, предъявляемыми конкретной ситуацией. Изменяемыми параметрами могут быть диапазон частот с наилучшим согласованием, диаграмма направленности, поляризация, а также различные комбинации этих параметров. В данной работе предложена конструкция реконфигурируемой микрополосковой антенны, перестраиваемой по частоте и поляризации, и исследованы ее диаграмма направленности и поляризационные свойства. Антенна имеет компактные размеры и может быть использована в беспроводных системах связи, работающих в диапазоне 2–7 ГГц. В конструкции антенны имеется 5 рпн-диодов, в зависимости от состояния которых изменяется резонансная частота и поляризация излучения антенны. Выполнено моделирование предлагаемой антенны в программе FEKO и получены ее основные параметры. Анализ результатов моделирования показал, что для нижней части исследуемого диапазона частот (2,05, 2,45 и 3,7 ГГц) поляризация линейна. При работе в более высоком поддиапазоне (5,4, 5,6 и 5,75 ГГц) антенна имеет круговую поляризацию, направление вращения которой изменяется в зависимости от состояния диодов. Возможность переключения поляризации на ортогональную на одной и той же частоте позволяет улучшить условия приема сигнала в условиях многолучевого распространения.

Микрополосковая антенна; реконфигурируемая антенна; поляризация; диаграмма направленности; коэффициент эллиптичности.

A.A. Vaganova, N.N. Kisel, A.I. Panychev

DIRECTIONAL AND POLARISATION PROPERTIES OF A MICROSTRIP RECONFIGURABLE ANTENNA WITH TUNABLE FREQUENCY AND POLARISATION

A reconfigurable antenna is an antenna with parameters that can be varied according to the requirements of a particular situation. Under variable parameters we can understand the operating frequency range, radiation pattern, polarization, as well as various combinations of these parameters. In this paper, we propose the design of a reconfigurable microstrip antenna with tunable frequency and polarization, and investigate its radiation pattern and polarisation properties. The antenna has compact dimensions and can be used in wireless communication systems operating in the 2–7 GHz range. 5 pin diodes are included into the antenna design to change the resonant frequency and polarization of the antenna by turning the diodes on and off. The simulation of the proposed antenna was performed in the FEKO program and the main parameters of the antenna were obtained. The analysis of the simulation results showed that for the lower part of the studied frequency range (2.05, 2.45 and 3.7 GHz), the polarization is linear. When operating in the higher sub-range (5.4, 5.6 and 5.75 GHz), the antenna is circularly polarized, and the direction of the polarisation vector rotation depends on the connection of the diodes. This ability to switch polarization to orthogonal at the same frequency allows to perform efficient signal reception in the conditions of multipath propagation.

Microstrip antenna; reconfigurable antenna; polarization; radiation pattern; axial ratio.

Введение. Реконфигурируемой антенной называется антенна, параметры которой могут изменяться в соответствии с требованиями, предъявляемыми конкретной ситуацией. Изменяемыми параметрами могут быть диапазон частот с наилучшим согласованием, диаграмма направленности, поляризация, а также различные комбинации этих параметров. Наиболее распространенным способом построения реконфигурируемой антенны является соединение различных частей ее конструкции с помощью радиочастотных переключателей, включение и выключение которых направляет ток по требуемому пути, за счет чего меняются характеристики антенны. При этом в качестве переключателей используются либо полупроводниковые элементы (рpn-диоды, варакторы, полевые транзисторы), либо MEMS (микроэлектромеханические системы). Выбор конкретной элементной базы и топологии зависит от решаемой проблемы, но в общем случае при выборе переключателя следует учитывать такие характеристики, как импеданс, нормальное состояние (включен или выключен), полосу частот, допустимую мощность и время срабатывания [1]. В настоящее время наиболее часто используется реконфигурирование с целью настройки по частоте [2–14]. Реконфигурируемые антенны с переключаемой поляризацией описаны в работах [15–25].

При проектировании антенны необходимо определить условия ее функционирования, сформулировать требования к ее характеристикам и создать конструкцию, позволяющую осуществить реконфигурацию в соответствии с этими требованиями. Целью данной статьи является моделирование антенны для устройств беспроводной связи (2–7 ГГц), способной перестраиваться по частоте и поляризации, что повысит эффективность приема сигнала в условиях множественных отражений [26–28].

Конструкция антенны. В основу разработки конструкции предложенной микрополосковой реконфигурируемой антенны положены результаты, описанные в статье [14], где предложена антенна, перестраиваемая по частоте, и способная работать в трех значительно отдаленных друг от друга частотных диапазонах беспроводной связи, и в статье [20], где рассмотрена антенна, перестраиваемая по поляризации. Выбор именно этих антенн обоснован их конструктивными и геометрическими параметрами, позволяющими создать на их основе единую конструкцию, исследованию свойств которой посвящена данная работа.

Модель антенны создана в САПР FEKO. В результате конструкционной оптимизации линии питания и геометрии излучающего элемента базового эскизного проекта, составленного на основе рекомендаций, приведенных в [14, 20], сформирована конструкция антенны, показанная на рис. 1 (основные размеры приведены в миллиметрах). Антенна расположена на подложке из диэлектрика FR4 с параметрами $\epsilon = 4,7$, $\text{tg}\delta = 0,018$. Общий размер антенны равен 20x29x1,6 мм. Излучающая часть антенны (рис. 1,а) представляет собой ромб, к которому с двух сторон может поступать сигнал из единственного порта. При этом в обе ветви, по которым может передаваться сигнал, встроены рpn-диоды (D1 и D2). В зависимости от режима работы диодов симметрия конструкции антенны может быть нарушена, и за счет этого может изменяться ее поляризация.

Заземляющая плоскость предлагаемой антенны показана на рис. 1,б. С помощью переключения трех рpn-диодов D3, D4 и D5 конструкция заземляющей плоскости может изменяться, что позволяет скачкообразно изменять резонансную частоту в диапазоне от 2 до 7 ГГц [14].

Поскольку разработка электрической принципиальной схемы переключения режимов работы антенны не проводилась, то с целью упрощения процесса моделирования учитывались лишь наиболее важные свойства элементов. Так, поскольку

ку pin-диоды имеют два состояния ON (+) и OFF (-), в данной модели они заменяются наличием или отсутствием металлического элемента (D1 – D5), создающего электрический контакт в рассматриваемом участке антенны.

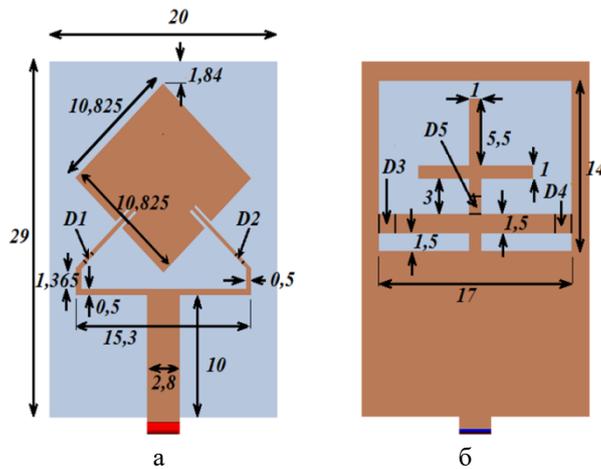


Рис. 1. Конструкция антенны: а – излучающая сторона, б – заземляющая плоскость

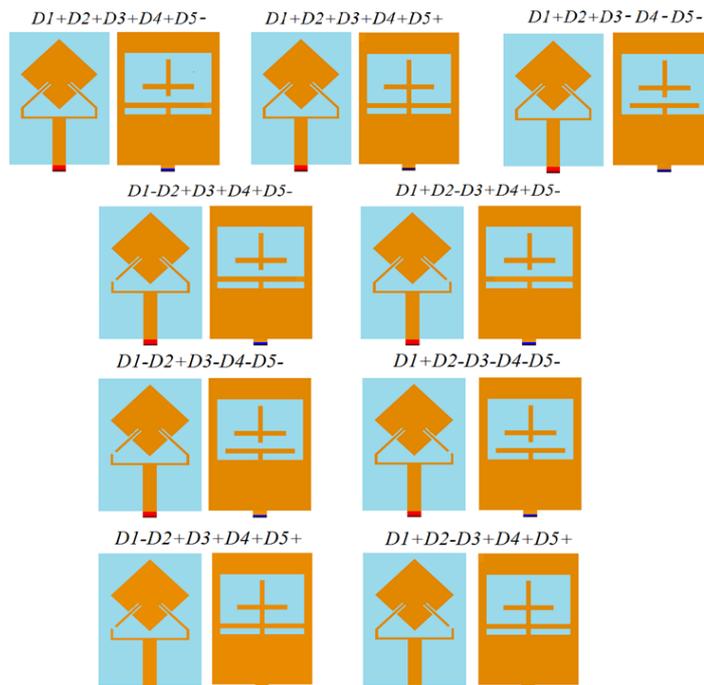


Рис. 2. Варианты конструкции антенны, соответствующие разным вариантам переключения диодов

На рис. 2 показано, как меняется геометрия антенны в зависимости от подключения тех или иных pin-диодов. Эти состояния соответствуют различным режимам работы антенны, которые будут проанализированы далее.

Расчет и анализ характеристик антенны. Для каждого варианта переключения была рассчитана зависимость параметра S_{11} от частоты, соответствующие графики приведены на рис. 3. Видно, что в зависимости от включения/выключения диодов резонансная частота антенны меняется в пределах от 2 до 7 ГГц.

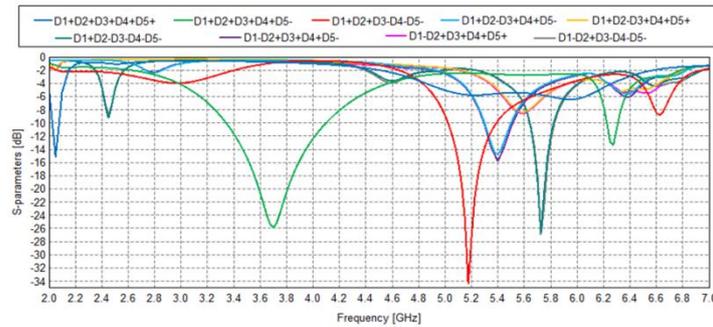


Рис. 3. Зависимость параметра S_{11} от частоты

Сравнение полученных результатов с данными, приведенными в статьях [14] и [20], показывает, что при объединении в одну конструкцию точные значения резонансных частот изменились, однако все они по-прежнему лежат в диапазоне, используемом для беспроводной связи, что является достаточным результатом в рассматриваемом случае. Оптимизация конструкции антенны для приведения ее в соответствие с каким-либо конкретным стандартом беспроводной связи представляется принципиально возможной и является задачей следующего этапа исследования.

Зависимость коэффициента стоячей волны (КСВ) по напряжению для различных режимов от частоты показана на рис. 4. Видно, что для каждого варианта достигаются минимальные значения КСВ (см. также табл. 1), но при этом ширина полосы частот, в которой $КСВ < 3$, существенно различается для разных режимов работы и достигает от ± 50 МГц до ± 600 МГц от резонансной частоты.

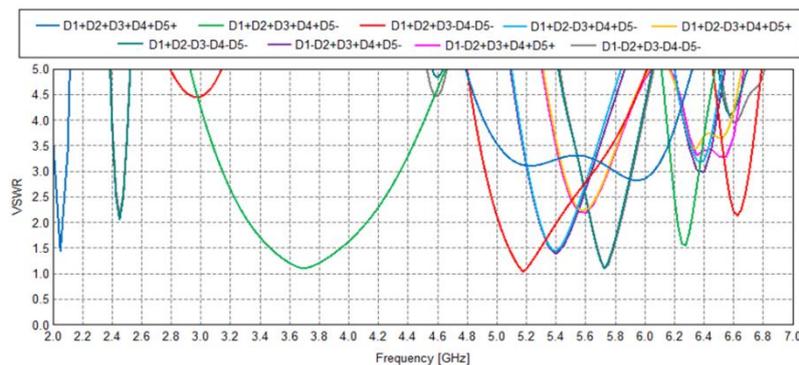


Рис. 4. Зависимость КСВН от частоты

При подключении одновременно двух pin-диодов D1 и D2 излучающая часть антенны является симметричной и антенна имеет линейную поляризацию. При этом резонансная частота антенны зависит от состояния остальных диодов

(D3 – D5). На рис. 5 показаны объемные диаграммы направленности антенны для случая, когда включены все диоды, кроме D5. Диаграммы приведены для двух частот, на которых достигается минимальное значение S_{11} . На рис. 6 приведены графики коэффициента эллиптичности антенны на этих частотах. Графики построены в двух перпендикулярных вертикальных плоскостях (считаем, что антенна расположена горизонтально и центр сферической системы координат совмещен с геометрическим центром подложки). При этом в качестве коэффициента эллиптичности (*Axial ratio*) взято отношение максимального радиуса эллипса поляризации к минимальному, то есть он равен 0 дБ в случае идеально круговой поляризации и стремится к бесконечности в случае идеально линейной поляризации. Рис. 6 показывает, что при использовании данной комбинации диодов антенна обладает линейной поляризацией на частоте 3,7 ГГц. Кроме того, КСВ антенны имеет приемлемое значение также на более высокой частоте (6,25 ГГц), но при этом ее поляризация линейна только в нескольких узких секторах углов, что позволяет говорить об эллиптической поляризации.

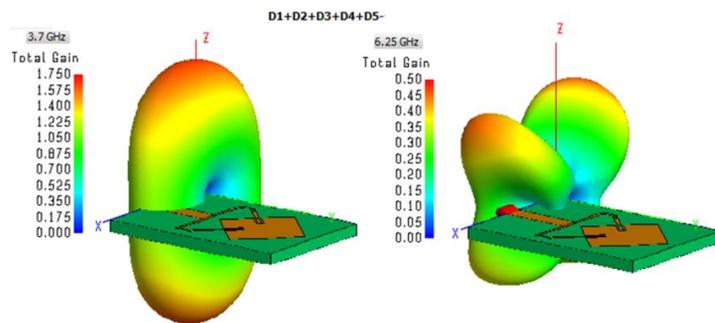


Рис. 5. Диаграмма направленности антенны в режиме $D1+D2+D3+D4+D5-$

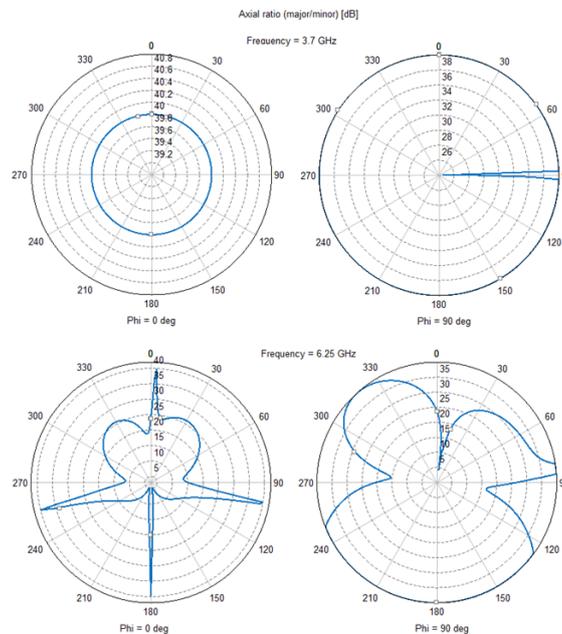


Рис. 6. Коэффициент эллиптичности антенны в режиме $D1+D2+D3+D4+D5-$

При подключении только одного из диодов D1 или D2 симметрия излучающей части нарушается. При этом резонансная частота по-прежнему определяется комбинацией диодов D3 — D5, а поляризация меняется на ортогональную в зависимости от того, какой из диодов (D1 или D2) подключен. Для иллюстрации этого явления на рис. 7, рис. 9 и рис. 11 показаны объемные диаграммы направленности антенн, а на рис. 8, рис. 10 и рис. 12 соответствующие им графики коэффициента эллиптичности.

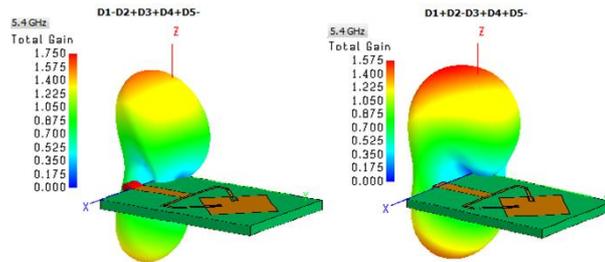


Рис. 7. Диаграммы направленности антенны в режимах $D1-D2+D3+D4+D5-$ и $D1+D2-D3+D4+D5-$

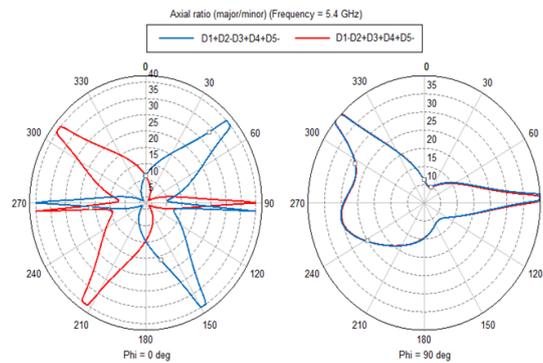


Рис. 8. Коэффициент эллиптичности антенны в режимах $D1-D2+D3+D4+D5-$ и $D1+D2-D3+D4+D5-$

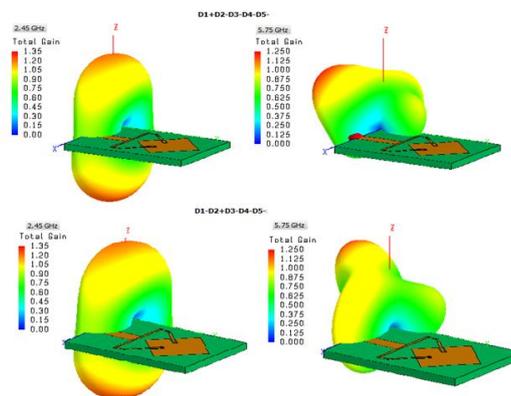


Рис. 9. Диаграммы направленности антенны в режимах $D1-D2+D3-D4-D5-$ и $D1+D2-D3-D4-D5-$

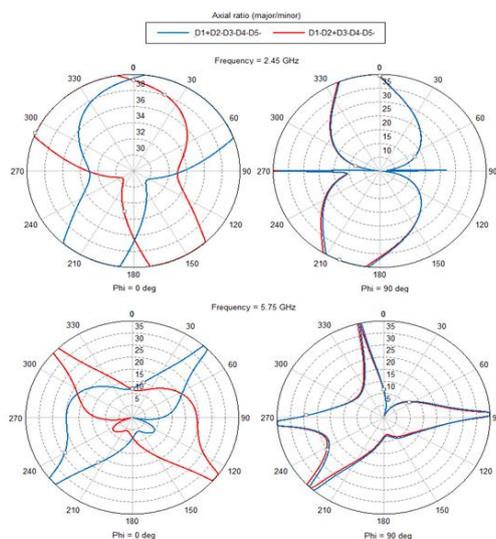


Рис. 10. Коэффициент эллиптичности антенны в режимах $D1-D2+D3-D4-D5-$ и $D1+D2-D3-D4-D5-$

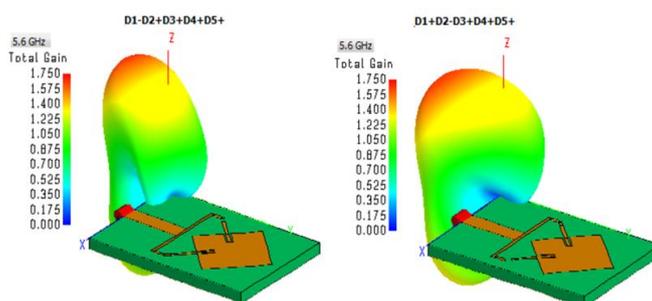


Рис. 11. Диаграммы направленности антенны в режимах $D1-D2+D3+D4+D5+$ и $D1+D2-D3+D4+D5+$

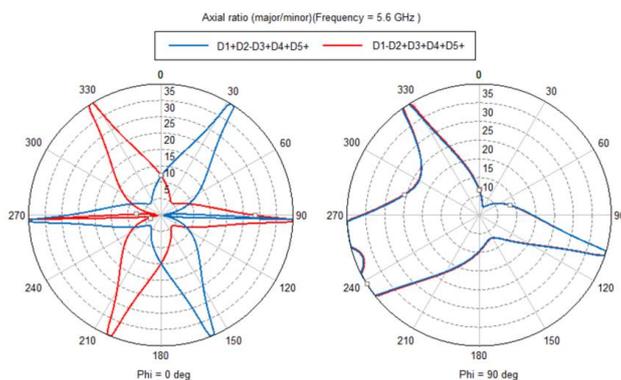


Рис. 12. Коэффициент эллиптичности антенны в режимах $D1-D2+D3+D4+D5+$ и $D1+D2-D3+D4+D5+$

По приведенным рисункам видно, что антенна в рассмотренных режимах, то есть в частотном диапазоне 5–6 ГГц, имеет круговую поляризацию везде, кроме узких секторов углов. Также на рис. 10 показан случай, где антенна при той же комбинации диодов может работать и на более низкой частоте 2,45 ГГц, но здесь обеспечивается линейная поляризация. Для каждой рассматриваемой пары режимов графики коэффициента эллиптичности одинаковы по форме, но противоположно ориентированы, что говорит о противоположном направлении вращения вектора поляризации. Таким образом видно, что выбор включения одного из диодов D1 или D2 позволяет изменить поляризацию антенны на одной частоте с правой на левую.

Полученные при моделировании параметры антенны для комбинаций pin-диодов, обеспечивающих минимальный КСВ, приведены в табл. 1.

Таблица 1

Параметры антенны в различных режимах работы

Комбинация pin-диодов	Резонансная частота	КСВ	S11	Поляризация	Axial ratio (max/min)
D1+D2+D3+ D4+D5-	3,7 ГГц	1,1	-26 дБ	Линейная	40 дБ
	6,25 ГГц	1,6	-13 дБ	Эллиптическая	5 дБ
D1+D2+D3+ D4+D5+	2,05 ГГц	1,5	-15 дБ	Линейная	29 дБ
D1+D2+D3- D4-D5-	5,2 ГГц	1,1	-27 дБ	Эллиптическая	18 дБ
	6,65 ГГц	2,2	-9 дБ	Эллиптическая	6 дБ
D1-D2+D3+ D4+D5-	5,4 ГГц	1,5	-16 дБ	Левая круговая	1,7 дБ
D1+D2-D3+ D4+D5-			-15 дБ	Правая круговая	1,7 дБ
D1-D2+D3- D4-D5-	2,45 ГГц	2	-9 дБ	Линейная	29 дБ
D1+D2-D3- D4-D5-					30 дБ
D1-D2+D3-D4- D5-	5,75 ГГц	1,2	-27 дБ	Левая круговая	0,5 дБ
D1+D2-D3- D4-D5-				Правая круговая	0,5 дБ
D1-D2+D3+ D4+D5+	5,6 ГГц	2,2	-9 дБ	Левая круговая	1 дБ
D1+D2-D3+ D4+D5+			-9 дБ	Правая круговая	1 дБ

Заключение. В данной работе предложена конструкция реконфигурируемой микрополосковой антенны, перестраиваемой по частоте и поляризации, и исследованы ее диаграммы направленности и угловые зависимости коэффициента эллиптичности. Антенна имеет компактные размеры и может быть использована в бес-

проводных системах связи, работающих в диапазоне 2–7 ГГц. В конструкции антенны имеется 5 рin-диодов, в зависимости от состояния которых изменяется резонансная частота и поляризация антенны. Так, для нижней части исследуемого частотного диапазона (2,05, 2,45 и 3,7 ГГц) антенна имеет линейную поляризацию. При работе в более высоком поддиапазоне (5,4, 5,6 и 5,75 ГГц) антенна имеет круговую поляризацию, направление вращения которой изменяется в зависимости от подключения диодов. Эта возможность переключения поляризации на ортогональную на одной и той же частоте позволяет принимать сигнал в условиях отражений.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Haupt R. L., Lanagan M. Reconfigurable antennas, *IEEE Antennas Propag. Mag.*, 2013, Vol. 55 (1), pp. 49-61.
2. Hassan. M.M., Zahid Z., Khan A.A., Rashid I., Rauf A., Maqsood M. Two element MIMO antenna with frequency reconfigurable characteristics utilizing RF MEMS for 5G applications, *Journal of Electromagnetic Waves and Applications*, 2020, Vol. 34, Issue 9: Massive MIMO Toward 5G, pp. 1210-1224.
3. Purisima M.C.L., Salvador M., Agustin S.G.P., Cunanan M.T. Frequency and pattern reconfigurable antennas for community cellular applications, *2016 IEEE Region 10 Conference (TENCON)*, 22-25 Nov. 2016, Singapore.
4. Constantine J., Tawk Y., Barbin S.E., et al. Reconfigurable antennas: design and applications, *Proc. IEEE*, 2015, Vol. 103 (3), pp. 424-437.
5. Wright M.D., Baron W., Miller J., et al. MEMS reconfigurable broadband patch antenna for conformal applications, *IEEE Trans. Antennas Propag.*, 2018, Vol. 66 (6), pp. 2770-2778.
6. Han L., Wang C., Chen X., et al. Compact frequency-reconfigurable slot antenna for wireless applications, *IEEE Antennas Wirel. Propag. Lett.*, 2016, Vol. 15, pp. 1795-1798.
7. Rajagopalan H., Kovitz J. M., Rahmat-Samii Y. MEMS reconfigurable optimized E-shaped patch antenna design for cognitive radio, *IEEE Trans. Antennas Propag.*, 2014, Vol. 62 (3), pp. 1056-1064.
8. Abdulraheem Y., Oguntala G. A., Abdullah A. S., et al. Design of frequency reconfigurable multi-band compact antenna using two pin diodes for WLAN/WiMAX applications, *IET Microwaves, Antennas Propag.*, 2017, Vol. 11 (8), pp. 1098-1105.
9. Pandit S., Mohan A., Ray P. Compact frequency-reconfigurable MIMO antenna for microwave sensing applications in WLAN and WiMAX frequency bands, *IEEE Sens Lett.*, 2018, No. 2 (2), pp. 1-4. Doi: 10.1109/LSENS.2018.2822598.
10. Zhao X., Riaz S. A dual-band frequency reconfigurable MIMO patch-slot antenna based on reconfigurable microstrip feedline, *IEEE Access.*, 2018, Vol. 6, pp. 41450-41457.
11. Hussain R., Sharawi M.S., Shamim A. 4-element concentric pentagonal slot-line-based ultra-wide tuning frequency reconfigurable MIMO antenna system, *IEEE Trans. Antennas Propag.*, 2018, Vol. 66 (8), pp. 4282-4287.
12. Hussain R., Khan M.U., Sharawi M.S. Design and analysis of a miniaturized meandered slot-line-based quad-band frequency agile MIMO antenna, *IEEE Trans. Antennas Propag.*, 2019, Vol. 68 (3), pp. 2410-2415.
13. Hussain R., Sharawi M.S. A cognitive radio reconfigurable MIMO and sensing antenna system, *IEEE Antennas Wirel. Propag. Lett.*, 2014, Vol. 14, pp. 257-260.
14. Borhani M., Rezaei P., Valizade A. Design of a Reconfigurable Miniaturized Microstrip Antenna for Switchable Multiband Systems, *IEEE Antennas and Wireless Propagation Letters*, 2016, Vol. 15, pp. 822-825.
15. Song T., Lee Y., Ga D., Choi J. A Polarization Reconfigurable Microstrip Patch Antenna using PIN Diodes. *Proceedings of APMC 2012, Kaohsiung, Taiwan, Dec. 4-7, 2012*, pp. 616-618.
16. Lee M.J., Sung Y., Kim S.H., Kim Y.S. Polarization Reconfigurable Microstrip Patch Antenna with Complementary Split Ring Resonator (CSRR), *Proceedings of APMC 2012, Kaohsiung, Taiwan, Dec. 4-7, 2012*, pp. 619-621.
17. Lee S.W., Sung Y., Lee S.J., Yoon H.S. and Park H.J. A reconfigurable patch antenna with symmetrical structure for polarization diversity., *2015 9th European Conference on Antennas and Propagation (EuCAP), 13-17 April 2015, Lisbon, Portugal*.

18. Ji L.Y., Qin P.Y., Guo Y.J., Fu G., Mittra R. A Wideband Polarization Reconfigurable Antenna for WLAN Applications, *2016 10th European Conference on Antennas and Propagation (EuCAP), 10-15 April 2016, Davos, Switzerland.*
19. Pan P., Guan B. A Wideband Polarization Reconfigurable Antenna with Six Polarization States. 2018 12th International Symposium on Antennas, *Propagation and EM Theory (ISAPE), 3-6 Dec. 2018, Hangzhou, China.*
20. Roseli W.I., Moktar N.H., Ali M.T. Polarization Reconfigurable Microstrip Patch Antenna for Wireless Communication Applications, *2018 International Symposium on Antennas and Propagation (ISAP), Busan, Korea (South), 23-26 Oct. 2018.*
21. Qi Ch., Cui Y., Li R. A Quad-Polarization Reconfigurable Omnidirectional Antenna, *2018 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting, 8-13 July 2018, Boston, MA, USA.*
22. McMichael I.T. A Mechanically Reconfigurable Patch Antenna With Polarization Diversity, *IEEE Antennas and Wireless Propagation Letters, vol. 17, issue 7, July 2018.*
23. Eqbal Z., Varma R., Ghosh J. A Compact Polarization Reconfigurable Square Slot Antenna for WiMAX, *2018 3rd International Conference on Microwave and Photonics (ICMAP 2018), 9-11 February, 2018.*
24. Wong H., Lin W., Wang X., Lu M. LP and CP Polarization Reconfigurable Antennas for Modern Wireless Applications, *2017 International Symposium on Antennas and Propagation (ISAP), 30 Oct.-2 Nov. 2017, Phuket, Thailand.*
25. Begum H., Wang X., Lu M. A Polarization-Reconfigurable Microstrip Antenna Design Based on Parasitic Pin Loading, *2017 IEEE Radio and Wireless Symposium (RWS), 15-18 Jan. 2017, Phoenix, AZ, USA.*
26. Vaganova A.A., Kisel N.N., Panychev A.I. Statistical Account of the Obstacle Position in a MIMO System Channel Matrix Indoors, *2020 Moscow Workshop on Electronic and Networking Technologies (MWENT), 11-13 March 2020, Moscow, Russia.*
27. Panychev A.I., Vaganova A.A. Efficient three-dimensional ray tracing and electromagnetic field intensity estimation algorithm for WLAN, *2017 IEEE East-West Design & Test Symposium (EWDTS), 29 Sept.-2 Oct. 2017, Novi Sad.*
28. Panychev A.I., Vaganova A.A. Approximate calculation of the diffraction coefficients in the analysis of WLAN coverage area. 26-30 June 2017, Divnomorskoe, Russia.

Статью рекомендовал к опубликованию д.т.н., профессор А.М. Макаров

Ваганова Анастасия Алексеевна – Южный федеральный университет; email: anastasia_vaganova@mail.ru; г. Таганрог, Россия; м.н.с.

Кисель Наталья Николаевна – e-mail: nnkisel@sfdedu.ru; кафедра антенн и радиопередающих устройств; к.т.н.; доцент.

Паньчев Андрей Иванович – e-mail: ruu2011@mail.ru; кафедра антенн и радиопередающих устройств; к.т.н.; доцент.

Vaganova Anastasia Alexeevna – Southern Federal University; e-mail: anastasia_vaganova@mail.ru; Taganrog, Russia; junior researcher.

Kisel Natalia Nikolayevna – e-mail: nnkisel@sfdedu.ru; the department of antennas and radio transmitters; cand. of emg. sc.; associate professor.

Panychev Andrey Ivanovich – e-mail: ruu2011@mail.ru; the department of antennas and radio transmitters; cand. of emg. sc.; associate professor.

А.Н. Зикий, А.С. Кочубей

ГЕНЕРАТОР ДЕЦИМЕТРОВОГО ДИАПАЗОНА

Гетеродин приёмника и задающий генератор передатчика являются важнейшими составными частями, определяя их стабильность и диапазонные свойства. В последние годы завод «Метеор» создал ряд новых микросхем генераторов, управляемых напряжением (ГУН), с высокими электрическими параметрами. Однако рекламные материалы предприятия не содержат ряд параметров, важных с точки зрения потребителя. Целью данной работы является исследование основных характеристик генератора, в том числе не задекларированных поставщиком: ширины спектра сигнала, средней крутизны модуляционной характеристики, уровня гармоник. Объектом исследования является микросхема ГУН382 в типовой схеме включения. Представлены результаты экспериментального исследования ГУН, работающего в области 1200 МГц. Дана оценка паразитных продуктов в спектре выходного сигнала. Спектр выходного сигнала демонстрирует малую ширину спектральной линии. Измерены модуляционные характеристики при изменении управляющего напряжения и напряжения питания, вычислена их средняя крутизна. Эти данные позволяют предъявить обоснованные требования к стабильности управляющего и питающего напряжений. Полученные результаты могут быть использованы в приёмно-передающей аппаратуре связи, навигации, радиоэлектронной борьбы. Статья расширяет представление о линейке генераторов завода «Метеор», демонстрирует их высокие электрические характеристики: диапазон рабочих частот 1200 ± 16 МГц; выходная мощность не менее 1 дБм; напряжение питания + 5 В; управляющее напряжение от 0 до 8 В; уровень второй и третьей гармоники не превышает минус 22 дБ по отношению к полезному сигналу.

Генератор; управляемый напряжением; дециметровый диапазон волн; эксперимент; модуляционная характеристика; ширина спектра; уровень гармоник.

A.N. Zikiy, A.S. Kochubey

DECIMETER RANGE GENERATOR

The receiver's heterodyne and the transmitter's master oscillator are the most important components, as they determine their stability and range properties. In recent years, the Meteor plant has created a number of new voltage-controlled generator chips with high electrical parameters. However, the advertising materials of this company do not contain a number of parameters that are important from the point of view of the consumer. Therefore, the purpose of this work is to study the main characteristics of the generator, including those not declared by the supplier: the width of the signal spectrum, the average steepness of the modulation characteristic, the level of harmonics. The object of the study is the GUN382 microchip in a typical switching circuit. The results of an experimental study of a VCO operating in the 1200 MHz region are presented. The estimation of parasitic products in the spectrum of the output signal is given. Photos of the spectrum of the output signal showing a small width of the spectral line are presented. The modulation characteristics are measured when the control voltage and supply voltage change, and their average steepness is calculated. These data allow us to make reasonable requirements for the stability of the control and supply voltages. The results obtained can be used in receiving and transmitting equipment for communication, navigation, and electronic warfare. The article expands the idea of the line of generators of the Meteor plant, demonstrates their high electrical characteristics: operating frequency range 1200 ± 16 MHz; output power of at least 1 dBm; supply voltage + 5V; control voltage from 0 to 8 V; the level of the second and third harmonics does not exceed minus 22 dB in relation to the useful signal.

Generator; voltage-controlled; decimeter wavelength range; modulation characteristics of the experiment; width of the spectrum; level of harmonics.

Введение. Генераторы, управляемые напряжением (ГУН), находят широкое применение в качестве гетеродинов приёмников, задающих генераторов передатчиков, используются в составе синтезаторов частот и измерительных генераторов. Их исследованию посвящена обширная литература, в том числе монографии [1–8], учебные пособия [9, 10], статьи [11–18]. Появление новой элементной базы [19, 20] стимулирует дальнейшие исследования. Объектом исследования является ГУН типа ГУН382-1-В-1200М-5 (изготовитель завод «Метеор»). Достоинством выбранной модели ГУН является малые габариты и масса, низкое токопотребление, высокая линейность модуляционной характеристики, низкий уровень фазовых шумов и гармоник основной частоты. Целью работы является исследование характеристик и параметров генератора, незадекларированных изготовителем.

Схема и конструкция. Схема включения ГУН приведена на рис. 1. Она соответствует типовой схеме включения этого генератора.

Конструкция генератора представляет собой печатную плату, с одной стороны которой установлены все навесные элементы, а монтаж выполнен на другой стороне платы. Плата изготовлена из фольгированного стеклотекстолита размером 70х30х1,5 мм. В качестве выходного СВЧ соединителя использована розетка типа СР-50-727. В качестве низкочастотного соединителя для подачи питания и управления использована вилка типа СНП346-10ВП-21-2-В. Используются конденсаторы типов КТ, КМ6 и К10-17а.

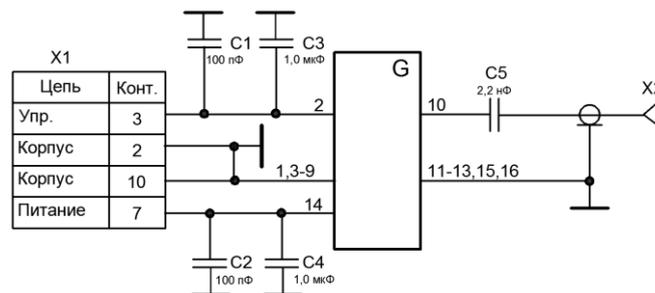


Рис. 1. Схема включения ГУН382-1-В-1200М-5

Эксперимент. Целью эксперимента является исследование основных параметров и характеристик генератора.

Эксперимент проводился на установке, структурная схема которой приведена на рис. 2.

В первом эксперименте исследовали зависимости частоты и мощности от управляющего напряжения. Результаты измерений сведены в табл. 1. По данным таблицы построены графики на рис. 3, из которых видно, что изменение частоты составило 32,7 МГц, а изменение мощности равно 1,85 дБ.

Во втором эксперименте снимались зависимости частоты и мощности ГУН от напряжения питания. Результаты измерений представлены в табл. 2. По данным таблицы построены графики на рис. 4. Из графиков видно, что изменение частоты составило 0,3 МГц, а изменение мощности равно 0,65 дБ.

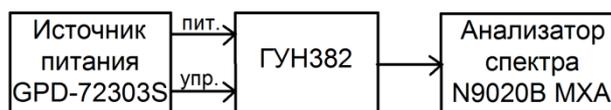


Рис. 2. Структурная схема измерительного стенда

В третьем эксперименте проводилось фотографирование спектра выходного сигнала при управляющем напряжении 4 В. График спектра выходного сигнала приведён на рис. 5. При этом полоса пропускания анализатора спектра составила 10 кГц. Ширина спектра сигнала на уровне минус 3 дБ от максимума составляет 10 кГц, что свидетельствует о высокой добротности резонансной системы. Для оценки уровня гармонических составляющих получен спектр выходного сигнала в широкой полосе. График спектра гармоник с указанными уровнями изображён на рис. 6.

Таблица 1

Зависимость частоты и мощности от управляющего напряжения

Управляющее напряжение $U_{упр}$, В	Частота f , МГц	Выходная мощность $P_{вых}$, дБм
0	1184	-4.76
0.5	1186.9	-5.09
1.0	1188.5	-4.85
1.5	1190.3	-4.37
2.0	1192.1	-4.67
2.5	1193.9	-4.59
3.0	1195.9	-4.32
3.5	1198.1	-4.25
4.0	1200.5	-3.62
4.5	1203.1	-3.94
5.0	1205.7	-3.76
5.5	1208.1	-3.24
6.0	1210.3	-3.47
6.5	1212.3	-3.69
7.0	1214.1	-4.12
7.5	1215.5	-4.18
8.0	1216.7	-4.17

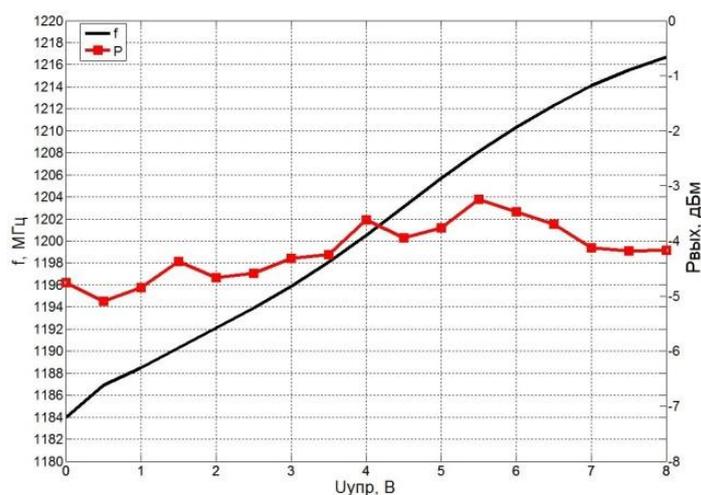


Рис. 3. Модуляционные характеристики ГУН

Таблица 2

Зависимость частоты и мощности от напряжения питания

Напряжение питания Uпит, В	Выходная частота f, МГц	Выходная мощность Pвых, дБм
4.75	1200.509	-3.96
4.8	1200.486	-3.94
4.85	1200.452	-3.87
4.9	1200.433	-3.76
4.95	1200.396	-3.81
5.0	1200.366	-3.66
5.05	1200.340	-3.62
5.1	1200.307	-3.53
5.15	1200.278	-3.58
5.2	1200.245	-3.37
5.25	1200.214	-3.31

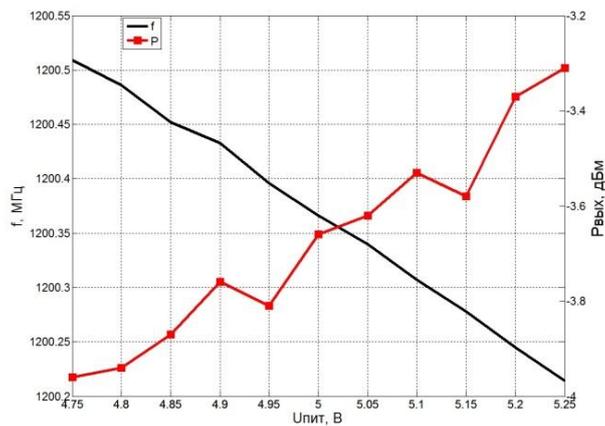


Рис. 4. Зависимость частоты и мощности от напряжения питания

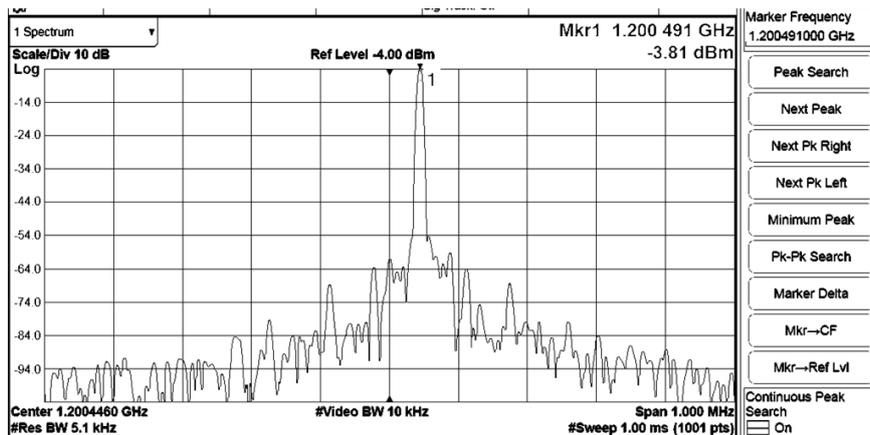


Рис. 5. Спектр выходного сигнала ГУН в ближней зоне

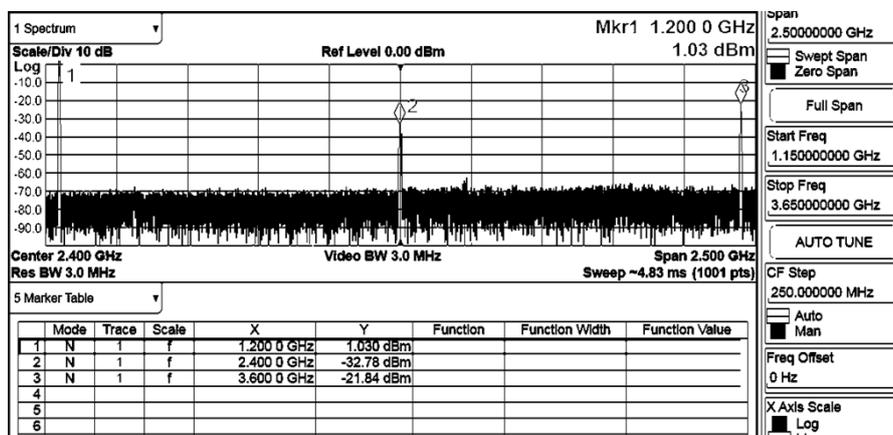


Рис. 6. Спектр гармоник выходного сигнала ГУН

Выводы. Данные табл. 3 позволяют сравнить заданные и достигнутые параметры ГУН.

Таблица 3

Основные параметры ГУН

Наименование параметра, размерность	Задано	Измерено
Номинальное значение частоты, МГц	1200	1200 ± 16
Диапазон электронной перестройки частоты, МГц	± 12	± 16
Выходная мощность, не менее, дБм	0	1
Нестабильность частоты от напряжения питания, при изменении напряжения питания на ± 5 % от номинального значения, не более, ppm	± 250	± 125
Ширина спектра выходного сигнала на уровне минус 3 дБ от максимума, кГц	–	10
Средняя крутизна модуляционной характеристики по частоте при изменении управляющего напряжения, МГц/В	–	4
Средняя крутизна модуляционной характеристики по частоте при изменении питающего напряжения, МГц/В	–	0,59
Уровень второй гармоники по отношению к первой, дБ	–	– 33
Уровень третьей гармоники по отношению к первой, дБ	–	– 22

Из таблицы видно, что все требования к генератору выполняются, с превышением. Генератор имеет высокий технологический запас и возможность модернизации.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Michal Odyniec*. RF and Microwave Oscillator Design. Artech House, Boston, London, 2002. – 398 p.
2. *Pieter L.D. Abrie*. Design of RF and Microwave Amplifiers and Oscillators. Artech House, Boston, London, 1999. – 480 p.
3. *Ali Hajimiri, Thomas H. Lee*. The Design of Low Noise Oscillators. Kluwer Academic Publishers, New York et al, 2003. – 207 p.
4. *Andrei Grebennikov*. RF and Microwave Transistor Oscillator Design. John Wiley & Sons, Ltd. – 2007.
5. *Соколинский В.Г., Шейнкман В.Г.* Частотные и фазовые модуляторы и манипуляторы. – М.: Радио и связь, 1983. – 192 с.
6. *Кочемасов В.Н., Белов Л.А., Оконешиников В.С.* Формирование сигналов с линейной частотной модуляцией. – М.: Радио и связь, 1983. – 192 с.
7. *Верещагин Е.М.* Модуляция в генераторах СВЧ. – М.: Сов. радио, 1972. – 304 с.
8. *Фаго Ж., Мань Ф.* Частотная модуляция. Теория и применение в радиорелейных линиях. – М.: Сов. радио, 1964. – 695 с.
9. *Зикий А.Н., Помазанов А.В.* Передатчики помех современным средствам связи: учеб. пособие. – Ростов-на-Дону – Таганрог: Изд-во ЮФУ, 2020. – 176 с.
10. *Белов Л.А.* Устройства формирования СВЧ сигналов и их компоненты: учеб. пособие. – М.: Издательский дом МЭИ, 2010. – 320 с.
11. *Зикий А.Н., Давтян А.Д.* Генератор, управляемый напряжением // Фундаментальные и прикладные научные исследования: инноватика в современном мире. Уфа, март 2020. – С. 23-29.
12. *Андрианов А.В., Давтян А.Д., Зикий А.Н.* Генератор ЧМ сигналов // Инженерный вестник Дона. – 2019. – № 2. – URL: ivdon.ru/ru/magazine/archive/n2y2019/5722/.
13. *Андрианов А.В., Зикий А.Н., Падалко А.Д.* Задающий генератор для передатчика скользких помех // Сб. научных трудов 14 международной конференции «Наука России: цели и задачи». Ч. 4. – Екатеринбург: НИЦ «Л-Журнал», 2019. – С. 5-8.
14. *Горевой А.В.* Генераторы диапазона 1–2 ГГц с повышенной крутизной регулировочной характеристики // Электроника, информационные технологии, радиотехника и связь: Доклады ТУСУРа. – 2011. – № 1923. – С. 44-49.
15. *Мальшиев В.М.* и др. Модель варикапа для разработки сверхширокополосных перестраиваемых генераторов СВЧ // Научно-технические ведомости СПбГПУ. – 2014. – № 2 (193). – С. 55-60.
16. *Никитин А.Б., Хабитуева Е.И.* Особенности разработки сверхширокополосных перестраиваемых генераторов СВЧ-диапазона в гибридном исполнении // Научно-технические ведомости СПбГПУ. – 2017. – Т. 10, № 4. – С. 41-50.
17. *Мазеев Е.В., Фурсаев М.А.* Проектирование СВЧ транзисторного генератора с варакторной перестройкой частоты и моделирование его электрических характеристик // Вестник СГТУ. – 2011. – № 3(57). – Вып. 1. – С. 68-73.
18. *Романюк В.* Проектирование СВЧ-генератора, управляемого напряжением // Современная электроника. – 2010. – № 2. – С. 46-51.
19. СВЧ генератор ГУН382-1-B-1-1200M-5. Этикетка завода «Метеор». – 2020. – 2 с.
20. Стратегия повышения надёжности. Каталог АО «Завод «Метеор». – 2020. – 84 с.

REFERENCES

1. *Michal Odyniec*. RF and Microwave Oscillator Design. Artech House, Boston, London, 2002, 398 p.
2. *Pieter L.D. Abrie*. Design of RF and Microwave Amplifiers and Oscillators. Artech House, Boston, London, 1999, 480 p.
3. *Ali Hajimiri, Thomas H. Lee*. The Design of Low Noise Oscillators. Kluwer Academic Publishers, New York et al, 2003, 207 p.
4. *Andrei Grebennikov*. RF and Microwave Transistor Oscillator Design. John Wiley & Sons, Ltd, 2007.
5. *Sokolinskiy V.G., Sheynkman V.G.* Chastotnye i fazovye modulyatory i manipulyatory [Frequency and phase modulators and manipulators]. Moscow: Radio i svyaz', 1983, 192 p.

6. Kochemasov V.N., Belov L.A., Okoneshnikov V.S. Formirovanie signalov s lineynoy chastotnoy modulyatsiey [Formation of signals with linear frequency modulation]. Moscow: Radio i svyaz', 1983, 192 p.
7. Vereshchagin E.M. Modulyatsiya v generatorakh SVCh [Modulation in microwave generators]. Moscow: Sov. radio, 1972, 304 p.
8. Fago Zh., Man' F. Chastotnaya modulyatsiya. Teoriya i primeneniye v radioreleynykh liniyakh [Frequency modulation. Theory and application in radio relay lines]. Moscow: Sov. radio, 1964, 695 p.
9. Zikiy A.N., Pomazanov A.V. Peredatchiki pomekh sovremennym sredstvam svyazi: ucheb. posobie [Transmitters of interference to modern means of communication: training manual]. Rostov-on-Don – Taganrog: Izd-vo YuFU, 2020, 176 p.
10. Belov L.A. Ustroystva formirovaniya SVCh signalov i ikh komponenty: ucheb. posobie [Devices for forming microwave signals and their components: textbook]. Moscow: Izdatel'skiy dom MEL, 2010, 320 p.
11. Zikiy A.N., Davtyan A.D. Generator, upravlyaemy napryazheniem [Generator controlled by voltage. In the collection], *Fundamental'nye i prikladnye nauchnye issledovaniya: innovatika v sovremennom mire* [Fundamental and applied scientific research: Innovation in the modern world]. Ufa, mart 2020, pp. 23-29.
12. Andrianov A.V., Davtyan A.D., Zikiy A.N. Generator ChM signalov [FM signal generator], *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don], 2019, No. 2. Available at: ivdon.ru/ru/magazine/archive/n2y2019/5722/.
13. Andrianov A.V., Zikiy A.N., Padalko A.D. Zadayushchiy generator dlya peredatchika skol'zyashchikh pomekh [Setting generator for a sliding interference transmitter], *Sb. nauchnykh trudov 14 mezhdunarodnoy konferentsii «Nauka Rossii: tseli i zadachi»* [Collection of scientific papers of the 14th International Conference "Science of Russia: goals and objectives"]. Part 4. Ekaterinburg, NITS «L-Zhurnal», 2019, pp. 5-8.
14. Gorevoy A.V. Generatory diapazona 1–2 GGts s povyshennoy krutiznoy regulirovochnoy kharakteristiki [Generators of the range 1-2 GHz with increased steepness of the adjustment characteristics], *Elektronika, informatsionnye tekhnologii, radiotekhnika i svyaz': Doklady TUSURa* [Electronics, information technologies, radio engineering and communications. Reports of TUSUR], 2011, No. 1923, pp. 44-49.
15. Malyshev V.M. i dr. Model' varikap dlya razrabotki sverkhshirokopolosnykh prestraivaemykh generatorov SVCh [Varicap model for the development of ultra-wideband tunable microwave generators], *Nauchno-tekhnicheskie vedomosti SPbGPU* [Scientific and technical vedomosti SPbGPU], 2014, No. 2 (193), pp. 55-60.
16. Nikitin A.B., Khabituieva E.I. Osobennosti razrabotki sverkhshirokopolosnykh prestraivaemykh generatorov SVCh-diapazona v gibridnom ispolnenii [Features of the development of ultra-wideband tunable microwave generators in hybrid design], *Nauchno-tekhnicheskie vedomosti SPbGPU* [Scientific and Technical Bulletin of SPbGPU], 2017, Vol. 10, No. 4, pp. 41-50.
17. Mazeev E.V., Fursaev M.A. Proektirovanie SVCh tranzistornogo generatora s varaktor-noy perestroikoy chastoty i modelirovanie ego elektricheskikh kharakteristik [Design of a microwave transistor generator with varactor frequency tuning and modeling of its electrical characteristics], *Vestnik SGTU* [Bulletin of SSTU], 2011, No. 3 (57), Issue 1, pp. 68-73.
18. Romanyuk V. Proektirovanie SVCh-generatora, upravlyaemogo napryazheniem [Designing a voltage-controlled microwave generator], *Sovremennaya elektronika* [Modern electronics], 2010, No. 2, pp. 46-51.
19. SVCh generator GUN382-1-V-1-1200M-5. Etiketka zavoda «Meteor» [Microwave generator GUN382-1-V-1-1200M-5. Label of the plant "Meteor"], 2020, 2 p.
20. Strategiya povysheniya nadezhnosti. Katalog AO «Zavod «Meteor» [Strategy for improving reliability. Catalog of JSC "Plant "Meteor"], 2020, 84 p.

Статью рекомендовал к опубликованию к.т.н. М.И. Дулин.

Зикий Анатолий Николаевич – Южный федеральный университет; e-mail: zikiy50@mail.ru; г. Таганрог, Россия; кафедра информационной безопасности телекоммуникационных систем; к.т.н.; с.н.с.; доцент.

Кочубей Алексей Сергеевич – Таганрогский научно-исследовательский институт связи; e-mail: l.co4ubey@yandex.ru; г. Таганрог, Россия; тел.: 89896270939; магистр, инженер-конструктор 2 категории.

Zikiy Anatoliy Nikolaevich – Southern Federal University; e-mail: zikiy50@mail.ru; Taganrog, Russia; the department of information security of telecommunication systems; cand. of eng. sc.; senior researcher; associate professor.

Kochubey Alexey Sergeevich – Taganrog Scientific Research Institute of Communications; e-mail: l.co4ubey@yandex.ru; Taganrog, Russia; phone: +79896270939; Master, design engineer of the 2nd category.

УДК 621.396.677.49

DOI 10.18522/2311-3103-2021-2-91-104

А.О. Касьянов

РЕЗУЛЬТАТЫ ЧИСЛЕННОГО ИССЛЕДОВАНИЯ ХАРАКТЕРИСТИК РАССЕЯНИЯ АНТЕННЫХ ОБТЕКАТЕЛЕЙ НА ОСНОВЕ МЕТАЛЛОДИЭЛЕКТРИЧЕСКИХ ДИФРАКЦИОННЫХ РЕШЕТОК

Статья посвящена анализу результатов численного исследования характеристик рассеяния печатного пространственного и частотно-избирательного антенного обтекателя на основе электродинамического моделирования частотно-избирательных поверхностей с диэлектрическими укрытиями. Задачи моделирования решены методом интегрального уравнения. С помощью разработанной математической модели выполнено всестороннее численное исследование микроволновых устройств пространственной, частотной и поляризационной трансформации и селекции, выполненных в виде многоэлементных мультипланарных печатных отражательных антенных решеток. На основе проведенных исследований найдены конструктивные решения для частотно-селективных структур в интегральном исполнении, реализующих функции пространственных и частотных фильтров. Рассмотрены вопросы конструктивного синтеза многослойных планарных частотно-избирательных поверхностей для создания пространственных и частотных фильтров, интегрируемых в излучающие и распределительные системы современных многоэлементных антенных решеток в печатном исполнении. Получены численные результаты электродинамического моделирования частотно-избирательных поверхностей и угловых фильтров с диэлектрическими укрытиями, которые могут быть использованы для выбора наиболее рациональных вариантов топологий частотно-избирательных поверхностей и угловых фильтров при разработке многофункциональных обтекателей в антенных системах микроволнового диапазона. На основе полученных численных данных рассматриваются возможности применения трех вариантов плоских дифракционных решеток в качестве частотных фильтров и два варианта решеток в качестве пространственных фильтров в составе антенных обтекателей. Предложенные в работе пространственные частотные фильтры в виде мультипланарных печатных решеток, входящих в состав антенных обтекателей, предназначены для обеспечения электромагнитной совместимости близкорасположенных радиоэлектронных средств, функционирующих в смежных частотных диапазонах, и содержащих сканирующие антенные решетки, устанавливаемые под антенными обтекателями. В то же время установка в антенных обтекателях, предложенных в работе угловых фильтров, выполненных в виде многослойных печатных дифракционных решеток проходного типа, позволяет устранить появление нежелательных побочных главных максимумов в диаграммах направленности разреженных цифровых антенных решеток перспективных радиотехнических комплексов микроволнового диапазона длин волн.

Частотно-избирательная поверхность; угловой фильтр; многофункциональный антенный обтекатель; метод интегрального уравнения; многослойные диэлектрические укрытия; поляризационная матрица рассеяния.

A.O. Kasyanov

RESULTS OF NUMERICAL STUDY OF SCATTERING CHARACTERISTICS IN ANTENNA RADOMES BASED ON METAL-DIELECTRIC GRATINGS

Mathematical model of multilayered printed frequency selective surfaces with dielectric covers is presented in this paper. The model is built on the suggestion of array infinity and perfect conductivity of microstrip elements. Such printed structures can be used as frequency selective surface and as covers with controllable characteristics (for example tunable filters, adaptive radar cover, electronically switched polarizers). Full-wave analysis is executed by the integral equation method of electromagnetics. The numerical solution of an integral equation has been obtained by Galerkin's method. Unknown distribution of surface magnetic currents has been approximated by roof-top basic functions. The generalized scattering matrix method was used for simulation of multilayered printed frequency selective surface. The paper presents the compound algorithm which combines the integral equation method with the method of generalized scattering matrix. A lot of numerical examples are presented proving the algorithm effectiveness. By means of this model there were synthesized multilayer frequency selective surface as periodic arrays of the printed elements, which have arbitrary shape of reradiators. It is known, that the printed elements of special shape ensure, as rejecting and as transacting of electromagnetic waves in the given frequencies, and have neglected angular sensitivity. The results of constructive synthesis of printed frequency selective surfaces as rejecting or transmitting filters, which have neglected angular sensitivity, are represented in paper. Such an algorithm is rather flexible and multiple repeats the basic problem solution. It makes the procedure of computer code preparing much more effective and do not require to change the problem decision itself.

Frequency selective surfaces; angular filters; multifunctional antenna's radomes; integral equation method; multilayer magnetodielectric covers; radar scattering matrix.

Введение. Антенны радиотехнических систем современных надводных и подводных судов, а также летательных аппаратов в процессе эксплуатации подвергаются самым разнообразным воздействиям окружающей среды, которые могут значительно ухудшить их характеристики. Для защиты антенн от воздействия окружающей среды используются диэлектрические оболочки – радиопрозрачные антенные обтекатели, которые должны вносить минимальные искажения в излучаемое поле [1]. Это означает, что задача создания современных радиотехнических комплексов требует постоянного совершенствования техники радиопрозрачных обтекателей.

Перспективными направлениями разработки радиопрозрачных обтекателей являются широкополосные обтекатели на основе многослойных структур. Введение дифракционных решеток в состав таких многослойных антенных обтекателей позволяет, наряду с обеспечением герметичности, решать с помощью обтекателей также и задачи пространственной, частотной и поляризационной селекции [2].

Вопросы разработки поляризационно-селективных антенных обтекателей ранее рассмотрены автором на примере антенн судовых радионавигационных станций [3-9]. В настоящей работе рассмотрим возможности применения и характеристики рассеяния металлодиэлектрических антенных обтекателей с расширенными функциями в части пространственной и частотной селекции.

Целью данной работы является разработка антенных обтекателей в виде армированных диэлектриком многослойных периодических решеток металлических пластин или отверстий в перфорированных экранах, выполняющих функции устройств пространственной и частотной селекции, а также определения возможности их использования в составе перспективных радиосистем.

2. Частотно-избирательные антенные обтекатели. Обзор современного состояния исследований и разработок в области частотно-избирательных поверхностей, представляющих собой решетки металлических пластин или апертур в про-

водящем экране, приведен автором в монографии [10], где показано, что при решении задач электродинамического моделирования многослойных обтекателей в большинстве работ [1, 11] используется подход, в основу которого положена замена слоистой металлодиэлектрической структуры обтекателя каскадным соединением эквивалентных линий передачи с включенными в них сосредоточенными элементами. В то же время, автор в работах [3, 12] при разработке многослойных обтекателей судовых антенн с улучшенными характеристиками опирался на модели электродинамического уровня, при построении которых использовался метод обобщенной матрицы рассеяния.

Результаты численного исследования характеристик рассеяния плоских частотно-избирательных антенных обтекателей в виде металлодиэлектрических дифракционных решеток, полученные с помощью вычислительных программ, разработанных автором на основе метода интегральных уравнений, рассмотрены в статье [12], где показано, что с помощью таких решеток удастся решить актуальную задачу разработки наиболее технологичных антенных обтекателей для развязки близкорасположенных антенн, работающих в смежных частотных диапазонах.

2.1. Постановка задачи конструктивного синтеза частотно-избирательного антенного обтекателя. Пусть имеется два радиоэлектронных средства (РЭС), в состав которых входят приемно-передающие антенные системы. Первое РЭС работает в двух узких диапазонах частот: в окрестности частоты f_0 – 1-й поддиапазон (Δf_1), и в окрестности частоты $1,17f_0$ – 2-й поддиапазон (Δf_2). Второе РЭС – в диапазоне частот от $0,87f_0$ до $0,93f_0$ – поддиапазон заграждаемых частот (Δf_3). В первом РЭС требуется обеспечить уровень пространственной развязки его приемной антенны от мешающего излучения, возбуждаемого передающей антенной второго РЭС, не менее чем на 23 дБ. При этом потери на прохождение волн сквозь антенный обтекатель на частотах первого РЭС не должны превышать 0,5 дБ.

2.2. Результаты численного исследования характеристик рассеяния металлодиэлектрических антенных обтекателей. В работах автора [12, 13] предложено два варианта возможной конструктивной реализации частотно-избирательного металлодиэлектрического антенного обтекателя. При разработке каждого из этих вариантов учитывалось требование по обеспечению частотной развязки близкорасположенных антенных систем приемно-передающих РЭС, работающих в смежных частотных диапазонах.

Несмотря на то, что описанные в [12, 13] конструкции антенных обтекателей полностью удовлетворяют предъявляемым к ним требованиям в части частотной избирательности, присутствие диэлектрических материалов в их составе снижает эксплуатационные характеристики таких обтекателей.

Отмеченный недостаток преодолевается в антенных обтекателях, выполненных на основе частотно-избирательных поверхностей (ЧИП) волноводного типа [14–19].

Ниже в качестве примера приводятся результаты численного исследования характеристик рассеяния, в качестве которых выступают элементы матрицы рассеяния [S], ЧИП волноводного типа, выполненной в виде плоской металлической дифракционной решетки.

Для целей пространственной развязки удобно использовать дифракционную решетку из отрезков квазиквадратных волноводов с резонансными диафрагмами. Схематическое изображение фрагмента такой ЧИП показано на рис. 1. Топология её единичной ячейки изображена на рис. 2.

Как показано на рис. 1, волноводные элементы решетки 2 размещены в узлах сетки с треугольной ячейкой. При этом ячейка Флоке решетки 1 имеет размеры: $A \times B = (0,6343 \times 0,5493) \cdot \lambda_0$, где $\lambda_0 = c / f_0$ – длина волны, c – скорость света в вакууме. Размеры апертуры 3 прямоугольного волновода: $a \times b = (0,5198 \times 0,5188) \cdot \lambda_0$. Размеры апертур каждой из резонансных диафрагм: $w \times h = (0,4312 \times 0,4433) \cdot \lambda_0$.

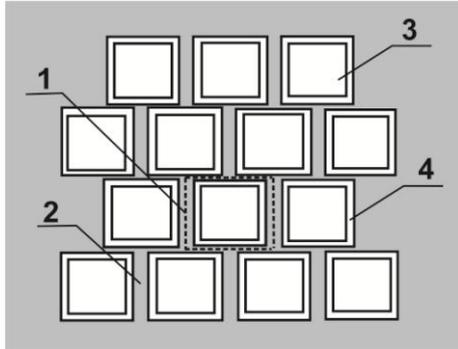


Рис. 1. Изображение фрагмента ЧИП волноводного типа в плановой проекции: 1 – канал Флоке; 2 – волноводный фланец; 3 – апертура прямоугольного волновода; 4 – резонансная диафрагма

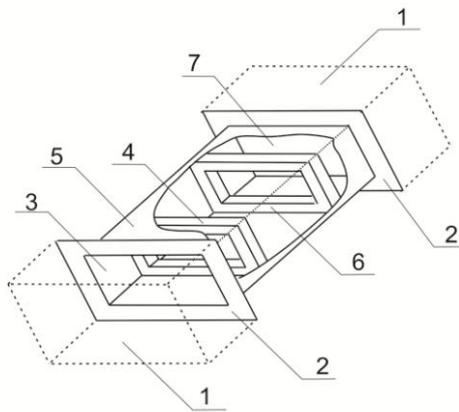


Рис. 2. Топология единичной ячейки ЧИП волноводного типа: 1 – канал Флоке; 2 – волноводный фланец; 3 – входная апертура отрезка волновода; 4 – входная резонансная диафрагма; 5 – отрезок волновода; 6 – входная резонансная диафрагма; 7 – выходная апертура отрезка волновода

Как показано на рис. 2, расстояние от апертуры отрезка прямоугольного волновода 5 до ближайшей резонансной диафрагмы: $t_1 = 1,1029 \cdot \lambda_0$. Расстояние между диафрагмами 4, расположенными в отрезке прямоугольного волновода: $t_2 = 1,0287 \cdot \lambda_0$. Толщина резонансной диафрагмы: $s = 0,0572 \cdot \lambda_0$.

На рис. 3 и 4 показаны, рассчитанные на частоте f_0 , распределения y – составляющей вектора напряженности электрического поля в двух поперечных и одной продольной плоскости единичной ячейки ЧИП волноводного типа, изображенной на рис. 2. При этом поперечные координаты x или y обозначены на рис. 3 буквой ξ , а продольная координата z – буквой ζ на рис. 4.

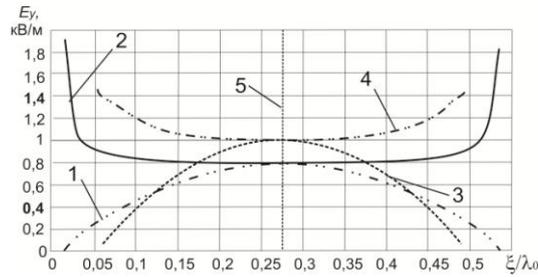


Рис. 3. Распределения y -составляющей вектора напряженности электрического поля в поперечной плоскости: 1 – в среднем сечении входной апертуры отрезка волновода для случая $\xi = x$; 2 – в той же апертуре, но при $\xi = y$; 3 – в среднем сечении ближайшей диафрагмы для случая $\xi = x$; 4 – в той же диафрагме, но при $\xi = y$; 5 – координата среднего сечения апертуры волновода / диафрагмы

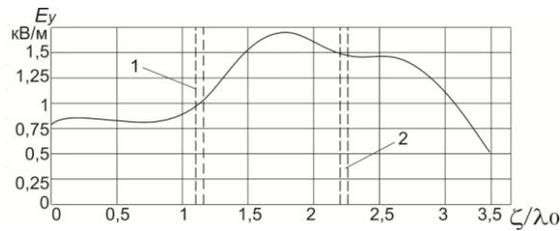


Рис. 4. Распределения y -составляющей вектора напряженности электрического поля в продольном сечении, проходящем через ось волновода: 1 – входная резонансная диафрагма; 2 – выходная диафрагма

Как видно из графиков, показанных на рис. 3, распределения поля в прямоугольных апертурах являются седловидными. При этом, как следует из рис. 4, уровень поля во входной диафрагме 4 примерно на четверть превышает уровень поля во входной апертуре 3 отрезка волновода, уровень поля в выходной диафрагме 6 примерно в полтора раза превышает уровень поля во входной диафрагме 4, а уровень поля в выходной апертуре 7 отрезка волновода 5 примерно в три раза ниже уровня поля в выходной диафрагме 6.

Частотные характеристики (ЧХ) диагональных элементов S_{pq} , $p = q$ поляризационной матрицы рассеяния (ПМР) [S] вышеописанной ЧИП показаны на рис. 5.

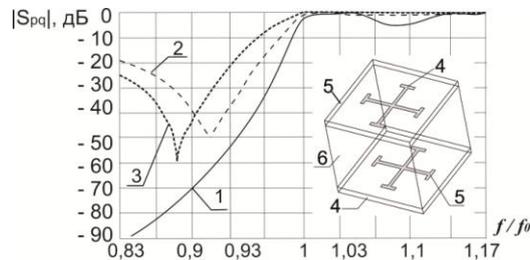


Рис. 5. Частотные характеристики элементов поляризационной матрицы рассеяния [S] ЧИП: 1 – ЧИП волноводного типа (рис. 1, 2); 2 – печатная решетка, изображенная во врезке рис. 5; 3 – печатная решетка, изображенная на рис. 6; 4 – печатные вибраторы; 5 – диэлектрические укрытия; 6 – сердцевина

Для сравнения на рис. 5 приведены ЧХ элементов ПМР печатных решеток, рассмотренных в работах автора [12, 13]. Излучатели обеих печатных решеток представляют собой планарные вибраторы, выполненные в виде иерусалимских крестов, нанесенных с обеих сторон на диэлектрическую подложку (сердцевину) и имеющих двухсторонние диэлектрические укрытия. Для решетки, топология которой изображена на врезке в рис. 5, сердцевина представляет собой четвертьволновый слой ($\lambda_0/4$), выполненный из однородного материала (Teflon 2,08). Для решетки, топология которой изображена на рис. 6, сердцевина является составной, выполненной в виде многослойной воздушно-диэлектрической структуры, эквивалентной однородному слою толщиной $\lambda_0/4$.

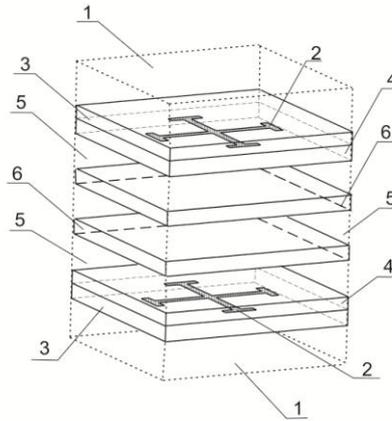


Рис. 6. Топология единичной ячейки металлдиэлектрической ЧИП: 1 – канал Флоке; 2 – печатные вибраторы; 3 – диэлектрическое укрытие; 4 – внешние слои диэлектрической сердцевины; 5 – воздушные прослойки сердцевины; 6 – внутренние слои диэлектрической сердцевины

Как видно из графиков, решетка волноводного типа по частотной избирательности значительно превосходит печатные, однако в случае её исполнения в цельнометаллическом варианте масса такой ЧИП оказывается существенно выше, чем у печатных ЧИП. Для преодоления указанного недостатка при изготовлении вышеописанной ЧИП волноводного типа потребуется применение технологии металлизации пластика.

3. Угловые фильтры. Исследуем возможность применения многослойных мультирешеточных ЧИП в качестве устройств пространственной селекции на частоте f_0 . Такие устройства называют *угловыми фильтрами* (УФ) [20–25]. Областью применения УФ являются сканирующие антенные решетки (АР) [26]. С помощью УФ удастся устранить побочные главные максимумы в ДН антенных решеток с повышенными межэлементными расстояниями. Таким образом, применение УФ способствует уменьшению числа излучателей в АР, что ведет к снижению их стоимости.

Например, УФ можно включить в состав многофункционального обтекателя судовой антенны [4], где для устранения паразитной составляющей вектора поляризации поля возбуждаемого волноводно-щелевой антенной решеткой (ВЩАР) из встречно-наклонных щелей, выполненных в узкой стенке волновода, используется поляризационный фильтр. Другим способом устранения нежелательной вертикальной составляющей поля ВЩАР является выполнение поперечных щелей на широкой стенке волновода. Однако такой излучатель имеет повышенные межэле-

ментные расстояния. При этом возрастает опасность появления побочных главных максимумов (ПГМ) в диаграмме направленности (ДН) антенны. Задачу подавления этих лепестков можно возложить на многофункциональный обтекатель. Как и в случае, когда обтекатель выполняет функции поляризационного фильтра [3–9], УФ представляет собой плоскую герметизирующую вставку, устанавливаемую в раскрытие рупора. Однако теперь эта вставка, помимо функций обтекателя, должна выполнять роль пространственного, или углового фильтра [4].

В качестве одного из вариантов конструктивной реализации такого фильтра рассмотрим типичную частотно-селективную структуру, а именно, перфорированный металлический экран. Интерес к исследованию ЧИП в качестве углового фильтра не является случайным. Дело в том, что, как отмечается в [27, 28], с одной стороны, при фиксированном угле падения поведение характеристик рассеяния углового фильтра в полосе частот ничем не отличается от аналогичных зависимостей обычной ЧИП, а, с другой стороны, при фиксированной частоте ЧИП ведет себя в секторе углов падения подобно угловому фильтру.

3.1. Постановка задачи конструктивного синтеза пространственно-избирательного антенного обтекателя. Пусть имеется мобильное РЭС, устанавливаемое на подвижном носителе. В состав данного РЭС должна входить остронаправленная антенная система (АС), предназначенная для излучения и приема электромагнитных волн в коротковолновой части сантиметрового диапазона длин волн на двух наклонных взаимно ортогональных линейных поляризациях. Необходимость установки РЭС на подвижном носителе исключает возможность применения крупногабаритной параболической антенны в качестве остронаправленной АС данного РЭС.

Как известно [26], помимо применения параболических антенн, для создания остронаправленного излучения могут быть использованы плоские антенные решетки. Например, рассмотрим возможность использования многоэлементной плоской решетки рупорных излучателей в качестве АС вышеописанного мобильного РЭС. На рис. 7 показан фрагмент такой плоской антенной решетки (АР) размером 5×3 излучателей. Как видно из рис. 7, в качестве излучателей в такой АР выступают пирамидальные рупорные антенны.

Для обеспечения требования одновременной работы на прием и передачу на двух наклонных взаимно ортогональных линейных поляризациях каждый из рупорных излучателей такой решетки должен быть двухполяризационным, как показано на рис. 8. Как видно из рис. 8, объединение/деление волн двух ортогональных поляризаций в фидерной системе каждого пирамидального рупора 1 реализуется на отрезке квадратного двухмодового волновода 2.

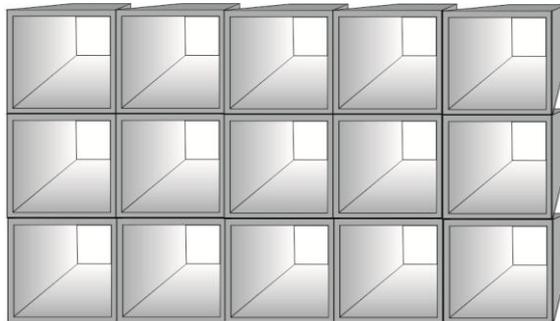


Рис. 7. Фрагмент плоской антенной решетки, составленной из рупорных излучателей

Для обеспечения высокого КПД каждый из прямоугольных волноводов 3 и 4, возбуждающих двухмодовый волновод 2, должен быть одномодовым и работать на основном типе волны.

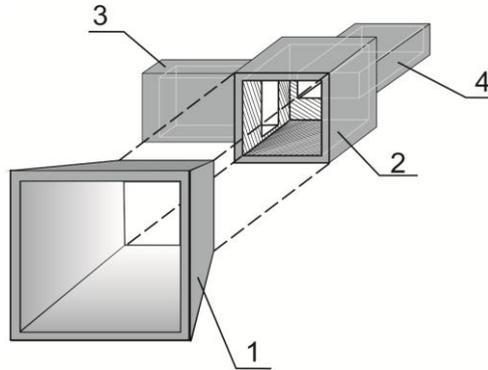


Рис. 8. Излучатель плоской двухполяризационной решетки рупорных антенн: 1 – пирамидальный рупор; 2 – отрезок квадратного волновода; 3 – отрезок прямоугольного волновода, предназначенный для возбуждения волны горизонтальной поляризации; 4 – отрезок прямоугольного волновода, предназначенный для возбуждения волны вертикальной поляризации

Всем вышеперечисленным требованиям удовлетворяет АС, имеющая следующие характеристики: апертура – прямоугольная с размерами 24×6 элементов, расположенных в узлах сетки с квадратной ячейкой; шаги решетки $A = B \approx 1,37 \cdot \lambda_0$, размеры апертуры рупора $a = b \approx 1,29 \cdot \lambda_0$, размеры горловины рупора $a_1 = b_1 \approx 0,744 \cdot \lambda_0$, длина раствора рупора $L \approx \lambda_0$, размеры апертуры отрезка квадратного волновода $a_2 = b_2 \approx 0,6659 \cdot \lambda_0$.

Сечения ДН описанной выше антенны азимутальной и угломестной плоскостями, рассчитанные на частоте f_0 , показаны соответственно на рис. 9 и рис. 10.

Как видно из графиков, показанных на рис. 9 и 10, в обеих плоскостях в ДН антенны присутствуют побочные главные максимумы при углах примерно равных 45° и 135° . Появление ПГМ в ДН антенны обусловлено тем, что данная АР является разреженной, поскольку шаги решетки в обеих плоскостях превышают рабочую длину волны λ_0 . Для устранения ПГМ в ДН антенны её обтекатель может быть выполнен в виде УФ.

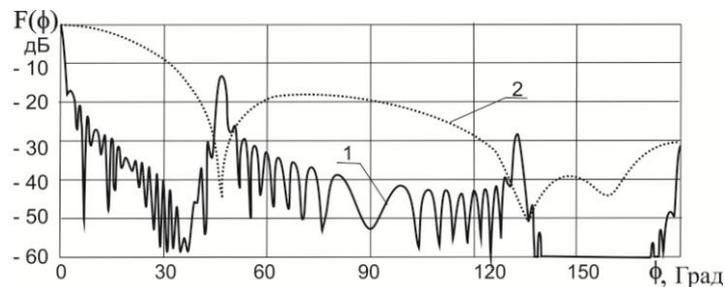


Рис. 9. Диаграмма направленности антенны в азимутальной плоскости: 1 – АР без углового фильтра; 2 – излучатель АР с угловым фильтром

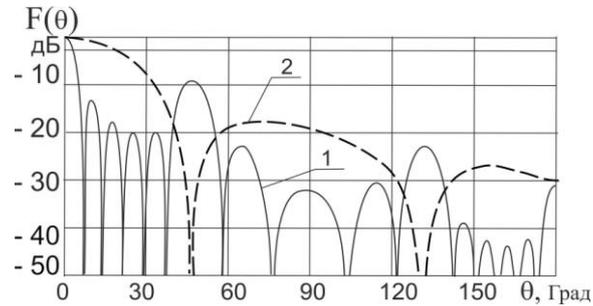


Рис. 10. Диаграмма направленности антенны в угломестной плоскости:
1 – AP без углового фильтра; 2 – излучатель AP с угловым фильтром

3.2. Результаты численного исследования характеристик направленности антенной решетки с металлодиэлектрическим обтекателем. Известно [29, 30], что при построении УФ могут быть использованы результаты конструктивного синтеза ЧИП. Опираясь на приведенные выше результаты конструктивного синтеза ЧИП, можно утверждать, что наиболее предпочтительным вариантом выполнения УФ является конструкция, в которой совмещены топологии, показанные на рис. 2 и 6. Пример конструктивной реализации такого УФ показан на рис. 11. Как видно из рис. 11, угловой фильтр является многослойным и состоит из трех решеток скрещенных металлических брусьев 3, образующих отрезки квадратных волноводов, и диэлектрических слоев укрытия 6, прокладок 5 и подложки 4.

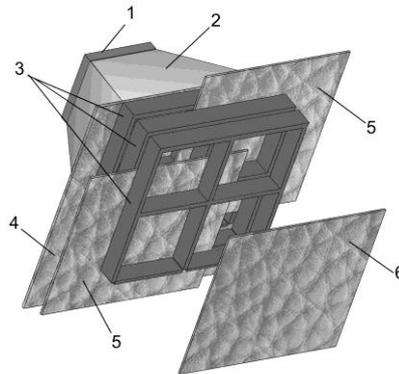


Рис. 11. Рупорный излучатель плоской антенной решетки с угловым фильтром:
1 – отрезок квадратного волновода в горловине рупора; 2 – раствор пирамидального рупора; 3 – дифракционные решетки из скрещенных металлических брусьев; 4 – диэлектрическая подложка; 5 – диэлектрические прокладки; 6 – диэлектрическое укрытие

Данный УФ имеет следующие параметры: число каскадно-включенных дифракционных решеток – 3; число диэлектрических слоев – 5; шаги решеток $A_1 = B_1 \approx 0,685 \cdot \lambda_0$; толщина брусьев $t_1 \approx 0,1567 \cdot \lambda_0$; ширина брусьев $w \approx 0,039267 \cdot \lambda_0$; толщины диэлектрического укрытия и каждой из прокладок $t_2 \approx 0,01 \cdot \lambda_0$; диэлектрическая подложка состоит из двух диэлектрических слоев, выполненных из такого же материала, как укрытие и прокладки. В качестве материала диэлектрических слоев УФ можно использовать Arlon AD410 (материал 1) или Arlon AD600 (материал 2).

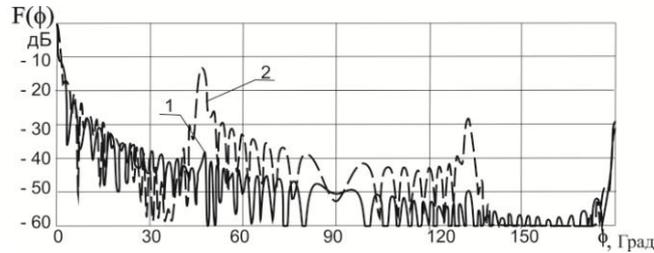


Рис. 12. Диаграмма направленности антенной решетки в азимутальной плоскости: 1 – АР без УФ; 2 – АР с УФ

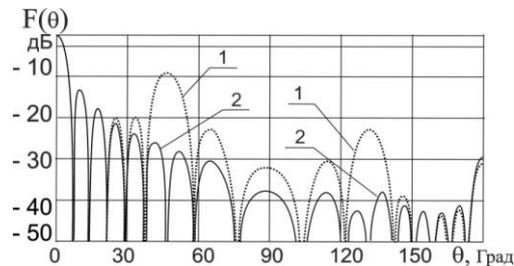


Рис. 13. Диаграмма направленности антенной решетки в угломестной плоскости: 1 – АР без УФ; 2 – АР с УФ

На рис. 9 и 10 показаны ДН одиночного излучателя из состава АР, в качестве антенных обтекателей которого выступают угловые фильтры, имеющие одинаковую конструкцию, показанную на рис. 11. При этом между собой УФ различаются только тем, что диэлектрические слои УФ в случае, изображенном на рис. 9, выполнены из материала 1, а в случае, изображенном на рис. 10, из материала 2. Как видно из графиков, показанных на рис. 12 и 13, введение УФ в состав АР позволяет подавить нежелательные ПГМ в её ДН.

Из приведенных результатов численных экспериментов следует, что размещением перфорированных экранов внутри диэлектрика и выбором параметров последнего можно синтезировать угловые характеристики такой электродинамической структуры, приближая их к виду, необходимому для реализации функций пространственного фильтра.

Заключение. На основе численных исследований характеристик рассеяния многоэлементных мультипланарных решеток печатных элементов выработаны рекомендации по конструкционным параметрам многослойных ЧИП и УФ, выступающих в качестве многофункциональных антенных обтекателей антенной решетки приемо-передающего РЭС. Проведенные численные исследования дают право утверждать, что использование микрополосковых отражательных антенных решеток с печатными элементами, имеющими сложную топологию на многослойных магнетодиэлектрических подложках, открывает дополнительные возможности по частотной и пространственной фильтрации полей прошедших сквозь такие решетки волн, обеспечивая возможность развязки источников, имеющих близкие рабочие частоты и подавление ПГМ в ДН разреженных АР, что позволяет рекомендовать их для использования в составе обтекателей перспективных РЭС.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Каплун В.А.* Обтекатели антенн СВЧ (Радиотехнический расчет и проектирование). – М.: Сов. радио, 1974. – 240 с.
2. *Касьянов А.О.* Частотно-избирательные поверхности. Методы проектирования и области применения: монография. – Таганрог: Изд-во ЮФУ, 2019. – 150 с.
3. *Касьянов А.О., Обуховец В.А.* Отражательные антенные решетки как микроволновые компоненты интеллектуальных покрытий // Антенны. – 2001. – Вып. 4 (50). – С. 12-16.
4. *Касьянов А.О., Обуховец В.А.* Численное исследование многофункциональных обтекателей антенн судовых РЛС // Матер. 3-й МНТК «Перспективные технологии в средствах передачи информации» – (ПТСПИ'99). – Владимир, 1999. – С. 160-166.
5. *Касьянов А.О., Кошкидько В.Г.* Применение электродинамических методов анализа при разработке антенных систем судовых РЛС // Известия ТРТУ. – 2000. – № 1 (15). – С. 19.
6. *Касьянов А.О.* Обтекатель антенны судовой радиолокационной станции // Тр. Международной научной конференции "Излучение и рассеяние электромагнитных волн" ИРЭМВ-2003, 16-20 июня 2003 г. – Таганрог: ТРТУ. – С. 96-99.
7. *Касьянов А.О., Обуховец В.А.* «Дифракционные решетки как устройства пространственно-частотной и поляризационной селекции» // Сб. трудов III Международного симпозиума «Конверсия науки – международному сотрудничеству» (Сибконверс'99). – Томск, 1999. – С. 406-408.
8. *Kasyanov A.O., Obukhovets V.A.* Polarizing filters, converters and modulators based on controllable microstrip diffraction arrays // Proceedings of III International Seminar/Workshop on DIPED'98, 2-5 November 1998, Tbilisi. – P. 46-49.
9. *Касьянов А.О., Обуховец В.А.* Поляризационные фильтры, преобразователи и модуляторы на основе управляемых микрополосковых дифракционных решеток // Всерос. конф. «Компьютерные технологии в инженер. и управлен. деятельности»: Сб. тр. Ч. 2. – Таганрог, 1999. – С. 121-126.
10. *Касьянов А.О., Обуховец В.А.* Частотно-избирательные поверхности. Основные области применения // Антенны. – 2005. – Вып. 9 (100). – С. 4-12.
11. *Красюк В.Н.* Антенны СВЧ с диэлектрическими покрытиями – Л.: Судостроение, 1986.
12. *Касьянов А.О.* Математическое моделирование и расчет характеристик рассеяния печатного частотно-избирательного антенного обтекателя // Известия ЮФУ. Технические науки. – 2020. – № 6 (216). – С. 129-139.
13. *Касьянов А.О.* Результаты численного исследования характеристик рассеяния печатного частотно-избирательного антенного обтекателя // Матер. 30-й Международной Крымской конференции «СВЧ-техника и телекоммуникационные технологии», 7-12 сентября 2020 г. – Севастополь, Крым: Изд-во СевГУ, 2020. – С. 165-166.
14. *Besso P., Bozzi M., Perregrini L., Salghetti Drioli L., and Nickerson W.* Deep Space Antenna for Rosetta Mission: Design and Testing of the S/X-Band Dichroic Mirror // IEEE Trans. Antennas Propag. – March 2003. – Vol. AP-51, No. 3. – P. 388-394.
15. *Bozzi M., Perregrini L., Weinzierl J., and Winnewisser C.* Efficient Analysis of Quasi-Optical Filters by a Hybrid MoM/BI-RME Method // IEEE Trans. Antennas Propag. – July 2001. – Vol. AP-49, No. 7. – P. 1054-1064.
16. *Bozzi M., Perregrini L., Weinzierl J., and Winnewisser C.* Design, Fabrication, and Measurement of Frequency-Selective Surfaces // Optical Engineering. – Aug. 2000. – Vol. 39, No. 8. – P. 2263-2269.
17. *Winnewisser C., Lewen F., Schall M., Walther M., and Helm H.* Characterisation and Application of Dichroic Filters in the 0.1 to 2 THz Region // IEEE Trans. Microwave Theory & Techniques. – April 2000. – Vol. MTT-48, No. 4. – P. 744-749.
18. *Besso P., Bozzi M., Brenner M., Junker S., Madde R., Perregrini L., and Salghetti Drioli L.* Electrical and Mechanical Performance of a S/X/Ka-Band Dichroic Mirror // Proc. of EuMC. Oct. 3-5, 2000. – Paris, France, 2000. – P. 290-293.
19. *Chen J.C.* Analysis of a thick dichroic plate with rectangular holes at arbitrary angles of incidence," TDA Progress Report 42-104, October-December 1990, Jet Propulsion Laboratory, California Institute of Technology.
20. *Franchi P.R. and Mailloux R.J.* Theoretical and experimental study of metal grid angular filters for sidelobe suppression // IEEE Trans. Antennas Propagat. – May 1983. – Vol. 31. – P. 445-450.

21. *Biosca J., Llorens D., and Vigano M.C.* Side-lobe reduction with overlapped beam-forming network for Ku-band hybrid antenna array // in Proc. 10th Eur. Conf. Antennas Propag., Davos, Switzerland, 2016. – P. 1-5.
22. *Lee Y.J., Jeong S.H., Park W.S., Yun J.S., and Jeon S.I.* Multilayer spatial angular filter with air gap tuners to suppress grating lobes of 4 x 1 array antenna // Electron. Lett. – Jan. 2003. – Vol. 39, No. 1. – P. 15-17.
23. *Martinez-Ros A.J., Gomez-Tornero J.L., and Goussetis G.* Multifunctional angular bandpass filter SIW leaky-wave antenna // IEEE Antennas Wireless Propag. Lett. – 2017. – Vol. 16. – P. 936-939.
24. *Mailloux R.J.* Synthesis of spatial filters with Chebyshev characteristics // IEEE Trans. Antennas Propagat. – Mar. 1976. – Vol. AP-24. – P. 174-181.
25. *Rope E.L., Tricoles G., Yue O.C.* Metallic, angular filters for array economy // in IEEE Antennas Propagat. Int. Symp. Dig. – 1976. – P. 155-157.
26. *Хансен Р.* Сканирующие антенные системы СВЧ. Т. 1-2. – М.: Советское радио, 1966. – Т. 1. – 536 с. – Т. 2. – 496 с.
27. *Rope E.L., Tricoles G., Yue O.C.* Metallic, Angular Filters For Arrays Economy // AP-S Session 5, 1610, Tuesday, October 12, Room 163-173. – P. 155-157.
28. *Kinowski D., Guglielmi M., Roederer A.G.* Angular Bandpass Filters: An Alternative Viewpoint Gives Improved Design Flexibility // IEEE Trans. Antennas and Propag. – 1995. – Vol. AP43, No. 4. – P. 390-395.
29. *Касьянов А.О., Обуховец В.А.* Металлодиэлектрические частотно-избирательные поверхности // Электромагнитные волны и электронные системы. Радиотехника. – 2009. – Т. 14, № 11. – С. 29-38.
30. *Касьянов А.О., Касьянова А.Н.* Электродинамический анализ и разработка САПР-ориентированных математических моделей печатных антенных решеток: монография. – Таганрог: Изд-во ЮФУ, 2017. – 300 с.

REFERENCES

1. *Kaplun V.A.* Obtekateli antenn SVCh (Radiotekhnicheskii raschet i proektirovanie) [Fairings of microwave antennas (Radio engineering calculation and design)]. Moscow: Sov. radio, 1974, 240 p.
2. *Kas'yanov A.O.* Chastotno-izbiratel'nye poverkhnosti. Metody proektirovaniya i oblasti primeneniya: monografiya [Frequency-selective surfaces. Design methods and application areas: monograph]. Taganrog: Izd-vo YuFU, 2019, 150 p.
3. *Kas'yanov A.O., Obukhovets V.A.* Otrazhatel'nye antennnye reshetki kak mikrovolnovye komponenty intellektual'nykh pokrytiy [Reflective antenna arrays as microwave components of smart coatings], *Antenny* [Antennas], 2001, Issue 4 (50), pp. 12-16.
4. *Kas'yanov A.O., Obukhovets V.A.* Chislennoe issledovanie mnogo-funktional'nykh obtekatel'nykh antenn sudovykh RLS [Numerical study of multifunctional radome antennas], *Mater. 3-y MNTK «Perspektivnye tekhnologii v sredstvakh peredachi informatsii» – (PTSPI'99)* [Materials of the 3rd ISTC "Perspective technologies in the media of information transmission" – (PTSPI'99)]. Vladimir, 1999, pp. 160-166.
5. *Kas'yanov A.O., Koshkid'ko V.G.* Primenenie elektrodinamicheskikh metodov analiza pri razrabotke antennnykh sistem sudovykh RLS [Application of electrodynamic methods of analysis in the development of ship radar antenna systems], *Izvestiya TRTU* [Izvestiya TSURE], 2000, No. 1 (15), pp. 19.
6. *Kas'yanov A.O.* Obtekatel' antenny sudovoy radiolokatsionnoy stantsii [Fairing of the ship radar station antenna], *Tr. Mezhdunarodnoy nauchnoy konferentsii "Izsluchenie i rasseyaniye elektromagnitnykh voln" IREMV-2003, 16-20 iyunya 2003 g.* [Proceedings of the International Scientific Conference "Radiation and scattering of electromagnetic waves" IREMV-2003, June 16-20, 2003]. Taganrog: TRTU, pp. 96-99.
7. *Kas'yanov A.O., Obukhovets V.A.* Difraktsionnye reshetki kak ustroystva prostranstvenno-chastotnoy i polarizatsionnoy selektsii [Diffraction gratings as devices of spatial-frequency and polarizing selection], *Sb. trudov III Mezhdunarodnogo simpoziuma «Konversiya nauki – mezhdunarodnomu sotrudnichestvu» (Sibkonvers'99)* [Proceedings of the III International Symposium "Conversion of science to International cooperation" (Sibconverse '99). Tomsk, 1999, pp. 406-408.

8. Kasyanov A.O., Obukhovets V.A. Polarizing filters, converters and modulators based on controllable microstrip diffraction arrays, *Proceedings of III International Seminar/Workshop on DIPED'98, 2-5 November 1998, Tbilisi*, pp. 46-49.
9. Kas'yanov A.O., Obukhovets V.A. Polyarizatsionnye fil'try, preobrazovateli i modulyatory na osnove upravlyaemykh mikropoloskovykh difraktsionnykh reshetok [Polarized filters, converters and modulators based on controlled microstrip diffraction gratings], *Vseros. konf. «Komp'yuternye tekhnologii v inzhener. i upravlen. deyatel'nosti»: Sb. tr.* [All-Russian Conference "Computer Technologies in Engineering and Business Management": Collection of works]. Part 2. Taganrog, 1999, pp. 121-126.
10. Kas'yanov A.O., Obukhovets V.A. Chastotno-izbiratel'nye poverkhnosti. Osnovnye oblasti primeneniya [Frequency-selective surfaces. Main application areas], *Antenny* [Antennas], 2005, Issue 9 (100), pp. 4-12.
11. Krasnyuk V.N. Antenny SVCh s dielektricheskimi pokrytiyami [Microwave antennas with dielectric coatings]. Leningrad: Sudostroenie, 1986.
12. Kas'yanov A.O. Matematicheskoe modelirovanie i raschet kharakteristik rasseyaniya pechatnogo chastotno-izbiratel'nogo antennogo obtekatelya [Mathematical modeling and calculation of the scattering characteristics of a printed frequency-selective antenna fairing], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2020, No. 6 (216), pp. 129-139.
13. Kas'yanov A.O. Rezul'taty chislennogo issledovaniya kharakteristik rasseyaniya pechatnogo chastotno-izbiratel'nogo antennogo obtekatelya [Results of a numerical study of the scattering characteristics of a printed frequency-selective antenna fairing], *Mater. 30-y Mezhdunarodnoy Krymskoy konferentsii «SVCh-tekhnika i telekommunikatsionnye tekhnologii», 7-12 sentyabrya 2020 g.* [Proceedings of the 30th International Crimean Conference "Microwave Technology and Telecommunications Technologies", September 7-12, 2020]. Sevastopol', Krym: Izd-vo SevGU, 2020, pp. 165-166.
14. Besso P., Bozzi M., Perregrini L., Salghetti Drioli L., and Nickerson W. Deep Space Antenna for Rosetta Mission: Design and Testing of the S/X-Band Dichroic Mirror, *IEEE Trans. Antennas Propag.*, March 2003, Vol. AP-51, No. 3, pp. 388-394.
15. Bozzi M., Perregrini L., Weinzierl J., and Winnewisser C. Efficient Analysis of Quasi-Optical Filters by a Hybrid MoM/BI-RME Method, *IEEE Trans. Antennas Propag.*, July 2001, Vol. AP-49, No. 7, pp. 1054-1064.
16. Bozzi M., Perregrini L., Weinzierl J., and Winnewisser C. Design, Fabrication, and Measurement of Frequency-Selective Surfaces, *Optical Engineering*, Aug. 2000, Vol. 39, No. 8, pp. 2263-2269.
17. Winnewisser C., Lewen F., Schall M., Walther M., and Helm H. Characterisation and Application of Dichroic Filters in the 0.1 to 2 THz Region, *IEEE Trans. Microwave Theory & Techniques*, April 2000, Vol. MTT-48, No. 4, pp. 744-749.
18. Besso P., Bozzi M., Brenner M., Junker S., Madde R., Perregrini L., and Salghetti Drioli L. Electrical and Mechanical Performance of a S/X/Ka-Band Dichroic Mirror, *Proc. of EuMC. Oct. 3-5, 2000*. Paris, France, 2000, pp. 290-293.
19. Chen J.C. Analysis of a thick dichroic plate with rectangular holes at arbitrary angles of incidence," TDA Progress Report 42-104, October-December 1990, Jet Propulsion Laboratory, California Institute of Technology.
20. Franchi P.R. and Mailloux R.J. Theoretical and experimental study of metal grid angular filters for sidelobe suppression, *IEEE Trans. Antennas Propag.*, May 1983, Vol. 31, pp. 445-450.
21. Biosca J., Llorens D., and Vigano M.C. Side-lobe reduction with overlapped beam-forming network for Ku-band hybrid antenna array, in *Proc. 10th Eur. Conf. Antennas Propag., Davos, Switzerland, 2016*, pp. 1-5.
22. Lee Y.J., Jeong S.H., Park W.S., Yun J.S., and Jeon S.I. Multilayer spatial angular filter with air gap tuners to suppress grating lobes of 4 x 1 array antenna, *Electron. Lett.*, Jan. 2003, Vol. 39, No. 1, pp. 15-17.
23. Martinez-Ros A.J., Gomez-Tornero J.L., and Goussetis G. Multifunctional angular bandpass filter SIW leaky-wave antenna, *IEEE Antennas Wireless Propag. Lett.*, 2017, Vol. 16, pp. 936-939.
24. Mailloux R.J. Synthesis of spatial filters with Chebyshev characteristics, *IEEE Trans. Antennas Propag.*, Mar. 1976, Vol. AP-24, pp. 174-181.

25. *Rope E.L., Tricoles G., Yue O.C.* Metallic, angular filters for array economy, in *IEEE Antennas Propagat. Int. Symp. Dig.*, 1976, pp. 155-157.
26. *Khansen R.* Skaniruyushchie antennnye sistemy SVCh [Scanning antenna systems of microwave]. Vol. 1-2. Moscow: Sovetskoe radio, 1966, Vol. 1, 536 p., Vol. 2, 496 p.
27. *Rope E.L., Tricoles G., Yue O.C.* Metallic, Angular Filters For Arrays Economy, *AP-S Session 5, 1610, Tuesday, October 12, Room 163-173*, pp. 155-157.
28. *Kinowski D., Guglielmi M., Roederer A.G.* Angular Bandpass Filters: An Alternative Viewpoint Gives Improved Design Flexibility, *IEEE Trans. Antennas and Propag.*, 1995, Vol. AP43, No. 4, pp. 390-395.
29. *Kas'yanov A.O., Obukhovets V.A.* Metallodielektricheskie chastotno-izbiratel'nye poverkhnosti [Metal-dielectric frequency-selective surfaces], *Elektromagnitnye volny i elektronnye sistemy. Radiotekhnika* [Electromagnetic waves and electronic systems. Radio engineering], 2009, Vol. 14, No. 11, pp. 29-38.
30. *Kas'yanov A.O., Kas'yanova A.N.* Elektrodinamicheskiy analiz i razrabotka SAPR-orientirovannykh matematicheskikh modeley pechatnykh antenykh reshetok: monografiya [Electrodynamics analysis and development of CAD-oriented mathematical models of printed antenna arrays: monograph]. Taganrog: Izd-tvo YuFU, 2017, 300 p.

Статью рекомендовал к опубликованию д.т.н., профессор Д.Д. Габриэлян.

Касьянов Александр Олегович - Южный федеральный университет, e-mail: kasyanovao@sfedu.ru; г. Ростов-на-Дону, Россия; кафедра радиотехнических и телекоммуникационных систем, д.т.н., профессор.

Kasyanov Alexander Olegovich – Southern Federal University; e-mail: kasyanovao@sfedu.ru; Rostov-on-Don, Russia; the department of radio engineering and telecommunication systems; dr. of eng. sc.; professor

УДК 004.657

DOI 10.18522/2311-3103-2021-2-104-112

П.А. Курапов

ГИБРИДНОЕ ИСПОЛНЕНИЕ ЗАПРОСОВ К АНАЛИТИЧЕСКИМ БАЗАМ ДАННЫХ

Для повышения эффективности системам исполнения аналитических запросов необходимо использовать все доступные ресурсы современных распределенных гетерогенных систем. Ускорители, сложная иерархия памяти и распределенность вычислений создают возможности для оптимизации производительности. В статье проводится обзор существующих подходов к реализации механизмов исполнения аналитических запросов к СУБД для данных в оперативной памяти с использованием аппаратных ускорителей, в частности, особое внимание уделено графическим ускорителям. За счет массивного параллелизма и высокой пропускной способности памяти устройства графические ускорители представляют перспективную альтернативу основному устройству исполнения аналитических запросов. Существующие методы не задействуют всех возможностей современной аппаратуры и обычно основываются на передаче данных по относительно медленной шине PCIe для исполнения ядер каждого отдельно взятого оператора. Другой проблемой существующих методов является явное разделение кодовой базы кодогенераторов реляционных операторов для ускорителей (графических процессоров) и центрального процессора, и невозможность переиспользования сгенерированного кода для других устройств в системе, что существенно ограничивает возможности их совместного использования с целью повышения производительности. В статье представлен метод эффективного исполнения запросов на примере системы из двух классов устройств (ЦПУ и графический процессор) при помощи компиляции с построением единого, независимого от устройства, промежуточного представления (SPIR-V) и подход к оптимизации соответствующего гибридного физического плана запроса на основе расширенного классического оператора "Exchange" с использованием гетерогенных вычислительных ресурсов и явным контролем уровня параллелизма.

лизма для каждого устройства. Для поиска оптимального физического плана предложен способ построения модели затрат на основе данных о поведении основных вычислительных паттернов реляционных и вспомогательных операторов. Потенциал прироста производительности за счет оптимизации запросов целиком для наилучшего с точки зрения производительности устройства оценивается с помощью эмпирических данных, полученных для коммерческой СУБД с открытым исходным кодом OmniSci DB. Предварительные результаты демонстрируют возможность ускорения обработки запросов в разы (3-8x) при выборе наиболее подходящего устройства исполнения.

СУБД; графический процессор; оптимизация аналитических запросов; обработка аналитических запросов; гетерогенные вычисления.

Р.А. Kurapov

HYBRID EXECUTION OF QUERIES TO ANALYTICAL DATABASES

Analytical database engines should benefit from evolving heterogeneous distributed architectures and utilize their resources efficiently: various accelerators, complex memory hierarchy, and distributed nature of systems bring performance improvement opportunities. The article reviews existing approaches for in-memory DBMS query executor implementation using hardware accelerators, especially GPUs. Massive parallelism and high on-device memory bandwidth make graphics processors a promising alternative as a core query evaluation device. Existing methods do not utilize all modern hardware capabilities and usually are bound, performance-wise, by relatively slow PCIe data transfer in a GPU-as-a-co-processor model for each kernel representing a single relational algebra operator. Another existing approaches' issue is explicit code base separation for relational algebra operators code generation (for CPU and GPU) that significantly limits possibilities of joint device usage for performance increase and make it less feasible. The article presents an efficient query execution method using an example of two device classes (CPU and GPU) by compiling queries into a single, device agnostic, intermediate representation (SPIR-V) and an approach for corresponding hybrid physical query plan optimization based on extended classical "Exchange" operator with explicit control over heterogeneous resources and parallelism level available. A cost model composition process using basic compute DBMS patterns benchmarking and buses bandwidth data for both relational and auxiliary operators is proposed. Potential processing speedup from holistic query optimization is estimated empirically with a commercial open source DBMS OmniSci DB. Preliminary results show significant (3-8x, depending on device choice) possible speedup even with just using the right device for the job.

DBMS; GPU; analytical query optimization; analytical query processing; in-memory databases.

Введение. В последние 10 лет, с развитием GPGPU (General-purpose computing on graphics processing units), СУБД начали активно использовать графические ускорители для обработки аналитических запросов. Многие исследования показали, что за счет массивного параллелизма графических ускорителей, векторных вычислений и высокой пропускной способностью памяти на чипе, выполнение операторов реляционной алгебры на ускорителе показывает значительный прирост в производительности, при условиях, что данные для операции помещаются в память устройства и возможности аппаратуры, такие как SIMD, разделяемая память, группы потоков, используются правильно. Время исполнения оператора реляционной алгебры может быть в разы меньше, чем у соответствующего алгоритма, реализованном на центральном процессоре.

Несмотря на все преимущества, использование графического ускорителя как основного устройства для исполнения запросов не лишено проблем [4]. Доступная память по-прежнему значительно уступает основной в размере. Это ограничение заставляет систему либо хранить данные в основной памяти и каждый раз загружать их в память устройства для обработки, либо организовывать распределенное хранение. И тот, и другой способы опираются на передачу данных через различ-

ные системы соединений, которые значительно уступают в скорости обращения в локальную для устройства память, и потому, становятся узким местом. Другая проблема – зависимость графического процессора от ЦПУ с точки зрения модели исполнения, то есть полагается на внешнюю аллокацию памяти и инициацию вычислений. Наконец, графические процессоры не приспособлены к выполнению кода со сложным ветвлением, интенсивному межпоточному взаимодействию и доступу к сложным структурам данных, что делает выгодным исполнение части операций на других устройствах.

В этой статье мы даем обзор подходов к использованию ускорителей для обработки аналитических запросов к реляционным базам данных, предлагаем архитектуру обработчика аналитических запросов для гетерогенных систем и приводим предварительные эмпирические данные для подтверждения потенциала повышения производительности на примере коммерческой базы данных OmniSci DB [17].

Обзор индустрии. С ростом вычислительной мощности графических процессоров и появлением более удобного способа их программирования (GPGPU), появились базы данных, использующие оптимизированные под архитектуру графических процессоров алгоритмы типичных операторов реляционной алгебры (например BlazingSQL, OmniSciDB, HeteroDB, Kinetica, SQream DB [1]). Для таких систем типичен механизм использования графического процессора для исполнения большинства DML (data manipulation language) запросов на устройстве. Обработка и оптимизация запроса, а также сериализация результатов, обычно, происходит на ЦПУ. Массивный параллелизм графических ускорителей во многих случаях дает значительный прирост производительности в типичных операторах баз данных [20]. Однако не все операции выгодно исполнять на графическом процессоре. Основным ограничением является пропускная способность шины передачи данных (в случае большинства баз данных использующих графический процессор – PCIe). За счет накладных расходов на копирование и использование кода на ЦПУ для запуска ядер для графического процессора (kernel), во многих случаях передавать данные становится неэффективно. Такую проблему можно было бы решить размещением базы данных в локальной памяти устройства, однако ее размер значительно уступает возможностям системной памяти (до 64Гб в самых дорогих решениях против >1Тб системной памяти), что не позволяет размещать реальные базы данных или делает их размещение слишком дорогим если используется множество графических процессоров.

Другое решение это конвейеризация – объединение нескольких операторов в один kernel с целью сокращения объема передачи данных между хостом и ускорителем. Генерация кода для оператора (цепочки операторов, оканчивающихся pipeline-breaker - оператором, требующим готовности всего результата до исполнения следующего), совместно с пакетной обработкой данных позволяет значительно сократить пересылку данных по медленной PCIe и, тем самым, добиться кратного роста скорости исполнения запроса [2]. Увеличение пропускной способности шины демонстрирует значительное повышение производительности исполнения запроса в целом [3], даже в случае входных данных, не помещающихся в память целиком (morsel-driven).

Использование графических процессоров для ускорения работы базы данных не ограничивается только исполнением отдельных операторов, например, исследовались подходы совместного использования с ЦПУ. Один из подходов [5] решает проблему высокой нагрузки на шину PCIe и недостаточности размера «быстрой» памяти графического процессора с помощью побитового разделения данных – горячие части данных (над которыми производятся вычислительно-емкие и независимые операции) хранятся в локальной памяти устройства, менее активно ис-

пользуемые – в системной памяти. Каждый запрос обрабатывается в 2 стадии: сначала неточная фильтрация записей, затем промежуточный результат передается на ЦПУ и объединяется с остаточными данными в основной памяти. Другой способ совместного использования ЦПУ и графического ускорителя – использование ускорителя для поиска оптимального плана исполнения запроса, а точнее, для оценки селективности с целью повышения качества физического плана [7].

Эффективная координация ЦПУ и графического процессора (и других ускорителей, таких как FPGA) имеет большой потенциал ([6–9]) повышения производительности систем, поэтому интерес к гибридному исполнению запросов растет. Одна из первых работ в этой области [10] показала 42% прироста производительности отдельно взятого простого запроса с помощью выбора подходящего, с точки зрения решаемой задачи, устройства.

С добавлением устройств и уровней иерархии памяти в систему, растет пространство перебора возможных вариантов исполнения запроса, и существующие техники оптимизации планов становятся менее эффективны, поэтому оптимизаторы гибридных запросов оказываются далеки от генерации оптимальных планов. Наиболее распространенным способом использования гетерогенных систем для обработки запросов является пооператорное планирование [11, 12] – для отдельно взятого оператора оценивается стоимость исполнения на ЦПУ и сравнивается со стоимостью передачи данных (туда и обратно) и время исполнения на ускорителе.

Например, Hawk [11] решает задачу гибридных запросов следующим образом. Реляционные операторы транслируются во внутреннее представление (pipeline programs), которое строится по принципам максимизации локальности по данным и коду. Последовательность реляционных операторов, не требующих полной материализации результатов (то есть, можно оставлять данные в регистрах), можно объединять и исполнять совместно [13]. Каждая «программа» выражается в виде специального оператора с платформу-зависимыми параметрами – модификации (варьируется способ доступа к памяти, стратегия исполнения, количество потоков), которые может производить оптимизатор для генерации планов специфичных какому-либо устройству. Планы оцениваются с помощью обучаемой модели затрат, и наилучший по стоимости выбирается в качестве оптимального. Проблема такого подхода заключается в том, что оптимизация локальна по отношению к оператору и не учитывает возможное ускорение цепочки операторов при отсутствии перемещения данных. К тому же, поскольку генератор кода Hawk выдает исходный OpenCL код, возникают дополнительные расходы на его компиляцию.

Избежать накладных расходов на генерацию исходного кода позволит введение «независимого» от устройства промежуточного представления. Помимо упрощения логики компиляции, промежуточное представление может абстрагировать архитектурные и микроархитектурные свойства аппаратуры. Так, [14] решает проблему адаптации обработчика для MonetDB с сохранением конкурентной производительности с помощью введения промежуточной алгебры, параметризованной уровнем параллелизма. Сама по себе алгебра поддерживает стандартные оптимизации, такие как смена размещения памяти (row/column store), различные стратегии материализации результатов, совместное использование общих результатов, векторизацию, слияние и разделение циклов. Как и в случае Hawk, генерацией кода и его исполнением занимается OpenCL. Операторы промежуточной алгебры разделяются по уровню параллелизма на «фрагменты» кода (термин «фрагмент» заимствован из [13]) – работа предлагает реализовать операторы реляционной алгебры через примитивы промежуточного представления компилятора и передавать результаты вычис-

лений напрямую по графу потребителю). Каждый фрагмент может быть скомпилирован в отдельный kernel и исполнен с максимально возможной эффективностью с точки зрения использования внутреннего параллелизма.

Использование ускорителя для исполнения оператора или запроса целиком оставляет ЦПУ незадействованным, что в случае серверной аппаратуры – значительная потеря вычислительной мощности. HetExchange [15] решает эту проблему введением дополнительных операторов физического плана, управляющих передачей данных с ЦПУ на графический процессор и обратно (device-crossing), разделением работы (оператор router – отвечает за уровень гетерогенного параллелизма), и контролем назначения работы нужному узлу (совместно с data-flow операторами, перемещающими гетерогенную память).

Материалы и методы. Параллельное исполнение операторов на разных устройствах требует соответствующих реализаций алгоритмов. С точки зрения разработки и поддержки системы наиболее эффективный путь – использование существующего стека технологий, такого как OpenCL для генерации кода под конкретные платформы. Например, CoGaDB генерирует исходный код на OpenCL C для запроса по статически заданным шаблонам. Такой подход использует преимущество существующих генераторов кода для устройств (ЦПУ, графический процессор), однако вносит дополнительные затраты на синтаксический и семантический анализ входного языка. В нашем исследовании мы используем более эффективный метод: сочетание использования гетерогенного плана запроса и прямой генерации промежуточного представления компилятора из абстрактного дерева операторов. Proteus (HAPe) [15] расширяет классический Exchange оператор, вводя дополнительные операторы передачи данных и контроля потока исполнения для утилизации гетерогенности и параллелизма вычислителей (рис. 1).

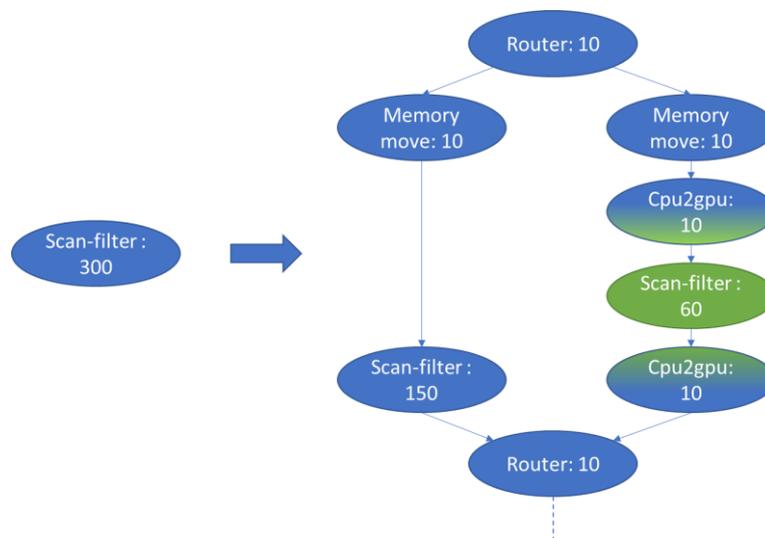


Рис. 1. Пример преобразования физического плана для ЦПУ в гетерогенный. Цветом отмечена смена устройства исполнения. Цифры – стоимость оператора в условных единицах времени исполнения

Мы исследуем возможность использования гетерогенного плана для создания промежуточного представления SPIR-V (Standard Portable Intermediate Representation) [18], которое не зависит от устройства и, как в случае с OpenCL, может выполняться на любом процессоре, для которого существует генератор ко-

да. Часть специфичной функциональности для устройства (к примеру, использование быстрой разделяемой памяти для графического ускорителя) компилируется заранее и соединяется на этапе генерации кода с запросом.

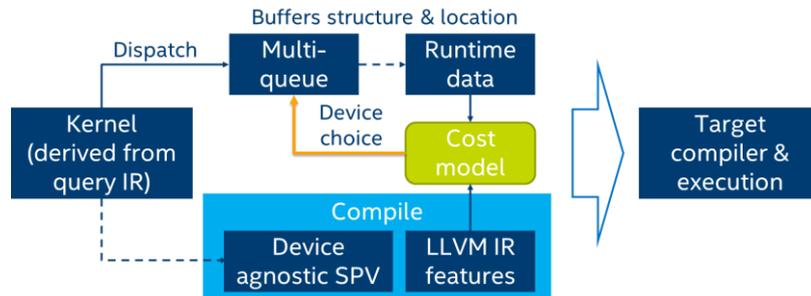


Рис. 2. Принципиальная архитектура исполнения сгенерированного кода на произвольной доступной платформе

В этой статье мы делаем первый шаг на пути к описанной схеме и предлагаем архитектуру генерации устройство-независимого представления (рис. 2), а также делаем предварительные замеры на коммерческой базе данных с открытым исходным кодом OmniSci DB с целью показать потенциальный выигрыш в производительности от использования гетерогенности. Для входного запроса строится абстрактное синтаксическое дерево (план), в узлах которого хранится информация о реляционных операторах. Исполнение запроса начинается с генерации промежуточного представления (LLVM IR) компилятора из плана запроса. К сгенерированному IR привязываются статически скомпилированные (до промежуточного представления) функции, реализующие стандартные алгоритмы (такие как hash-join), применяются классические оптимизации и, наконец, модуль транслируется в переносимое представление (SPIR-V). На этом этапе, с небольшими оговорками, полученный модуль может выполняться на любом доступном в системе устройстве.

Результаты. HAPE показывает значительный прирост производительности в случае использования гетерогенных планов исполнения запросов (до 3.5 раз). Здесь мы приводим эмпирическую оценку потенциального выигрыша в производительность от использования подходящего устройства для запроса, когда оптимизация происходит на уровне всего запроса, а не отдельных операторов.

OmniSci DB позволяет выполнять Just-In-Time компиляцию запроса для ЦПУ и графического процессора. На графике (рис. 3) представлены результаты замеров времени исполнения 4х запросов к данным NYC Taxi Benchmark [19], исполненных на Intel(R) Core(TM) i7-6950X CPU @ 3.00GHz и Nvidia GeForce GTX 1080 GPU. Исходный код SQL для запросов Q1-Q4 находится в приложении. Цифры отражают время исполнения сгенерированного кода в миллисекундах с накладными расходами на вызов процедур графического процессора и передачей данных, если это необходимо. Время компиляции не учитывается. Полученные значения – геометрическое среднее для 9 измерений, 2 из которых отбрасываются (максимум и минимум). Как показывает график (рис. 3), несмотря на то, что СУБД оптимизирована для исполнения запросов на графическом процессоре, некоторые запросы (Q2), выгоднее обрабатывать на ЦПУ, если его оптимизация производится на уровне всего запроса.

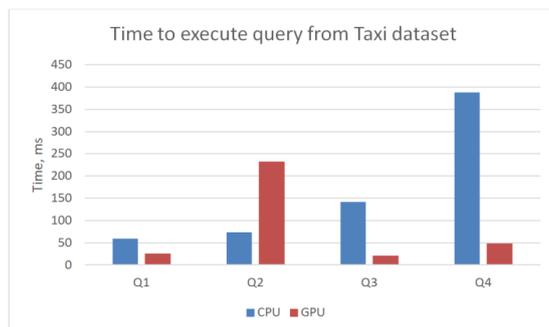


Рис. 3. График зависимости времени исполнения запросов к OmniSci DB в режимах ЦПУ и GPU для датасета NYC Taxi

Обсуждение. Предварительные результаты показывают возможность получения значительного (3–8 раз, в зависимости от устройств) прироста производительности, даже при использовании только подходящего для запроса устройства. Предложенная архитектура обработчика запросов позволяет абстрагировать имплементации операторов для различных устройств с помощью существующих стеков компиляции и генерации кода для различных платформ.

С увеличением количества операторов и контекстов их использования расширяется пространство перебора возможных физических планов исполнения запроса. Модель затрат для операторов физического плана позволит исследовать пространство поиска оптимального (достаточно эффективного) решения. ProGraML [16] показывает, как с помощью извлечения свойств промежуточного представления получить модель выбора наиболее эффективного устройства исполнения. В случае аналитических запросов задача упрощается, поскольку основные паттерны исполнения известны заранее – каждый оператор соответствует некоторому паттерну с точки зрения исполнения и профилю обращения в память. На основе данных о поведении таких паттернов можно построить модель затрат. На этапе оптимизации обработчику будет необходимо вычислить оценку стоимости вычисления оператора на конкретном устройстве, стоимость перемещения данных, которая зависит от скорости шины, и учесть накладные расходы на использование операторов, поддерживающих гетерогенное исполнение. Поскольку в нашем подходе генерируемый код практически не меняется в зависимости от конечного устройства, обучаемая модель, опирающаяся на характеристики промежуточного представления и на данные о загруженности устройства и их размере, будет устойчива к изменениям.

Заключение. В статье рассмотрена проблема оптимизации выполнения аналитических запросов для гетерогенных систем, рассмотрены существующие подходы к использованию аппаратных акселераторов и освещены проблемы этих методов. Для решения задачи совместной утилизации ресурсов предложена архитектура обработчика запросов, которая позволяет минимизировать затраты на поддержку и разработку системы за счет унификации кодогенерации, в отличие от существующих решений, без ущерба производительности, как, например, в случае генерации исходного кода OpenCL в качестве промежуточного представления. Предложенная архитектура в том числе позволяет проще расширять список поддерживаемых аппаратных ускорителей и дает возможность реализации единой модели затрат. В статье так же предложена схема построения такой модели на основе основных вычислительных паттернов. Приводятся эмпирические подтверждения потенциального повышения производительности коммерческой системы за счет использования гете-

рогенности. Следующим шагом нашего исследования станет построение модели оптимизации гетерогенного плана запроса на основе данных об относительной стоимости исполнения операторов на различных устройствах.

Приложение

Q1:

```
SELECT cab_type,
       count(*)
FROM trips
GROUP BY cab_type;
```

Q2:

```
SELECT passenger_count,
       avg(total_amount)
FROM trips
GROUP BY passenger_count;
```

Q3:

```
SELECT passenger_count,
       extract(year from pickup_datetime) AS pickup_year,
       count(*)
FROM trips
GROUP BY passenger_count,
       pickup_year;
```

Q4:

```
SELECT passenger_count,
       extract(year from pickup_datetime) AS pickup_year,
       cast(trip_distance as int) AS distance,
       count(*) AS the_count
FROM trips
GROUP BY passenger_count,
       pickup_year,
       distance
ORDER BY pickup_year,
       the_count desc;
```

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Hawon Chu, Seounghyun Kim, Joo-Young Lee, and Young-Kyoon Suh*. Empirical evaluation across multiple GPU-accelerated DBMSes, In *Proceedings of the 16th International Workshop on Data Management on New Hardware (DaMoN '20)*. Association for Computing Machinery, New York, NY, USA, 2020, Article 16, pp. 1–3. Doi: <https://doi.org/10.1145/3399666.3399907>.
2. *Henning Funke, Sebastian Breß, Stefan Noll, Volker Markl, and Jens Teubner*. Pipelined Query Processing in Coprocessor Environments. In *Proceedings of the 2018 International Conference on Management of Data (SIGMOD '18)*. Association for Computing Machinery, New York, NY, USA, 2018, 16031618. Doi: <https://doi.org/10.1145/3183713.3183734>.
3. *Clemens Lutz, Sebastian Breß, Steffen Zeuch, Tilmann Rabl, and Volker Markl*. 2020. Pump Up the Volume: Processing Large Data on GPUs with Fast Interconnects. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (SIGMOD '20)*. Association for Computing Machinery, New York, NY, USA, pp. 1633-1649. Doi:<https://doi.org/10.1145/3318464.3389705>.
4. *Breß S., Heimel M., Siegmund N., Bellatreche L., Saake G*. GPU-Accelerated Database Systems: Survey and Open Challenges. In: *Hameurlain A. et al. (eds) Transactions on Large-Scale Data- and Knowledge-Centered Systems XV. Lecture Notes in Computer Science, Vol. 8920*. Springer, Berlin, Heidelberg, 2014. Doi: https://doi.org/10.1007/978-3-662-45761-0_1.

5. Breß S., Heibel M., Siegmund N., Bellatreche L., Saake G. Exploring the Design Space of a GPU-Aware Database Architecture. In: Catania B. et al. (eds) *New Trends in Databases and Information Systems. Advances in Intelligent Systems and Computing*. Springer, Cham, 2014, Vol. 241. Doi: https://doi.org/10.1007/978-3-319-01863-8_25.
6. Bingsheng He, Mian Lu, Ke Yang, Rui Fang, Naga K. Govindaraju, Qiong Luo, and Pedro V. Sander. Relational query coprocessing on graphics processors. *ACM Trans. Database Syst.* 34, 4, Article 21 (December 2009), 39 p. Doi: <https://doi.org/10.1145/1620585.1620588>.
7. Heibel M., Markl V. A First Step Towards GPU-assisted Query Optimization. In: Bordawekar, R. and Lang, C.A. (eds.) *ADMS@VLDB*, 2012, pp. 33-44.
8. Peter Bakkum and Kevin Skadron. Accelerating SQL database operations on a GPU with CUDA. In *Proceedings of the 3rd Workshop on General-Purpose Computation on Graphics Processing Units (GPGPU-3)*. Association for Computing Machinery, New York, NY, USA, 2010. pp. 94-103. Doi: <https://doi.org/10.1145/1735688.1735706>.
9. Johns Paul, Jiong He, and Bingsheng He. GPL: A GPU-based Pipelined Query Processing Engine. In *Proceedings of the 2016 International Conference on Management of Data (SIGMOD '16)*. Association for Computing Machinery, New York, NY, USA, 2016, pp. 1935-1950. Doi: <https://doi.org/10.1145/2882903.2915224>.
10. Ilic, Aleksandar & Pratas, Frederico & Trancoso, Pedro & Sousa, Leonel. High-Performance Computing on Heterogeneous Systems: Database Queries on ЦПВ and GPU, 2011. Doi 10.3233/978-1-60750-803-8-202.
11. Sebastian Breß, Bastian Köcher, Henning Funke, Steffen Zeuch, Tilmann Rabl, and Volker Markl. Generating custom code for efficient query execution on heterogeneous processors, *The VLDB Journal* 27, 6 (December 2018), pp. 797-822. Doi: <https://doi.org/10.1007/s00778-018-0512-y>.
12. Owaida M., Sidler D., Kara K. and Alonso G. Centaur: A Framework for Hybrid ЦПВ-FPGA Databases, *2017 IEEE 25th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), 2017*, pp. 211-218. Doi: 10.1109/FCCM.2017.37.
13. Thomas Neumann. Efficiently compiling efficient query plans for modern hardware. *Proc. VLDB Endow.* 4, 9 (June 2011), pp. 539-550. DOI:<https://doi.org/10.14778/2002938.2002940>.
14. Holger Pirk, Oscar Moll, Matei Zaharia, and Sam Madden. Voodoo - a vector algebra for portable database performance on modern hardware. *Proc. VLDB Endow.* 9, 14 (October 2016), 1707-1718. Doi: <https://doi.org/10.14778/3007328.3007336>.
15. Periklis Chrysogelos, Manos Karpathiotakis, Raja Appuswamy, and Anastasia Ailamaki. HetExchange: encapsulating heterogeneous ЦПВ-GPU parallelism in JIT compiled engines. *Proc. VLDB Endow.* 12, 5 (January 2019), pp. 544-556. Doi: <https://doi.org/10.14778/3303753.3303760>.
16. Cummins C., Fisches Z.V., Ben-Nun T., Hoefler T., and Leather H. Programl: Graph-based deep learning for program optimization and analysis, *arXiv preprint arXiv:2003.10536*, 2020.
17. Christopher Root and Todd Mostak. MapD: a GPU-powered big data analytics and visualization platform. In *ACM SIGGRAPH 2016 Talks (SIGGRAPH '16)*. Association for Computing Machinery, New York, NY, USA, 2016, Article 73, pp. 1-2. Doi: <https://doi.org/10.1145/2897839.2927468>.
18. <https://www.khronos.org/registry/spir-v/>.
19. New York (N.Y.). Taxi and Limousine Commission. New York City Taxi Trip Data, 2009-2018. Inter-university Consortium for Political and Social Research [distributor], 2019-02-20. <https://doi.org/10.3886/ICPSR37254.v1>.
20. Shanbhag, Anil and Madden, Samuel and Yu, Xiangyao. A Study of the Fundamental Performance Characteristics of GPUs and CPUs for Database Analytics, *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, Association for Computing Machinery, New York, NY, USA, 2020*, pp. 1617-1632.

Статью рекомендовал к опубликованию д.т.н., профессор В.М. Курейчик.

Курапов Петр Александрович – МФТИ; e-mail: petr.kurapov@gmail.com; Москва, Россия; кафедра микропроцессорных технологий в интеллектуальных системах управления; аспирант.

Kurapov Petr Alexandrovich – Moscow Institute of Physics and Technology; e-mail: petr.kurapov@gmail.com; Moscow, Russia; the department of microprocessor technologies in intelligent control systems; graduate student.

П.А. Пантелюк

ДЕТЕКТИРОВАНИЕ ВЫБРОСОВ В МЕТОДЕ ПРЯМОЙ СТЕРЕО-ВИЗУАЛЬНОЙ ОДОМЕТРИИ НА БАЗЕ ИЕРАРХИЧЕСКОЙ КЛАСТЕРИЗАЦИИ

Представляется подход к стерео-визуальной одометрии без явного вычисления оптического потока. Визуальная одометрия – метод получения навигационной информации путем обработки последовательности кадров с бортовых камер. Существует два подхода к обработке видеоинформации – используя хорошо локализуемые участки изображения – признаковые точки и используя все высококонтрастные пиксели – прямой метод. Прямой метод работает, используя интенсивности всех высококонтрастных пикселей изображения, что позволяет снизить вычислительную сложность, затраченную на поиск, описание, сопоставление признаковых точек и повысить точность оценки движения. Однако методы подобного класса обладают недостатком – наличие движущихся объектов в кадре существенно снижают точность оценки параметров движения. Для избегания этого применяются методы детектирования выбросов. Классические методы детектирования выбросов во входных данных, такие как RANSAC плохо применимы, и имеют высокие вычислительные затраты из-за вычислительно сложной функции рейтингования гипотез. Целью данной работы является описание и демонстрация подхода детектирования выбросов на базе алгоритма иерархической кластеризации, который выделяет статистически наиболее вероятное решение, минуя этап рейтингования каждой гипотезы, что значительно снижает вычислительную сложность. Для иерархической кластеризации предлагается мера расстояния между гипотезами с низкой чувствительностью к ошибкам оценки параметров движения. Также предлагается расширение алгоритма стерео-визуальной одометрии для работы в более сложных условиях видимости благодаря переходу от интенсивностного представления изображения к многоканальному бинарному. Перевод изображения к многоканальному бинарному представлению дает инвариантность к изменениям яркости изображения, однако, требует модификации алгоритмов нелинейной оптимизации для работы с бинарными дескрипторами. В результате работы показано, что предложенный алгоритм детектирования выбросов способен работать в реальном масштабе времени на мобильных устройствах, и может служить менее ресурсоёмкой заменой алгоритма RANSAC в задачах визуальной одометрии и вычисления оптического потока. Качественные метрики предложенного решения демонстрируются на датасете KITTI. Приведены зависимости качества работы алгоритма от параметров алгоритма.

Навигация; визуальная одометрия; оптический поток; детектирование выбросов.

Р.А. Panteliuk

OUTLIERS DETECTION IN THE STEREO-VISUAL ODOMETRY BASED ON HIERARCHICAL CLUSTERIZATION

This paper presents an approach to stereo-visual odometry without explicitly calculating optical flow. Visual odometry is a method of obtaining navigation information by processing a sequence of frames from onboard cameras. There are two approaches to processing video information - using well-localized areas of the image - feature points and using all high-contrast pixels - a direct method. The direct method works by using the intensities of all high-contrast pixels in the image, making it possible to reduce the computational complexity spent searching, describing, and matching feature points and increasing the accuracy of motion estimation. However, methods of this class have a drawback - the presence of moving objects in the frame significantly reduces the accuracy of the estimation of motion parameters. Outlier detection techniques are used to avoid this problem. Classical methods for detecting outliers in input data, such as RANSAC, are poorly applicable and have high computational costs due to the computationally complex function of rating hypotheses. This work aims to describe and demonstrate an outlier detection approach

based on the hierarchical clustering algorithm, which selects the statistically most probable solution, bypassing the stage of rating each hypothesis, which significantly reduces the computational complexity. For hierarchical clustering, a measure of the distance between hypotheses with low sensitivity to errors in estimating motion parameters is proposed. An extension of the stereo-visual odometry algorithm is also proposed to work in more complex visibility conditions due to the transition from the intensity representation of the image to a multichannel binary one. Transforming an image to a multichannel binary representation gives invariance to changes in image brightness. However, it requires modification of nonlinear optimization algorithms to work with binary descriptors. As a result of the work, it has been shown that the proposed outlier detection algorithm can operate in real-time on mobile devices and can serve as a less resource-intensive replacement for the RANSAC algorithm in problems of visual odometry and optical flow eviction. Qualitative metrics of the proposed solution are demonstrated on the KITTI dataset. The dependences of the performance of the algorithm on the parameters of the algorithm are given.

Navigation; visual odometry; optical flow; outliers detection.

Введение. Оценка собственного движения объекта управления на основе последовательности кадров с бортовых камер называется визуальная одометрия [1–4]. Это одна из основных тем исследований в области робототехники и компьютерного зрения. Визуальная одометрия широко применяется в большом количестве задач, в числе которых детектирование и сопровождение объектов и навигация.

В частности, визуальная одометрия на базе стереокамеры активно исследуется поскольку позволяет точно и быстро оценить свое собственное движение в метрическом масштабе, что имеет большое значение в автономном вождении [5].

Существующие и активно развивающиеся методы визуальной одометрии можно разделить на две группы: базирующиеся на признаковых точках и прямые. Методы, базирующиеся на признаковых точках, используют в расчетах детектированные и сопровождаемые на последовательности кадров признаковые точки. Функция невязки для такого метода определяется как ошибки проецирования признаковых точек, а оценка параметров движения осуществляется путем минимизации функции невязки [6]. Однако, при таком подходе локальное изменение внешности признаковых точек на последовательности кадров приводит к дрейфу позиций признаковых точек от начальных позиций, что приводит к накоплению ошибки оценки собственного движения [7]. Напротив, прямые методы оценивают параметры движения путем минимизации разностей интенсивностей пикселей между двумя последовательными кадрами [8]. Использование большего количества пикселей в оценке параметров движения. Недостатком прямого метода является сложность детектирования и обработки выбросов, а также чувствительность к изменению яркости кадров [9]. По этой причине большинство новых методов визуальной одометрии использует поход на основе признаковых точек и детектированию выбросов алгоритмом RANSAC [10–14], последовательно выбирая минимальных набор признаковых точек, вычислением оптимальных параметров движения на их основе, и рейтингованию вычисленных параметров движения по оставшимся признаковым точкам.

В данной работе предлагается подход, который последовательно выбирает минимальный набор признаковых точек на изображении, на подобии того, как это происходит в подходах на базе признаковых точек, и вычисляет параметры движения используя интенсивности пикселей в окрестностях этих точек как в прямых методах. Рейтингование гипотез осуществляется не на основе фотометрической разности оставшиеся окрестностей признаковых точек, а на основе принадлежности гипотез некому множеству, большинству гипотез со схожими признаками.

Предложенный метод. Функцию невязки прямого метода визуальной одометрии можно записать в следующем виде:

$$error(R, \mathbf{t}) = \sum_i^N \sum_j^W \|T(K \times \mathbf{p}_i + win(j)) - I(K \times (R \times \mathbf{p}_i + \mathbf{t}) + win(j))\|^2, \quad (1)$$

где R – матрица поворота камеры, \mathbf{t} – вектор трансляции камеры, K – матрица внутренних параметров камеры, описывающая камеру Обскура, \mathbf{p}_i – позиция обзереваемой точки в пространстве, $win(j)$ – функция, задающая окрестность проекции точки \mathbf{p}_i с аппаратурой W , $I(\mathbf{x}\mathbf{y})$ – функция задающая текущий кадр изображения, $T(\mathbf{x}\mathbf{y})$ – функция задающая предыдущий кадр изображения.

Уравнение (1) описывает процесс движения камеры в пространстве. В момент времени t_0 камера снимает изображение T , в момент времени t_1 камера снимает изображение I . За время $t_1 - t_0$ камера поворачивается (это поворот описывается матрицей поворота R) и смещается на вектор \mathbf{t} . При этом камера наблюдает статическую точку окружающего пространства \mathbf{p}_i . Координаты точки \mathbf{p}_i вычисляются посредством стерео-триангуляции алгоритмом стерео-зрения [15]. В идеальном случае, если отсутствуют какие-либо шумы и искажения камеры, точке \mathbf{p}_i будет соответствовать пиксели на изображениях I и T которые будут иметь одинаковую яркость. Поскольку камера обзеревает большое количество точек пространства (по сути, все что снимает камера) и эти точки имеют разное свечение (наблюдаемая сцена обладает контрастностью и неоднородностью), и работа камеры вносит шумы и искажения в наблюдаемые яркости, можно выразить R, \mathbf{t} как параметры движения, которые минимизируют функцию (1).

Выразив матрицу поворота как функцию от углов Родрига [16]:

$$R = Rodrig(\alpha, \beta, \gamma), \quad (2)$$

задача оценки движения сводится к задаче минимизации нелинейной функции:

$$\alpha, \beta, \gamma, \mathbf{t} = \arg \min[error(Rodrig(\alpha, \beta, \gamma), \mathbf{t})]. \quad (3)$$

Сравнение результата работы вычисления оптического потока классическим методом [17] и базирующемся на предложенном методе приведены на рис. 1.

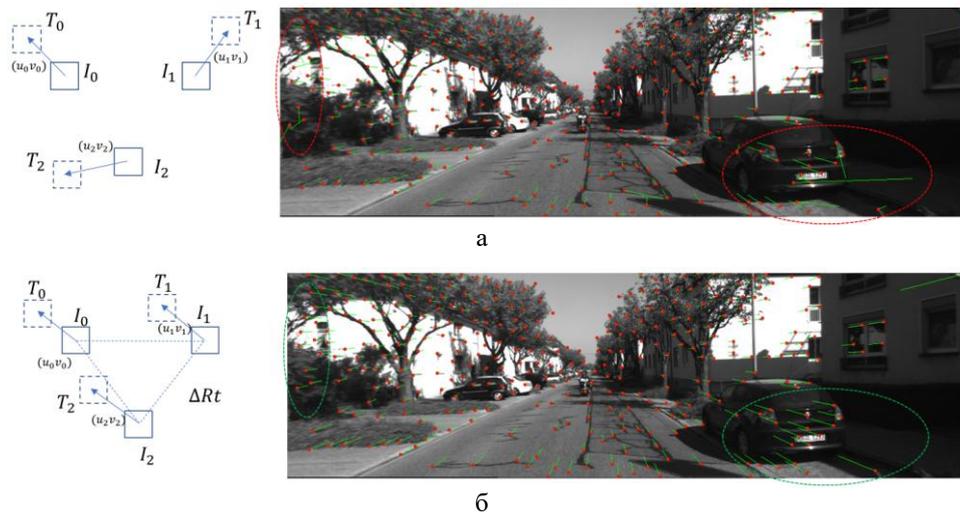


Рис. 1. Вычисление оптического потока классическим методом (а), вычисление оптического потока на базе предложенного метода (б)

Как показано на рис. 1, вычисление оптического потока путем минимизации фотометрической ошибки и определение оптимальных параметров движения – $\Delta R \mathbf{t}$ дает меньше выбросов по сравнению с методом вычисления оптического потока, который минимизирует фотометрическую ошибку и определяет оптимальные смещения – (u_i, v_i) индивидуально для каждой признаковой точки на изображении.

Цифровые видеокамеры искажают измеряемое свечение нелинейной передаточной функцией сенсора камеры, поскольку не способны охватить весь динамический диапазон, это выражается в том, что, если камера обзрывает окружение с неизменяемой «яркостью», полученные последовательные кадры могут значительно отличаться по суммарной яркости. Это явление нарушает предположение о постоянстве яркости обзриваемых точек, на основе которого задана функция (1). В случаях, когда два последовательных кадра будут сильно отличаться по яркости из-за резкого изменения времени экспонирования сенсора камеры, оценка параметров $\alpha, \beta, \gamma, \mathbf{t}$ будет содержать значительную ошибку.

Для нивелирования влияния описанных выше эффектов предлагается использовать непараметрическое преобразования изображения известное как Цензус-преобразование (*англ. Census transform*) [18], которое обладает инвариантностью к локальным изменениям яркости изображения, и зарекомендовало себя в алгоритмах компьютерного зрения. Цензус-преобразование строит бинарный дескриптор для каждого пикселя изображения путем сравнения яркости пикселя с яркостями его локального соседства (восемь соседних пикселей). В результате каждого сравнения формируется один бит дескриптора. Цензус-преобразованные изображения сравниваются, используя расстояние Хемминга, в отличие от оригинальных изображений, которые сравниваются, используя евклидово расстояние.

Расстояние по Хэммингу не применимо в нелинейной оптимизации в исходном виде, однако расстояние Хемминга можно линеаризовать. Представив каждый бит Цензус-дескриптора как отдельный канал в изображении, Уравнение (1) можно записать в следующем виде:

$$error(R, \mathbf{t}) = \sum_c^8 \sum_i^N \sum_j^W \left\| T(K \times \mathbf{p}_i + win(j))_c - I(K \times (R \times \mathbf{p}_i + \mathbf{t}) + win(j))_c \right\|^2, \quad (4)$$

где c – номер канала изображения.

Функция (4) обладает инвариантностью к локальным изменениям яркости изображения, а также более устойчива в условиях плохой видимости камерами.

Поскольку обзриваемые точки \mathbf{p}_i могут принадлежать динамическим объектам, или они не обзриваются в оба момента времени t_0 и t_1 , их использование в выборке для минимизации функции (4) приводит к значительному увеличению ошибки оценки параметров движения.

Для уменьшения влияния выбросов (точек \mathbf{p}_i принадлежащих динамическим объектам или не обзриваемых в оба момента времени t_0 и t_1) выполняется следующая процедура:

- 1) из множества признаков точек выбирается подмножество случайных $M = 3$ точек;
- 2) выполняется минимизация функции (4);
- 3) оцененные параметры движения $\alpha, \beta, \gamma, \mathbf{t}$ заносятся в список.

Выполнив описанную процедуру несколько раз (в реальных экспериментах число повторений равно 1000), формируется набор гипотез параметров движения, часть из которых будет построена с использованием верных точек \mathbf{p}_i , а часть на выбросах.

Для того что бы разделить условно верные гипотезы от условно ложных, выполняется иерархическая кластеризация гипотез [19]. Гипотезы, которые были получены и использованием «истинных» точек p_i будут образовывать кластер «истинных» гипотез, которые отражают параметры движения камеры. Используя предположение что «истинных» точек p_i больше, чем ложных (иначе вычислить параметры движения практически невозможно), «истинных» гипотез будет тоже больше, и кластер с наибольшим количеством гипотез является кластером «истинных» гипотез.

Выделив кластер «истинных» гипотез, гипотезы входящие в этот кластер усредняются, что дает оценку параметров движения камеры. Точки p_i , которые использовались с построением «истинных» гипотез, взвешиваются согласно частоте их использования, и используются в дальнейшем уточнении параметров движения камеры.

Принцип работы метода детектирования ложных гипотез приведен на рис. 2.

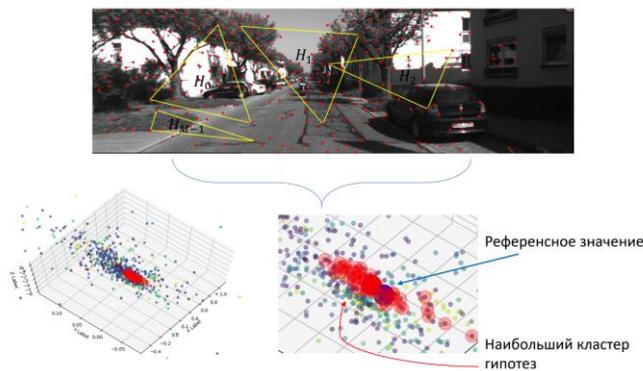


Рис. 2. Принцип работы метода детектирования ложных гипотез

Гипотеза H_i – набор из трех точек используемых для оценки параметров движения. Каждая точка на нижнем графике – параметры движения, построенные на основе данных гипотез H_i . Гипотезы, принадлежащие наибольшему кластеру выделены красным цветом

Испытание работы алгоритма. Испытание точности работы алгоритма осуществлялось датасете KITTI [20]. Датасет предоставляет истинную траекторию движения автомобиля и изображения с бортовой стереокамеры. Изображения используются для оценки параметров движения предложенным алгоритмом, а истинная траектория используется для оценки точности работы метода.

Точность оценки параметров движения зависит от количества используемых гипотез. График зависимости точности оценки позиции от количества гипотез приведен на рис. 3.

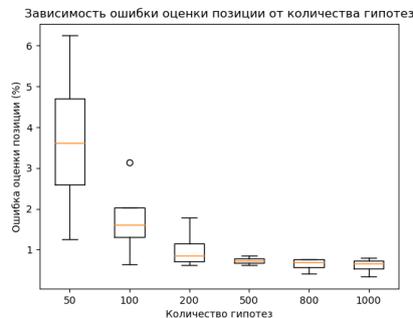


Рис. 3. Зависимость точности оценки позиции от количества гипотез

Из графика видно, что для данной конфигурации достаточно 500 гипотез
Траектория движения, полученная предложенным методом для датасета
KITTI с номером 00 приведена на рис. 4.

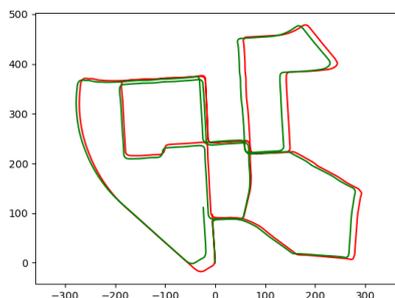


Рис. 4. траектория движения для датасета KITTI с номером 00. Красная линия – референсная траектория, зеленая линия – полученная траектория.
Ошибка в конечной точке – 0.6 % от пройденного пути

Заключение. В ходе проведенной работы был предложен метод оценки параметров движения камеры, который способен работать в условиях ухудшенной видимости, а также меньше подвержен влиянию выбросов во входных данных что ведет к меньшей ошибке оценке параметров движения. Время генерации 1000 гипотез и выделение «истинных» гипотез, с использованием функции невязки (1) составляет 50 мс. на процессоре core i7 2.6 ГГц, с использованием функции невязки (4) – 270 мс. в однопоточном режиме, что позволяет оценивать параметры движения в реальном масштабе времени. На датасетах KITTI предложенный метод достиг в среднем 2% ошибки от пройденного пути. В дальнейших работах по развитию предложенного метода предлагается улучшить способ выбора подмножества входных точек, выбирать такие подмножества точек, которые дадут «истинную» гипотезу с большей вероятностью. Это позволит формировать кластера гипотез с более четкими границами, что приведет к более точному выбору «истинных» гипотез и повысит точность оценки параметров движения.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Howard A. Real-time stereo visual odometry for autonomous ground vehicles, 2008 *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2008 зз. 3946-3952.
2. Scaramuzza D., Fraundorfer F. Visual odometry [tutorial], *IEEE robotics & automation magazine*, 2011, Vol. 18, No. 4, pp. 80-92.
3. Fraundorfer F., Scaramuzza D. Visual odometry: Part ii: Matching, robustness, optimization, and applications, *IEEE Robotics & Automation Magazine*, 2012, Vol. 19, No. 2, pp. 78-90.
4. Maimone M., Cheng Y., Matthies L. Two years of visual odometry on the mars exploration rovers, *Journal of Field Robotics*, 2007, Vol. 24, No. 3, pp. 169-186.
5. Aqel M.O.A. et al. Review of visual odometry: types, approaches, challenges, and applications, *SpringerPlus*, 2016, Vol. 5, No. 1, pp. 1-26.
6. Badino H., Yamamoto A., Kanade T. Visual odometry by multi-frame feature integration, *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 222-229.
7. Liu H. et al. Navigational drift analysis for visual odometry, *Computing and Informatics*, 2014, Vol. 33, No. 3, pp. 685-706.
8. Usenko V. et al. Direct visual-inertial odometry with stereo cameras, 2016 *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1885-1892.
9. Usenko V. et al. Direct visual-inertial odometry with stereo cameras, 2016 *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1885-1892.

10. *Derpanis K.G.* Overview of the RANSAC Algorithm, *Image Rochester NY*, 2010, Vol. 4, No. 1, pp. 2-3.
11. *Zuliani M.* RANSAC for Dummies, *Vision Research Lab, University of California, Santa Barbara*, 2009.
12. *Nistér D.* Preemptive RANSAC for live structure and motion estimation, *Machine Vision and Applications*, 2005, Vol. 16, No. 5, pp. 321-329.
13. *Raguram R., Frahm J. M., Pollefeys M.* A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus, *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, 2008, pp. 500-513.
14. *Li H. et al.* An efficient image matching algorithm based on adaptive threshold and RANSAC, *IEEE Access*, 2018, Vol. 6, pp. 66963-66971.
15. *Haller I., Nedeveschi S.* GPU optimization of the SGM stereo algorithm, *Proceedings of the 2010 IEEE 6th International Conference on Intelligent Computer Communication and Processing*. IEEE, 2010, pp. 197-202.
16. *Piña E.* Rotations with Rodrigues' vector, *European journal of physics*, 2011, Vol. 32, No. 5, pp. 1171.
17. *Suhr J.K.* Kanade-lucas-tomasi (klt) feature tracker, *Computer Vision (EEE6503)*, 2009, pp. 9-18.
18. *Hafner D., Demetz O., Weickert J.* Why is the census transform good for robust optic flow computation?, *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, Berlin, Heidelberg, 2013, pp. 210-221.
19. *Johnson S.C.* Hierarchical clustering schemes, *Psychometrika*, 1967, Vol. 32, No. 3, pp. 241-254.
20. *Geiger A. et al.* Vision meets robotics: The kitti dataset, *The International Journal of Robotics Research*, 2013, Vol. 32, No. 11, pp. 1231-1237.

Статью рекомендовал к опубликованию д.т.н., профессор В.В. Курейчик.

Пантелюк Павел Александрович – Южный федеральный университет; e-mail: pavel-panteluk@mail.ru; г. Таганрог, Россия; тел.: +79185693138; кафедра синергетики и процессов управления; аспирант.

Pantelyuk Pavel Alexandrovich – Southern Federal University; e-mail: pavel-panteluk@mail.ru; Taganrog, Russia; phone: +79185693138; the department of synergetics and management processes; postgraduate student.

Раздел III. Моделирование процессов и систем

УДК 621.396.969

DOI 10.18522/2311-3103-2021-2-120-129

В.А. Деркачев

МОДЕЛЬ РАССЕЯНИЯ РАДИОЛОКАЦИОННЫХ СИГНАЛОВ ОТ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

Рассматривается модель рассеяния радиолокационных сигналов от беспилотных летательных аппаратов (БПЛА) мультироторного типа для формирования обучающих данных нейросетевого классификатора. В последнее время к изучению вопроса обнаружения и классификации малоразмерных беспилотных летательных аппаратов (БПЛА) наблюдается повышенный интерес, который связан с развитием номенклатуры БПЛА в продаже и производстве. Помимо развития БПЛА увеличение производительности вычислителей позволило создавать классификаторы с использованием новых нейросетевых алгоритмов. Данная модель осуществляет формирование радиолокационных изображений, получаемых в результате отражения от беспилотного летательного аппарата радиолокационного сигнала с линейной частотной модуляцией с учетом конфигурации, характеристик, текущего местоположения и параметров полета наблюдаемого объекта. При расчете отраженного сигнала учитываются углы поворота БПЛА (тангажа, крена и рыскания), скорости полета, размера и местоположения винтов в текущей конфигурации БПЛА. Полученная модель может быть полезна для формирования обучающего набора классификатора беспилотных летательных аппаратов мультироторного типа, построенного с использованием сверточных нейронных сетей. Необходимость использования модели, формирующей данные для нейронной сети, обуславливается требованием к большому числу обучающих и верифицирующих выборок, а также большим разнообразием конфигураций беспилотных летательных аппаратов, что сильно увеличивает сложность и стоимость создания обучающего датасета с применением экспериментальных измерений. Помимо собственно обучения нейронной сети, данную модель можно применить для оценки возможности обнаружения и классификации различных видов мультироторных БПЛА, в разработке специализированной радиолокационной станции обнаружения данного вида объектов.

Математическая модель; беспилотные летательные аппараты; БПЛА; FMCW; дальностно-скоростной портрет; ДСП.

V.A. Derkachev

MODEL OF SCATTERING OF RADAR SIGNALS FROM UAV

In this article, a model of scattering of radar signals from unmanned aerial vehicles (UAVs) of a multi-rotor type is considered for the formation of training data for a neural network classifier. Recently, there has been an increased interest in studying the issue of detecting and classifying small unmanned aerial vehicles (UAVs), which is associated with the development of the UAV range in sales and production. In addition to the development of UAVs, an increase in the performance of computers made it possible to create classifiers using new neural network algorithms. This model generates radar images obtained as a result of the reflection of a chirp radar signal from an unmanned aerial vehicle, taking into account the configuration, characteristics, current location and flight parameters of the observed object. When calculating the reflected signal, the angles of rotation of the UAV (pitch, roll and yaw), flight speed, size and location of propellers in the current UAV configuration are taken into account. The resulting model can be useful for the formation of a training set of a classifier of unmanned aerial vehicles of a multi-rotor type, built

using convolutional neural networks. The need to use a model that generates data for a neural network is due to the requirement for a large number of training and verification samples, as well as a wide variety of configurations of unmanned aerial vehicles, which greatly increases the complexity and cost of creating a training dataset using experimental measurements. In addition to training the neural network itself, this model can be used to assess the detection and classification of various types of multi-rotor UAVs, in the development of a specialized radar station for detecting this type of objects.

Mathematical model; unmanned aerial vehicles; UAV; FMCW; range-speed portrait; RSP.

Введение. Одной из проблем классификации БПЛА с использованием радиолокации является выбор метода предварительной подготовки радиолокационных данных и методики построения классификатора.

Применение различных методов подготовки данных может давать различные результаты классификации, что связано с различным временным и частотным разрешением получаемых радиолокационных изображений. Одним из наиболее популярных методов анализа является применение выделения микродоплеровской сигнатуры (MDS), формируемой вследствие вращения лопастей БПЛА мультироторного типа с высокой скоростью [1–9]. Данный метод состоит из двух этапов: преобразования принятых сигналов с использованием частотного-временного анализа и формирования представления сигнала, подходящего для извлечения данных. В большинстве источников для обработки полученных отраженных зондирующих сигналов применяется метод оконного преобразования Фурье (STFT) [4, 10–13]. Однако метод STFT не дает высокого разрешения радиолокационного изображения, поэтому в некоторых работах предложено использовать методы высокого частотно-временного разрешения, такие как MUSIC [14], сглаженного преобразования Вигнера-Вилля (SPWVD) [15] и итеративного адаптивного подхода (IAA) [16, 17]. Основной проблемой применения данных методов связано с высокой вычислительной сложностью, что, соответственно, приводит к малой применимости в устройствах классификации БПЛА в реальном масштабе времени. Другой методикой анализа данных может быть использование разложения по эмпирическим модам (EMD) [2, 18], однако применение данного метода может быть затруднено из-за возможного смешения мод и проблемы определения границ сигнала. Исходя из данных, полученных в проанализированных источниках, был осуществлен выбор типа обработки с использованием оконного преобразования Фурье и последующим формированием дальностно-скоростного портрета целевого объекта.

Основной задачей данной модели является формирование набора радиолокационных изображений для дальнейшего обучения классификатора, построенного на основе нейронной сети. Одной из проблем в использовании нейронных сетей является создание репрезентативных обучающих данных. Для создания данных можно использовать натурные измерения, однако для обучения требуются тысячи изображений одного БПЛА с разных ракурсов, что повышает трудозатратность и стоимость создания классификатора. Помимо этого, для классификации обычно используют несколько классов объектов, соответственно количество измерений увеличивается кратно. Одним из вариантов решения данной проблемы может быть создание имитационной модели рассеяния радиолокационного сигнала от БПЛА, что позволит создать большое количество радиолокационных изображений пригодных для обучения НС за сравнительно небольшой срок. Созданная модель достаточно гибкая и позволяет смоделировать различные ситуации, связанные с различным взаимным расположением объектов на радиосцене, а также имеет небольшое количество ограничений в конфигурациях моделируемых БПЛА.

Целью данной работы является разработка модели рассеяния радиоволн радиолокационной станции непрерывного излучения с частотной модуляцией от верхностей БПЛА.

Описание модели. Данная модель была построена с учетом нахождения приемника и передатчика РЛС в одной точке, для БПЛА допустимо изменение параметров местоположения без ограничений углов взаимного расположения БПЛА и РЛС, векторов скоростей БПЛА и скоростей вращения лопастей. Геометрия модели представлена на рис. 1. Для каждого отсчета времени производится вычисление координат (x_c, y_c, z_c) для центра БПЛА и (x_n, y_n, z_n) для двигателей, где n – номер двигателя. Угол φ_n – угол между плоскостями роторов двигателей и направлением на РЛС, где n – номер двигателя. Расстояние до центра БПЛА – r_c , а расстояния до двигателей БПЛА подписаны как r_m , где n – номер двигателя.

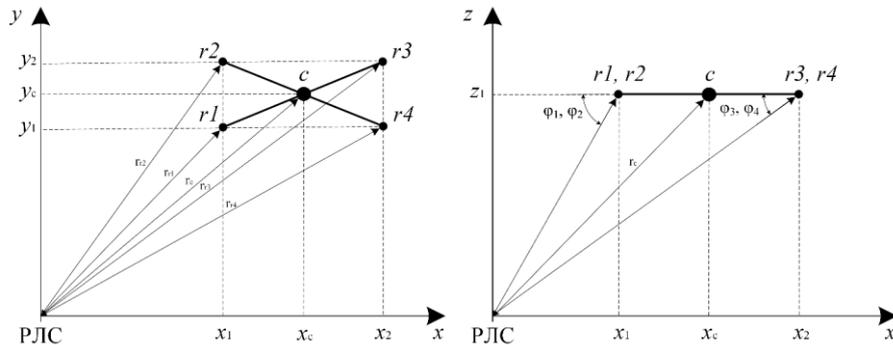


Рис. 1. Геометрия модели БПЛА

Далее на рис. 2 показана геометрия отдельного двигателя вместе с лопастями. Координаты двигателя (x_r, y_r, z_r) , а координаты концов лопастей (x_n, y_n, z_n) , где n – номер лопасти. Угол Θ_n – между нормальными лопастей двигателя \vec{n} и направлением на радиолокационную станцию, где n – номер двигателя. Расстояние до центра БПЛА – r_r , а расстояния до концов лопастей БПЛА обозначены как r_{bn} .

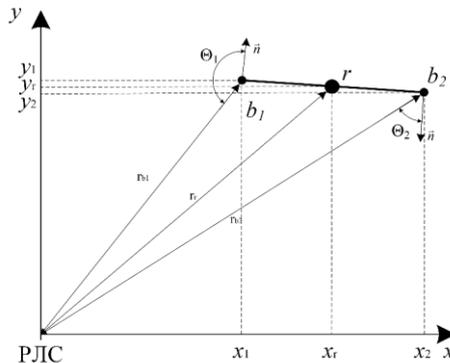


Рис. 2. Геометрия модели для отдельного двигателя

Алгоритм работы модели показан на рис. 3. Первоначально задаются параметры радиолокационной станции и параметров БПЛА. Далее осуществляется формирование первичного радиолокационного сигнала. Затем осуществляется вычисление откликов, данный этап осуществляется путем последовательного выполнения следующих операций:

- ♦ задание начальных координат винта и лопастей в системе координат P''' ;

- ◆ перенос полученных координат в систему координат P'' (с центром в центре каждого винта);
- ◆ перенос координат в систему координат P' (с учетом тангажа, крена и рыскания);
- ◆ перенос координат в конечную систему координат P (с центром в точке с координатами РЛС);
- ◆ нахождение соответствующих каждой лопасти, винту и центральной части БПЛА дальностей до РЛС;
- ◆ нахождение углов φ между плоскостями роторов двигателей и направлением на РЛС;
- ◆ нахождение углов Θ между нормальными лопастей и направлением на РЛС;
- ◆ вычисление соответствующих откликов от каждой лопасти и их сложение в общий отклик от БПЛА совместно с откликом от центральной части БПЛА.

После вычисления отраженного сигнала производится децимация отсчетов, с целью экономии вычислительных ресурсов и приведения параметров выходных данных модели к параметрам реальной радиолокационной станции. Данный этап требуется ввиду особенностей обработки данных в РЛС, связанных с ограниченной частотой дискретизации АЦП, которая сильно меньше используемой полосы разностного сигнала РЛС непрерывного излучения с частотной модуляцией (FMCW), используемого в моделировании.

Полученные отсчеты сигнала позволяют осуществить формирование дальностно-скоростного портрета (ДСП) путем последовательного применения быстрого преобразования Фурье (БПФ) сначала по «быстрому» времени (внутри импульса сигнала), а затем применением БПФ по «медленному» времени (межпериодная обработка сигнала).



Рис. 3. Алгоритм работы модели

Формирование откликов на зондирующий сигнал. Наибольший интерес вызывает вопрос формирования отраженного от БПЛА радиолокационного сигнала. Вычисление откликов сигнала от БПЛА можно свести к следующему выражению:

$$S_2 = S_{body} + S_{blades}, \quad (1)$$

где s_2 – суммарный отклик от БПЛА, s_{body} – отраженный сигнал от центральной части БПЛА, s_{blades} – отраженный сигнал от лопастей.

Рассмотрим формирование отраженного сигнала от центральной части БПЛА:

$$s_{body} = s_{1_b} \cdot \exp\left(-\frac{j \cdot 2 \cdot \pi \cdot f_0 \cdot r_{body} \cdot t}{c}\right), \quad (2)$$

где s_{1_b} – сигнал от РЛС, задержанный по времени в соответствии с временной задержкой распространения радиоволн, f_0 – несущая частота, r_{body} – расстояние до центра БПЛА, t – вектор времени, c – скорость света.

Рассмотрим формирование отраженного сигнала от лопастей БПЛА:

$$s_{blades} = \sum_{r=1}^R s_r, \quad (3)$$

$$s_r = \sum_{n=0}^{N-1} (L-l) \cdot A_n \cdot B_n \cdot C_n, \quad (4)$$

где R – число роторов, r – номер ротора, N – число лопастей на роторе, n – номер лопасти, A_n – фазовая составляющая, B_n – амплитудная составляющая, C_n – задержанный сигнал РЛС с учетом фазового набега обусловленного радиальной скоростью движения БПЛА.

Далее показана методика вычисления каждой из составляющих, формирующих отраженный сигнал от лопастей. Фазовая составляющая сигнала вычисляется по следующей формуле [16]:

$$A_n = \exp\left\{j \cdot \left[-\frac{4 \cdot \pi \cdot f_0}{c} \cdot r_{blade} + \frac{2 \cdot \pi \cdot f_0}{c} \cdot (L+l) \cdot \sin(\theta_n) \cdot \cos(\varphi)\right]\right\}, \quad (5)$$

где f_0 – несущая частота РЛС сигнала, r_{blade} – дальность до лопасти в текущий момент времени, L – длина лопасти, l – радиус ротора, φ – текущий угол между плоскостью вращения лопасти и направляющим вектором на РЛС, Θ_n – текущий модифицированный угол между нормалью лопасти и направляющим вектором на РЛС. Угол Θ_n вычисляется следующим образом [16]:

$$\Theta_n = \Theta_0 + \frac{(n-1) \cdot 2 \cdot \pi}{N}, \quad (6)$$

где Θ_0 – текущий между нормалью лопасти и направляющим вектором на РЛС, который ранее был вычислен (Θ_k).

Амплитудная составляющая сигнала вычисляется следующим образом [17]:

$$B_n = \alpha \frac{\sin\left(\frac{2 \cdot \pi \cdot f_0}{c} \cdot (L-l) \cdot \cos(\varphi) \cdot \sin(\theta_n)\right)}{\frac{2 \cdot \pi \cdot f_0}{c} \cdot (L-l) \cdot \cos(\varphi) \cdot \sin(\theta_n)}, \quad (7)$$

где α – интенсивность отраженного сигнала, численно равная текущему ЭПР лопасти.

Составляющая, описывающая сигнал от РЛС с учетом временной задержки и фазового набега представлена ниже:

$$C_n = s_{s_blade} \cdot \exp\left(-\frac{j \cdot 2 \cdot \pi \cdot f_0 \cdot r_{blade} \cdot t}{c}\right), \quad (8)$$

где S_{s_blade} – сигнал от РЛС, задержанный по времени в соответствии с временной задержкой распространения радиоволн, r_{blade} – дальность до лопасти в текущий момент времени.

Далее осуществляется выделение разностного сигнала (излучаемого сигнала РЛС и принятого отраженного сигнала от БПЛА). Затем производится децимация до частоты 10 МГц, которая позволяет имитировать сигнал s_3 на выходе АЦП с соответствующей частотой дискретизации.

После накопления необходимого количества отсчетов в матричном виде осуществляется формирование дальностно-временного портрета (ДВП) путем применения БПФ по периодам зондирующего сигнала:

$$s_4(k, :) = fft(s_3(k, :)), k = (1, 2, \dots, N_t), \quad (9)$$

где k – номер периода сигнала (дальностного профиля), s_3 – накопленный разностный сигнал, N_t – число периодов зондирующего сигнала. Далее производится вычисление дальностно-скоростного портрета (ДСП) путем внутривыборочной БПФ:

$$s_5(:, n) = fft(s_4(:, n)), n = (1, 2, \dots, N_{ADC}), \quad (10)$$

где n – номер периода сигнала (дальностного профиля), s_4 – ДВП, N_{ADC} – число отсчетов АЦП.

Моделирование. Модель в рамках данной работы настроена под конкретную связку БПЛА (*Xiaomi Mi Drone Mini*) и РЛС (*TI IWR1642*), данный набор был выбран для дальнейшей верификации модели путем сравнения данных, полученных моделью и экспериментальными данными. Ввиду того что натурный эксперимент еще не осуществлен, далее показана валидация модели путем осуществления моделирования одиночного винта большего размера и обладающей малой скоростью вращения, что позволяет получить отклики на дальностно-скоростном портрете в пределах однозначного измерения радиальной скорости вращения лопастей.

В табл. 1 показаны основные параметры модели, используемые для дальнейшего моделирования. Прототипом РЛС, используемой для моделирования, является *TI IWR1642* совместно с отладочным модулем *TI DCA1000*.

Таблица 1

Наименование параметра	Величина
Несущая частота	77 ГГц
Скорость изменения частоты	29,982 МГц/мкс
Ширина спектра сигнала (полная)	1,7989 ГГц
Ширина спектра сигнала (используемая)	767,54 МГц
Частота дискретизации	1,6 ГГц
Число периодов ЛЧМ импульсов	128
Длительность ЛЧМ импульса	60 мкс
Период ЛЧМ импульсов	160 мкс
Частота дискретизации АЦП	10 МГц
Число отсчетов АЦП за период ЛЧМ импульса	256
Максимальная дальность обнаружения	50 м
Разрешение по дальности	0,195 м

Наименование параметра	Величина
Максимальная скорость объекта	6 м/с
Разрешение по скорости	0,09 м/с
Размеры лопасти (отладочные)	2 м
Частота вращения лопасти (отладочная)	0,0796 Гц
Полное модельное время	20,3 мс

Параметры модели подобраны таким образом чтобы показать работоспособность модели в рамках формирования откликов сигнала.

Для имитационного моделирования БПЛА использовались параметры реального БПЛА *Xiaomi Mi Drone Mini*, параметры модели показаны в табл. 2.

Таблица 2

Наименование параметра	Величина
Число лопастей на роторе	2
Число роторов	4
Размеры лопасти	0,038 м
Радиус оси ротора	0,0025 м
Расстояние от геометрического центра до винта	0,0575 м
Частота вращения лопасти	± 100 Гц

На рис. 4 показан результат моделирования с учетом только одного винта (длина лопасти 2 м) с очень малой частотой вращения (0,0796 Гц), для моделирования использованы параметры БПЛА из табл. 1. Использование данных параметров моделирования позволяет показать правильность моделируемых дальностно-скоростных портретов и необходимо для верификации модели. Радиальные скорости краев лопастей при начальном угле поворота 20^0 составили 1 м/с и минус 0,8 м/с, что можно увидеть на ДСП и его сечении по уровню минус 15 дБ. Помимо скоростных характеристик на сечении ДСП видно разнесение концов лопастей на большой дальности (9,38 м и 10,9 м для каждой из лопастей, при нахождении центра винта на расстоянии 10 м), что связано с большой длиной лопасти.

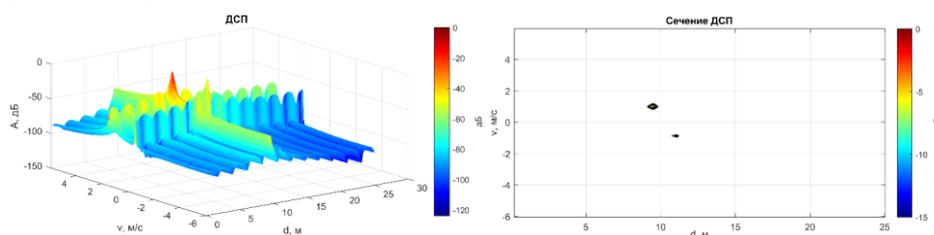


Рис. 4. ДСП одиночного винта большого размера с малой скоростью вращения

На рис. 5 показаны результаты моделирования с учетом одного винта малого размера (длина лопасти 0,038 м) и высокой частотой вращения (100 Гц). Параметры, использованные для моделирования приведены в табл. 2. Большое число составляющих по скорости связано с эффектом алиасинга из-за выхода радиальной скорости концов лопастей за пределы однозначно определяемых РЛС скоростей. Из-за малых размеров лопастей отклики сконцентрированы на одной дальности.

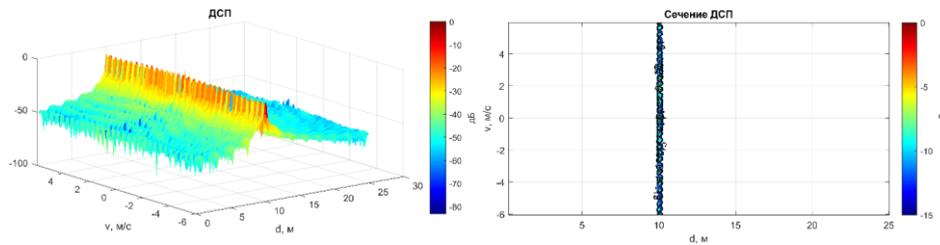


Рис. 5. ДСП одиночного винта малого размера с высокой скоростью вращения

На рис. 6 показаны результаты моделирования с учетом четырех винтов малого размера (длина лопасти 0,038 м) и высокой частотой вращения (100 Гц, -100 Гц, 100 Гц). Параметры, использованные для моделирования приведены в табл. 2.

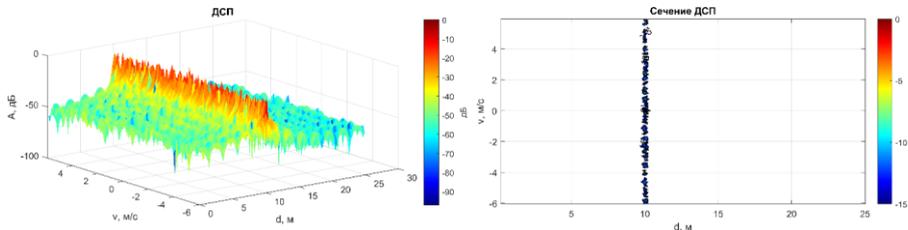


Рис. 6. ДСП 4 винтов малого размера с высокой скоростью вращения

Выводы. Предложенная модель рассеяния радиолокационных сигналов от беспилотных летательных аппаратов позволяет генерировать радиолокационные изображения с произвольными параметрами зондирующего радиолокационного сигнала, характеристиками беспилотного летательного аппарата мультироторного типа и текущими параметрами полета данного аппарата. Применение данной модели с целью синтеза большого числа радиолокационных изображений может позволить сформировать обучающую выборку для последующего обучения нейросетевого классификатора РЛИ. Для обучающих выборок несомненным плюсом будет возможность формирования радиолокационных изображений БПЛА с параметрами и характеристиками, изменяемыми в широких пределах, что несомненно может ускорить и уменьшить стоимость разработки классификатора с большим числом классифицируемых классов объектов. Дальнейшая работа должна быть направлена на ускорение работы модели и перенос вычислений на графические процессоры, для получения возможности формирования крупных обучающих выборок за короткие сроки.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Harmanny R.I.A., de Wit J.J.M., Premel-Cabic G.* Radar micro-Doppler mini-UAV classification using spectrograms and cepstrograms // International Journal of Microwave and Wireless Technologies. – 2015. – Vol. 7, No. 3-4. – P. 469.
2. *Oh B.S. et al.* An EMD-based micro-Doppler signature analysis for mini-UAV blade flash reconstruction // 2017 22nd International Conference on Digital Signal Processing (DSP). – IEEE, 2017. – P. 1-5.
3. *Tahmoush D.* Detection of small UAV helicopters using micro-Doppler // Radar Sensor Technology XVIII. – International Society for Optics and Photonics, 2014. – Vol. 9077. – P. 907717.

4. *Molchanov P. et al.* Classification of small UAVs and birds by micro-Doppler signatures // *International Journal of Microwave and Wireless Technologies*. – 2014. – Vol. 6, No. 3-4. – P. 435-444.
5. *De Wit J.J.M., Harmanny R.I.A., Molchanov P.* Radar micro-Doppler feature extraction using the singular value decomposition // *2014 International Radar Conference*. – IEEE, 2014. – P. 1-6.
6. *Fuhrmann L. et al.* Micro-Doppler analysis and classification of UAVs at Ka band // *2017 18th International Radar Symposium (IRS)*. – IEEE, 2017. – P. 1-9.
7. *Zhang P. et al.* Classification of drones based on micro-Doppler signatures with dual-band radar sensors // *2017 Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL)*. – IEEE, 2017. – P. 638-643.
8. *Kim B.K., Kang H.S., Park S.O.* Drone classification using convolutional neural networks with merged Doppler images // *IEEE Geoscience and Remote Sensing Letters*. – 2016. – Vol. 14, No. 1. – P. 38-42.
9. *Martinez J. et al.* Convolutional neural network assisted detection and localization of UAVs with a narrowband multi-site radar // *2018 IEEE MTT-S International Conference on Microwave for Intelligent Mobility (ICMIM)*. – IEEE, 2018. – P. 1-4.
10. *Stankovic L., Daković M., Thayaparan T.* Time-frequency signal analysis with applications. – Artech house, 2014.
11. *Stankovic L., Djurovic I., Thayaparan T.* Separation of target rigid body and micro-Doppler effects in ISAR imaging // *IEEE Transactions on Aerospace and Electronic Systems*. – 2006. – Vol. 42, No. 4. – P. 1496-1506.
12. *Stankovic L., Thayaparan T., Dakovic M.* Signal decomposition by using the S-method with application to the analysis of HF radar signals in sea-clutter // *IEEE Transactions on Signal Processing*. – 2006. – Vol. 54, No. 11. – P. 4332-4342.
13. *Fioranelli F. et al.* Classification of loaded/unloaded micro-drones using multistatic radar // *Electronics Letters*. – 2015. – Vol. 51, No. 22. – P. 1813-1815.
14. *Schmidt R.* Multiple emitter location and signal parameter estimation // *IEEE transactions on antennas and propagation*. – 1986. – Vol. 34, No. 3. – P. 276-280.
15. *Tan R. et al.* Improved micro-Doppler features extraction using smoothed-pseudo Wigner-Ville distribution // *2016 IEEE Region 10 Conference (TENCON)*. – IEEE, 2016. – P. 730-733.
16. *Yardibi T. et al.* Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares // *IEEE Transactions on Aerospace and Electronic Systems*. – 2010. – Vol. 46, No. 1. – P. 425-443.
17. *Sun H. et al.* Improving the Doppler resolution of ground-based surveillance radar for drone detection // *IEEE Transactions on Aerospace and Electronic Systems*. – 2019. – Vol. 55, No. 6. – P. 3667-3673.
18. *Bai X. et al.* Imaging of micromotion targets with rotating parts based on empirical-mode decomposition // *IEEE transactions on geoscience and remote sensing*. – 2008. – Vol. 46, No. 11. – P. 3514-3523.
19. *Ziyue T., Yongliang W., Zhiwen W.* STAP scheme to detection of hovering helicopter // *WCC 2000-ICSP 2000. 2000 5th International Conference on Signal Processing Proceedings. 16th World Computer Congress 2000*. – IEEE, 2000. – Vol. 3. – P. 1921-1924.
20. *Слюсарь Н.М.* Вторичная модуляция радиолокационных сигналов динамическими объектами. – Смоленск: ВА ВПВО СВ РФ, 2006.

REFERENCES

1. *Harmanny R.I.A., de Wit J.J.M., Premel-Cabic G.* Radar micro-Doppler mini-UAV classification using spectrograms and cepstrograms, *International Journal of Microwave and Wireless Technologies*, 2015, Vol. 7, No. 3-4, pp. 469.
2. *Oh B.S. et al.* An EMD-based micro-Doppler signature analysis for mini-UAV blade flash reconstruction, *2017 22nd International Conference on Digital Signal Processing (DSP)*. IEEE, 2017, pp. 1-5.
3. *Tahmoush D.* Detection of small UAV helicopters using micro-Doppler, *Radar Sensor Technology XVIII*. International Society for Optics and Photonics, 2014, Vol. 9077, pp. 907717.
4. *Molchanov P. et al.* Classification of small UAVs and birds by micro-Doppler signatures, *International Journal of Microwave and Wireless Technologies*, 2014, Vol. 6, No. 3-4, pp. 435-444.

5. De Wit J.J.M., Harmanny R.I.A., Molchanov P. Radar micro-Doppler feature extraction using the singular value decomposition, *2014 International Radar Conference*. IEEE, 2014, pp. 1-6.
6. Fuhrmann L. et al. Micro-Doppler analysis and classification of UAVs at Ka band, *2017 18th International Radar Symposium (IRS)*. IEEE, 2017, pp. 1-9.
7. Zhang P. et al. Classification of drones based on micro-Doppler signatures with dual-band radar sensors, *2017 Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL)*. IEEE, 2017, pp. 638-643.
8. Kim B.K., Kang H.S., Park S.O. Drone classification using convolutional neural networks with merged Doppler images, *IEEE Geoscience and Remote Sensing Letters*, 2016, Vol. 14, No. 1, pp. 38-42.
9. Martinez J. et al. Convolutional neural network assisted detection and localization of UAVs with a narrowband multi-site radar, *2018 IEEE MTT-S International Conference on Micro-Stakes for Intelligent Mobility (ICMIM)*. IEEE, 2018, pp. 1-4.
10. Stankovic L., Daković M., Thayaparan T. Time-frequency signal analysis with applications. Artech house, 2014.
11. Stankovic L., Djurovic I., Thayaparan T. Separation of target rigid body and micro-Doppler effects in ISAR imaging, *IEEE Transactions on Aerospace and Electronic Systems*, 2006, Vol. 42, No. 4, pp. 1496-1506.
12. Stankovic L., Thayaparan T., Dakovic M. Signal decomposition by using the S-method with application to the analysis of HF radar signals in sea-clutter, *IEEE Transactions on Signal Processing*, 2006, Vol. 54, No. 11, pp. 4332-4342.
13. Fioranelli F. et al. Classification of loaded/unloaded micro-drones using multistatic radar, *Electronics Letters*, 2015, Vol. 51, No. 22, pp. 1813-1815.
14. Schmidt R. Multiple emitter location and signal parameter estimation, *IEEE transactions on antennas and propagation*, 1986, Vol. 34, No. 3, pp. 276-280.
15. Tan R. et al. Improved micro-Doppler features extraction using smoothed-pseudo Wigner-Ville distribution, *2016 IEEE Region 10 Conference (TENCON)*. IEEE, 2016, pp. 730-733.
16. Yardibi T. et al. Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares, *IEEE Transactions on Aerospace and Electronic Systems*, 2010, Vol. 46, No. 1, pp. 425-443.
17. Sun H. et al. Improving the Doppler resolution of ground-based surveillance radar for drone detection, *IEEE Transactions on Aerospace and Electronic Systems*, 2019, Vol. 55, No. 6, pp. 3667-3673.
18. Bai X. et al. Imaging of micromotion targets with rotating parts based on empirical-mode decomposition, *IEEE transactions on geoscience and remote sensing*, 2008, Vol. 46, No. 11, pp. 3514-3523.
19. Ziyue T., Yongliang W., Zhiwen W. STAP scheme to detection of hovering helicopter, *WCC 2000-ICSP 2000. 2000 5th International Conference on Signal Processing Proceedings. 16th World Computer Congress 2000*. IEEE, 2000, Vol. 3, pp. 1921-1924.
20. Slyusar' N.M. Vtorichnaya modulyatsiya radiolokatsionnykh signalov dinamicheskimi ob"ektami [Secondary modulation of radar signals by dynamic objects]. Smolensk: VA VPVO SV RF, 2006.

Статью рекомендовал к опубликованию д.т.н. А.В. Ковалев.

Деркачев Владимир Александрович – Южный федеральный университет; e-mail: vderkachev@sfedu.ru; г. Таганрог, Россия; тел.: +79614154733; кафедра радиотехнических и телекоммуникационных систем, ассистент.

Derkachev Vladimir Aleksandrovich – Southern Federal University; e-mail: vderkachev@sfedu.ru; Taganrog, Russia; phone +79614154733; the department of radio engineering & telecommunication systems; associate professor.

В.В. Янчич

**ИССЛЕДОВАНИЕ ВЛИЯНИЯ КРЕПЛЕНИЯ ЭЛЕКТРОМЕХАНИЧЕСКОГО
ПРЕОБРАЗОВАТЕЛЯ НА ХАРАКТЕРИСТИКИ
ПЬЕЗОЭЛЕКТРИЧЕСКОГО ДАТЧИКА УСКОРЕНИЯ**

Исследования выполнены с целью получения данных, необходимых для повышения точности расчета и оптимизации конструкций пьезоэлектрических датчиков механических величин, которые широко используются для контроля, мониторинга и диагностики сложного оборудования и инженерных сооружений в различных областях техники. Задача исследований состоит в изучении особенностей рабочих деформаций в пьезоэлектрическом электромеханическом преобразователе в области крепления к основанию датчика и оценке их влияния на основные метрологические характеристики. Объектом исследования является электромеханический преобразователь в виде цилиндрического монолитного блока из пьезоэлектрической керамики с отношением высоты к диаметру от 0,33 до 2, закрепленный на основании датчика, на которое воздействует поступательное ускорение вибрационных колебаний со стороны контролируемого объекта. С использованием программного пакета ANSYS Multiphysics исследована математическая модель преобразователя с двумя принципиально разными типами крепления – «свободно-скользящего» и «жесткого». При этом выявлен механизм поперечного механического шунтирования деформации преобразователя в области границы жесткого крепления. Для количественной оценки влияния условий крепления на характеристики преобразователя предложены «коэффициент влияния крепления» и формула для его определения при различных отношениях высоты к диаметру преобразователя. Для определения влияния крепления преобразователя в реальных конструкциях датчиков разработана методика и проведены экспериментальные исследования с учетом свойств используемых на практике конструкционных материалов и наиболее часто применяемого упругого поджатия элементов. Установлено, что свойства материала основания и соотношения размеров преобразователя в реальной конструкции датчика могут вызывать изменения его коэффициента преобразования по напряжению до 15 %, коэффициента преобразования по заряду до 22 %, электрической емкости до 9 % и частоты продольного резонанса до 16 %. Влияние граничных условий крепления снижается одновременно с возрастанием относительной высоты преобразователя. Экспериментально получены данные для расчета коэффициента влияния крепления при выполнении основания датчиков из металлов с модулем упругости от 74 до 300 ГПа и плотностью от 2,7 до 17,5 т/м³. Результаты проведенных исследований могут учитываться при конструировании пьезоэлектрических датчиков механических величин.

Пьезоэлектрический датчик; электромеханический преобразователь, математическая модель; крепление; ускорение; вибрация; деформация.

V.V. Yanchich

**RESEARCH OF ELECTROMECHANICAL TRANSDUCER'S FASTENING
INFLUENCE TO THE ACCELERATION PIEZOSENSOR'S
CHARACTERISTICS**

The research is carried out to obtain data necessary to improve calculation accuracy and for construction optimization in piezosensors of mechanical values. These sensors are widely used for control, monitoring and diagnostics of complex equipment and engineering structures. The task of the research is to study the features of working deformations in electromechanical piezotransducer in area of fastening to sensor base and to estimate their influence to the main metrological characteristics. The object of the research is electromechanical transducer in the form of cylindrical monolithic block made of piezoelectric ceramics with height-to-diameter ratio 0.33 to 2. This transducer is fixed on the sensor's base which is affected by progressive acceleration of vibration oscillations from the controlled object. Using the ANSYS Multiphysics software

package a mathematical model of the transducer with two fundamentally different types of fastening – “free sliding” and “rigid” – has been investigated. At the same time the mechanism of transverse mechanical shunting of transducer’s deformation in limit region of the rigid fastening was revealed. “Fastening effect coefficient” and its determination formula for various transducer’s height-to-diameter ratios are proposed for quantitative estimate of fastening conditions influence to transducer’s characteristics. Taking into account the properties of structural materials used in practice and the most frequently used elastic compression of elements a method was developed and experimental studies were carried out to determine the effect of transducer’s fastening in real sensor designs. It was found that base material properties and transducer dimensions ratio in real sensor design can cause changes in voltage conversion coefficient up to 15%, the charge conversion coefficient up to 22%, electric capacity up to 9% and longitudinal resonance frequency up to 16%. The influence of boundary fastening conditions decreases simultaneously with the increase of transducer’s relative height. The calculating data were obtained experimentally for the fastening effect coefficient when making the sensor’s base of metals with elastic modulus of 74–300 GPa and density of 2700–17700 kg/m³. The results of research carried out can be taken into account when designing piezosensors of mechanical values.

Piezosensor; electromechanical transducer; mathematical model; fastening; acceleration; vibration; deformation.

Введение. Пьезоэлектрические датчики механических величин (вибрационного ускорения, динамического давления, силы) на протяжении последних десятилетий, благодаря своим преимуществам, широко используются для контроля, мониторинга и диагностики сложного оборудования и инженерных сооружений в различных областях техники [1–3]. Несмотря на постоянное совершенствование таких датчиков, их метрологические и эксплуатационные характеристики не всегда удовлетворяют предъявляемым к ним требованиям [4, 5]. В этой связи, актуальным остается решение задач, направленных на дальнейшее развитие методов расчета и проектирование пьезоэлектрических датчиков.

Целью работы является получение дополнительных данных, необходимых для повышения точности расчета и оптимизации конструкций пьезоэлектрических датчиков механических величин, которые могут учитываться при их проектировании [6].

Задача исследований состоит в изучении особенностей рабочих деформаций в пьезоэлектрическом электромеханическом преобразователе в области крепления к основанию датчика и оценке их влияние на основные метрологические характеристики датчика.

Данные, получены по результатам исследований пьезоэлектрического электромеханического преобразователя, который может использоваться в качестве базовой конструкции для пьезоэлектрических датчиков различного типа, включая многопараметровые и многофункциональные [7, 8].

Исследование математической модели. В общем случае, в зависимости от основного измеряемого параметра и конструктивных особенностей, на электромеханический преобразователь датчика может действовать сила, ускорение или одновременно оба этих воздействия, вызывающие различный характер его деформации [8–10]. Так как работа пьезоэлектрического преобразователя во многом обусловлена характером распределения возникающих в нем механических напряжений, изучению этого вопроса уделено особое внимание.

Математическое моделирование проводилось с использованием программного пакета ANSYS Multiphysics для конечно-элементных математических моделей, который позволяет учитывать пьезоэлектрические свойства базового материала преобразователей и обеспечивает широкие возможности их численного расчета [6, 11]. Частота продольного установочного резонанса определялась методом модального анализа ANSYS.

Объектом исследования являлся преобразователь в виде монолитного блока диаметром D и высотой H с электродами на торцах, выполненный из пьезоэлектрической керамики ЦТС-19 и наполяризованный в осевом направлении [12]. Один из торцов блока закреплен на базе – эквиваленте основания датчика. Считалось, что модуль упругости базы и ее масса в 10 раз больше чем у блока.

Исследовались случаи движения базы вдоль продольной оси блока при ускорении $1 \text{ м}\cdot\text{с}^{-2}$. Рассматривалась квазистатическая осесимметричная задача для двух типов крепления блока к сопрягаемой поверхности базы. При первом типе крепления («свободно-скользящее» крепление) исключалось перемещение точек блока относительно основания по продольной оси, но допускалось их свободное скольжение по сопрягаемой поверхности базы. При втором типе крепления («жесткое» крепление) на сопрягаемой поверхности исключалось относительное смещение точек базы и блока.

На рис. 1, в качестве примера для этих случаев крепления, показаны характер деформации блока $\varnothing 10 \times 10 \text{ мм}$ (недеформированное состояние обозначено пунктиром, масштаб деформации для наглядности увеличен) и возникающие при этом механические напряжения на частоте много меньшей собственной частоты системы.

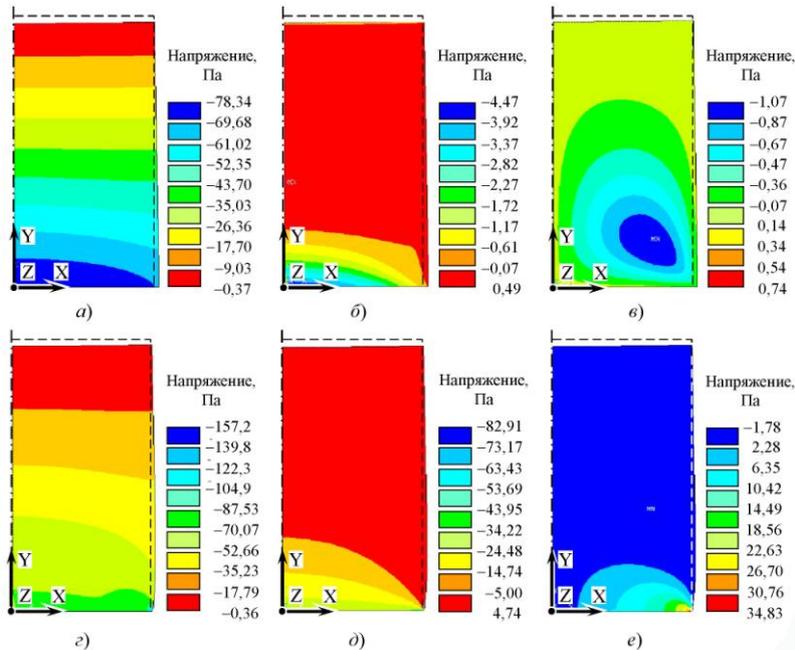


Рис. 1. Характер деформации блока и механические напряжения под воздействием продольного ускорения при свободно-скользящем (а–в) и жестком (г–е) способах крепления блока: а, г – продольные напряжения σ_z ; б, д – поперечные напряжения σ_x ; в, е – сдвиговые напряжения σ_{xy}

Видно, что граничные условия крепления блока оказывают влияние на характер его деформации и на возникающие в нем механические напряжения, особенно вблизи контакта с базой. Как показали результаты расчета, влияющие отличия деформаций блока находятся в слое высотой $h \approx 0,333D$ (D диаметр блока) (рис. 2), что соответствует принятому при расчете значению коэффициента Пуассона [13] пьезоэлектрической керамики.

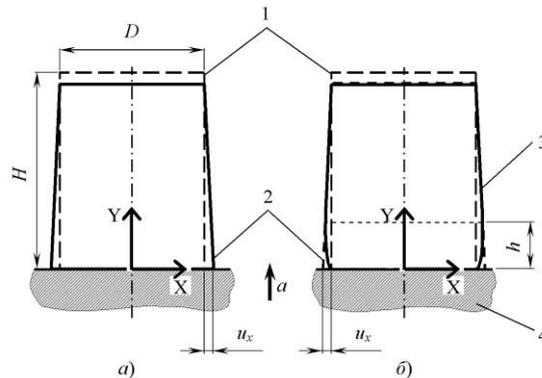


Рис. 2. Деформация монолитного блока при свободно-скользящем (а) и жестком (б) креплении на базе: 1 – исходная конфигурация блока; 2 – конфигурация блока при воздействии ускорения a в случае свободно-скользящего крепления; 3 – конфигурация блока при воздействии ускорения a в случае жесткого крепления; 4 – база; u_x – смещение точек блока

Поперечное механическое шунтирование деформации блока при его жестком креплении приводит к снижению коэффициента преобразования по заряду, электрической емкости и частоты продольного установочного резонанса, которая определяет верхнюю границу рабочего диапазона частот пьезоэлектрического датчика [14–17]. Наблюдаемое возрастание коэффициента преобразования по напряжению связано со снижением электрической емкости вследствие эффекта «зажатия кристалла» [18].

Для количественной оценки влияния на характеристики преобразователя поперечного механического шунтирования предложен коэффициент влияния крепления (КВК) $k_k = A_{жк} / A_{ск}$ (здесь $A_{жк}$ – характеристика преобразователя при жестком креплении, $A_{ск}$ – соответствующая характеристика при свободно-скользящем креплении).

В табл. 1 приведены рассчитанные значения КВК, а на рис. 3 показана зависимость характеристик от отношения высоты H к диаметру блока D .

Таблица 1

Рассчитанные значения КВК при различных отношениях высоты H к диаметру D блока

Отношение H/D	КВК для характеристик преобразователя			
	коэффициента преобразования по напряжению	коэффициента преобразования по заряду	электрической емкости	частоты продольного установочного резонанса
0,333	0,823	0,757	0,930	1,118
0,50	0,882	0,838	0,953	1,079
0,750	0,921	0,892	0,969	1,052
1,0	0,941	0,919	0,977	1,039
1,25	0,953	0,935	0,981	1,031
1,50	0,961	0,946	0,984	1,026
1,75	0,966	0,954	0,985	1,022
2,0	0,971	0,959	0,988	1,020

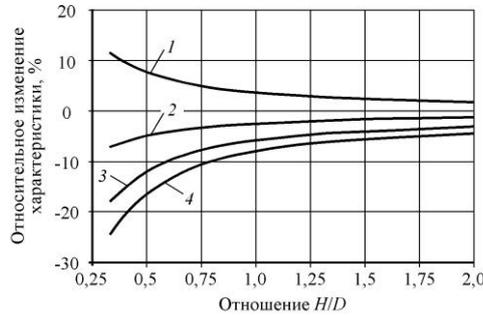


Рис. 3. Влияние отношения высоты H к диаметру D блока на характеристики преобразователя: 1 – частота продольного установочного резонанса; 2 – электрическая емкость; 3 – коэффициент преобразования по напряжению; 4 – коэффициент преобразования по заряду

При близких поперечном и продольном размерах блока отличия его характеристик для рассмотренных случаев крепления не превышают 10 %. Однако, при снижении отношения высоты блока к его диаметру, разница становится все более существенной. Значение КВК для блока с отношением $H/D \geq 0,333$ при жестком креплении к базе может быть определено из формулы:

$$k_{\kappa,i} = \frac{h(k_{h,i} + \frac{H}{h} - 1)}{H} = \frac{hk_{h,i} + H - h}{H}, \quad (1)$$

где $h = 0,333D$, $k_{h,i}$ – КВК для слоя высотой h (индекс « i » обозначает соответствующую характеристику преобразователя).

Рассчитанные значения $k_{h,i}$ составляют: $k_{h,U} = 0,823$ для коэффициента преобразования по напряжению, $k_{h,Q} = 0,757$ для коэффициента преобразования по заряду, $k_{h,C} = 0,9298$ для электрической емкости, $k_{h,f} = 1,1179$ для частоты продольного установочного резонанса.

Так как численное значение $h/D = 0,333$ соответствует среднему справочному значению коэффициента Пуассона пьезоэлектрической керамики ЦТС-19, формула (1) может быть представлена в виде:

$$k_{\kappa,i} = \frac{\mu D(k_{h,i} + \frac{H}{\mu D} - 1)}{H} = \frac{\mu Dk_{h,i} + H - \mu D}{H}, \quad (2)$$

где μ – коэффициент Пуассона ПКМ.

Следует отметить, что на практике условия крепления преобразователя всегда будут отличаться от принятых в рассмотренной модели. При этом его деформируемое состояние будет зависеть от механических свойств контактирующих тел, продольной и сдвиговой контактных жесткостей на границе сопрягаемых поверхностей, геометрических параметров основания и других факторов, которые в полном объеме невозможно учесть при моделировании.

Экспериментальное исследование. Целью экспериментального исследования являлась количественная оценка КВК для случаев, приближенных к практическому выполнению датчиков.

Поскольку для определения КВК необходимы данные, соответствующие свободно-скользящему и жесткому креплению преобразователя, в идеальном случае не реализуемые на практике, разработаны способы, позволяющие максимально приблизиться к таким типам крепления.

Для измерения коэффициентов преобразования и электрической емкости при имитации свободно-скользящего крепления испытуемый преобразователь в виде продольно поляризованного монолитного блока 4 (рис. 4,а) из пьезоэлектрической керамики с электродами 5 на торцах. Блок упруго поджат к основанию 7 через прокладку 6 резьбовой шпилькой 1 с гайкой через пружину 2, изолированную от электрода изолятором 3. Прокладка 6 высотой $h_n > 3,33D$, выполненная из той же неполяризованной пьезоэлектрической керамики, практически исключает поперечное механическое шунтирование в примыкающей к ней области преобразователя [19] и обеспечивает условия, максимально приближенные к его свободно-скользящему креплению.

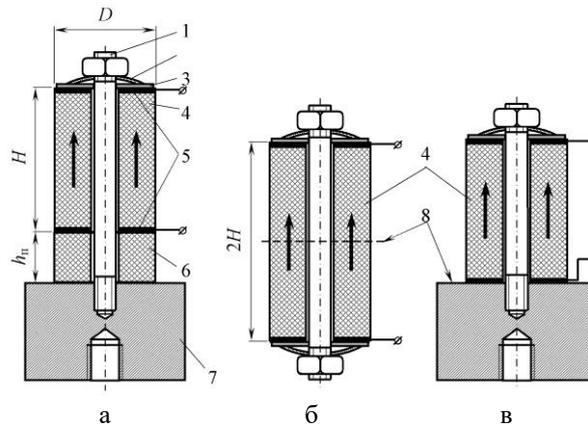


Рис. 4. Конструктивные схемы исследуемых образцов при испытании в режимах свободно-скользящего (а, б) и жесткого (в) крепления; направление поляризации показано стрелками

При определении частоты продольного установочного резонанса преобразователей в режиме, имитирующем свободно-скользящее крепление, измерялась частота резонанса блока с симметрично расположенными элементами крепления (рис. 4,б), который является полуволновым вибратором высотой $2H$ с узловой плоскостью 8. Частота его резонанса равна частоте четвертьволнового вибратора высотой H при свободно-скользящем креплении. При измерении характеристик преобразователя с реально выполненным жестким креплением использовалась схема согласно рис. 4,в.

Образцы исследуемых преобразователей имели следующие параметры: монолитный блок 4 преобразователя диаметром $D = 10$ мм с осевым отверстием диаметром 3,2 мм и воженными серебряными электродами 5 толщиной 10–12 мкм выполнен из пьезоэлектрической керамики ЦТС-19; дискретные значения высоты H составляли 3,3, 5, 7,5, 10, 15, 20 мм; прокладка 6 $\varnothing 10 \times 5$ мм; основание 7 $\varnothing 18 \times 10$ мм из конструкционных материалов, наиболее часто используемых в пьезоэлектрических датчиках (табл. 2); пружины 2 из сплава 36НХТЮ; изоляторы 3 из слюды 0,5 мм. Усилие упругого поджатия элементов составляло 80 и 120 Н, шероховатость сопрягаемых поверхностей элементов $R_a 0,63$.

Таблица 2

Параметры конструкционных материалов, используемых в исследуемых образцах (усредненные справочные значения)

Материал	Модуль Юнга, ГПа	Модуль сдвига, ГПа	Плотность, кг/м ³	Коэффициент Пуассона
ПКМ ЦТС-19	73	27,24	7600	0,340
Дюралюминий Д16Т	74	27,51	2700	0,345
Латунь Л62	97	35,93	8430	0,350
Сплав титановый ВТ3-1	115	43,56	4500	0,320
Сталь 12Х18Н10Т	200	76,34	7770	0,310
Сплав ВНМ 3-2	300	115,38	17700	0,30

Измерение коэффициентов преобразования образцов при двух типах крепления проводилось с использованием системы эталонной калибровки CS18STF производства СРЕКТРА (ФРГ). Электрическая емкость определялась как отношение $C = K_Q / K_U$ (здесь K_Q – коэффициент преобразования по заряду, K_U – коэффициент преобразования по напряжению).

Частоты продольного резонанса образцов, выполненных в соответствии с рис. 5б и 5в определялась по частотной характеристике модуля проводимости с присоединенной к основанию массой 80 г.

В табл. 3 приведены экспериментально полученные значения КВК для блока высотой $0,333D$ при усилении упругого поджатия 80 Н.

Таблица 3

Экспериментально полученные значения коэффициентов $k_{h,i}$

Коэффициент $k_{h,i}$	Материал основания				
	Д16Т	Л62	ВТ 3-1	12Х18Н10Т	ВНМ 3-2
$k_{h,U}$	0,997	0,998	0,993	0,894	0,857
$k_{h,Q}$	1,030	1,026	0,953	0,869	0,811
$k_{h,C}$	1,033	1,028	1,021	0,972	0,946
$k_{h,f}$	0,924	0,941	0,963	1,022	1,071

На рис. 5 показана зависимость характеристик преобразователя от отношения его высоты H к диаметру D и материала основания при усилении упругого поджатия 80 Н.

Из полученных экспериментальных данных следует, что параметры материалов основания, используемых в реальных конструкциях датчиков, приводят к заметному изменению их характеристик. Коэффициент преобразования по напряжению может снижаться до 14 %, коэффициент преобразования по заряду изменяется приблизительно от 3 до минус 19 %, электрическая емкость – от 3 до минус 5 % и частота продольного установочного резонанса – от минус 8 до 7 %.

Влияние граничных условий крепления, применительно к свойствам конструкционных материалов, снижается одновременно с возрастанием относительной высоты преобразователя. Зависимость характеристик от отношения продольного размера преобразователя к поперечному (рис. 5) соответствует формулам (1) и (2).

Характерным является тот факт, что конструкционные материалы с более высокими значениями модуля упругости, вызывают снижение коэффициентов преобразования и электрической емкости, способствуя росту частоты установоч-

ного резонанса. Материалы, по модулю упругости соизмеримые с пьезоэлектрической керамикой, приводят к обратному эффекту, вызываемому расширением приграничного слоя за счет пониженной контактной жесткости [20].

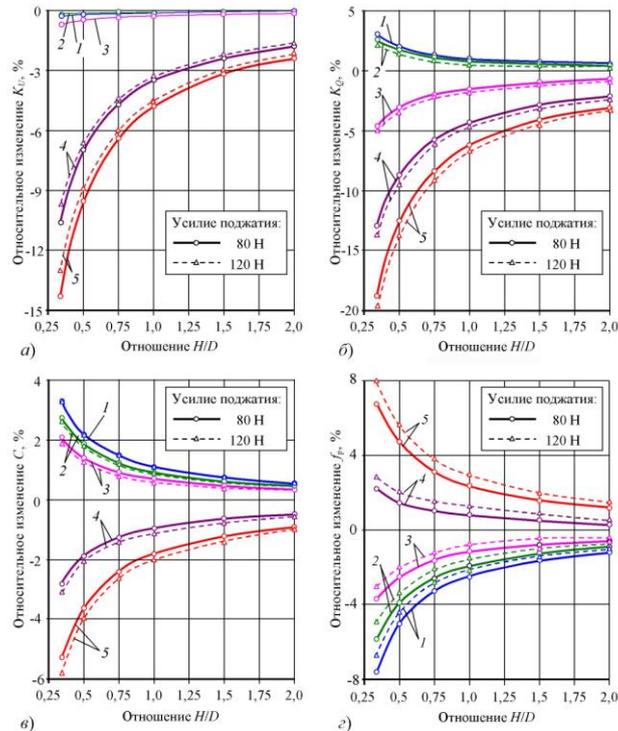


Рис. 5. Зависимость коэффициентов преобразования по напряжению K_U (а), по заряду K_Q , (б), электрической емкости C (в) и частоты установочного резонанса f_p (е) от отношения H к D блока при различных материалах основания: 1 – Д16Т, 2 – Л62, 3 – ВТ 3-1, 4 – 12Х18Н10Т, 5 – ВНМ 3-2

Сравнение экспериментальных результатов с полученными при моделировании показывает, что фактическая контактная жесткость на границе сопряжения составляет 0,85–0,9 от абсолютного значения. Это так же подтверждается некоторым влиянием усилия упругого поджатия на характеристики преобразователя.

Заключение. Проведенные исследования позволили количественно оценить влияние геометрических параметров преобразователя и свойств конструкционных материалов основания датчика на его основные характеристики. Так, использование сплава ВНМ 3-2 с максимальными значениями модуля упругости и плотности по сравнению с Д16Т, обладающим минимальными значениями этих параметров, позволяет повысить частоту установочного резонанса до 15 %. Однако это неизбежно приводит к снижению коэффициентов преобразования до 14 % по напряжению и до 22 % по заряду.

Экспериментально полученные значения КВК, применительно к различным конструкционным материалам, и аналитические выражения, связывающие КВК с относительными размерами преобразователя в виде монолитного блока, могут использоваться с целью повышения точности расчета основных метрологических характеристик пьезоэлектрических датчиков механических величин на 5–20 %, а также оптимизации их конструкций.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Датчики для измерения, контроля, диагностики и управления физических и технологических процессов: справочник: в 2 т. / под общ. ред. А.В. Гориша / под ред. А.Г. Дмитриенко, А.Н. Котова, Ю.Н. Макарова, С.А. Пономарева. Т. 1. Кн. 1. – М.: ФГБОУ ВПО МГУЛ, 2012. – 576 с.
2. *Crocker M.J., Arenas J.P.* Engineering Acoustics: Noise and Vibration Control. – Springer, 2012. – 706 p.
3. *Костюков В.Н., Науменко А.П.* Основы виброакустической диагностики и мониторинга машин. – Омск: ОмГТУ, 2011. – 360 с.
4. *Мокров Е.А.* Состояние, проблемы и пути развития датчикоостроения // Датчики и системы – 2006: Сб. трудов Всероссийской научно-практической конференции (Москва, 30–31 мая 2006 г.). – Пенза: ФНПЦ ФГУП «НИИ физических измерений», 2006. – С. 6-14.
5. *Гориш А.В., Панич А.Е., Свирская С.Н., Янчик В.В.* Перспективы развития пьезоэлектрических датчиков механических величин для РКТ и других областей // Сб. тр. науч.-техн. конф. «Информационно-измерительная техника». – М.: РУНД, 2014. – С. 282-293.
6. *Богущи М.В.* Проектирование пьезоэлектрических датчиков на основе пространственных электроупругих моделей. (Пьезоэлектрическое приборостроение Т. IX.). – М.: Техносфера, 2014. – 312 с.
7. *Янчик В.В., Панич А.Е.* Двухпараметровый пьезоэлектрический датчик поступательного и углового ускорения // Известия ЮФУ. Технические науки. – 2016. – № 2. – С. 191-199.
8. *Янчик В.В., Янчик Вл.В.* Преобразователи пьезоэлектрических датчиков механических величин (конструкции и пути развития). – Saarbrücken, Deutschland: LAP, 2013. – 142 с.
9. *Serridge M., Licht T.R.* Piezoelectric accelerometers and vibration preamplifiers: theory and application handbook. – Nerum, Denmark: Brüel and Kjær, 1987. – 150 p.
10. *Levinzon F.* Piezoelectrics Accelerometers with Integral Electronics. – NYC: Springer, 2015. – 168 p.
11. *Митько В.Н., Крамаров Ю.А., Панич А.А.* Математическое моделирование физических процессов в пьезоэлектрическом приборостроении. (Пьезоэлектрическое приборостроение. Т. 6). – Ростов-на-Дону: Изд-во ЮФУ, 2009. – 240 с.
12. *Янчик Вл.В., Митько В.Н.* Исследование механических напряжений и электрических полей в пьезоэлектрическом преобразователе акселерометра методом конечных элементов // Сб. тез. VIII Всерос. науч.-техн. конф. «Актуальные проблемы пьезоэлектрического приборостроения». – Ростов-на-Дону: Изд-во ЮФУ, 2012. – С. 50-54.
13. *Han-Chin Wu.* Continuum Mechanics and Plasticity. – Taylor & Francis, 2004. – 704 p.
14. *Янчик В. В.* Пьезоэлектрические виброизмерительные преобразователи (акселерометры). (Пьезоэлектрическое приборостроение. Т. 7). – Ростов-на-Дону: Изд-во ЮФУ, 2010. – 304 с.
15. *Liu B., Tcherniak D., Jacobson N., Olsen M.* Smart Setap and Accelerometer Mounting Check for Vibration Measurements // Sensors and Instrumentation. Proceedings of the 34th IMAC, A Conference and Exposition on Structural Dynamics. – 2016. – Vol. 5. – P. 9-34.
16. *Veber M.* Piezoelektrische Beschleunigungs aufnehmer. – MFF, 2012. – 55 s.
17. Acoustic and vibration transducers. Reliability and precision based on superior technology. – Denmark: Brüel & Kjær. 1996. – 60 p.
18. *Панич А.Е., Тополов В.Ю.* Физика сегнето- и пьезоэлектриков. – Ростов-на-Дону: Изд-во ЮФУ, 2009. – 171 с.
19. А.с. SU 1103161. Устройство для измерения продольного пьезомодуля / В.К. Доля, О.П. Крамаров, Ю.А. Крамаров. 1984, бюл. № 26.
20. *Robert M. Jones.* Deformation Theory of Plasticity. – Bull Ridge Publ, 2009. – 621 p.

REFERENCES

1. Datchiki dlya izmereniya, kontrolya, diagnostiki i upravleniya fizicheskikh i tekhnologicheskikh protsessov: spravochnik: v 2 t. [Sensors for measurement, control, diagnostics and management of physical and technological processes: handbook in 2 vol.], under the general ed. A.V. Gorisha, ed. by A.G. Dmitrienko, A.N. Kotova, Yu.N. Makarova, S.A. Ponomareva. Vol. 1. B. 1. Moscow: FGBOU VPO MGUL, 2012, 576 p.

2. Crocker M.J., Arenas J.P. Engineering Acoustics: Noise and Vibration Control. Springer, 2012, 706 p.
3. Kostyukov V.N., Naumenko A.P. Osnovy vibroakusticheskoy diagnostiki i monitoringa mashin [Basics of vibroacoustic diagnostics and monitoring of the machines: tutorial]. Omsk: OmGTU, 2011, 360 p.
4. Mokrov E.A. Sostoyaniye, problemy i puti razvitiya datchikostroeniya [Condition, problems and development ways of sensor engineering], *Datchiki i sistemy – 2006: Sb. trudov Vserossiyskoy nauchno-prakticheskoy konferentsii (Moskva, 30–31 maya 2006 g.)* [Sensors and Systems-2006: Proceedings of the All-Russian Scientific and Practical Conference (Moscow, May 30-31, 2006)]. Penza: FNPTS FGUP «NII fizicheskikh izmereniy», 2006, pp. 6-14.
5. Gorish A.V., Panich A.E., Svirskaya S.N., Yanchich V.V. Perspektivy razvitiya p'ezoelektricheskikh datchikov mekhanicheskikh velichin dlya RKT i drugikh oblastey [The development prospects of piezosensors of mechanical values for rocket and space technique and other fields], *Sb. tr. nauch.-tekhn. konf. «Informatsionno-izmeritel'naya tekhnika»* [Proceedings of the scientific and technical conference "Information and Measurement Technology"]. Moscow: RUND, 2014, pp. 282-293.
6. Bogush M.V. Proektirovaniye p'ezoelektricheskikh datchikov na osnove prostranstvennykh elektrouprugikh modeley. (P'ezoelektricheskoye priborostroeniye T. IX.) [Design of piezoelectric sensors based on spatial electroelastic models. (Piezoelectric instrument making Vol. IX)]. Moscow: Tekhnosfera, 2014, 312 p.
7. Yanchich V.V., Panich A.E. Dvukhparametrovyy p'ezoelektricheskyy datchik postupatel'nogo i uglovogo uskoreniya [Two-parameter piezoelectric sensor of progressive and angular acceleration], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2016, No. 2, pp. 191-199.
8. Yanchich V.V., Yanchich V.I. Preobrazovateli p'ezoelektricheskikh datchikov mekhanicheskikh velichin (konstruktsii i puti razvitiya) [The transducers of mechanical values piezosensors (constructions and development ways)]. Saarbrücken, Deutschland: LAP, 2013, 142 p.
9. Serridge M., Licht T.R. Piezoelectric accelerometers and vibration preamplifiers: theory and application handbook. Nerum, Denmark: Brüel and Kjær, 1987, 150 p.
10. Levinzon F. Piezoelectrics Accelerometers with Integral Electronics. NYC: Springer, 2015, 168 p.
11. Mit'ko V.N., Kramarov Yu.A., Panich A.A. Matematicheskoye modelirovaniye fizicheskikh protsessov v p'ezoelektricheskoye priborostroeniye. (P'ezoelektricheskoye priborostroeniye. T. 6) [Mathematical modeling of physical processes in piezoelectric instrumentation. (Piezoelectric instrumentation. Vol. 6)]. Rostov-on-Don: Izd-vo YuFU, 2009, 240 p.
12. Yanchich V.I., Mit'ko V.N. Issledovaniye mekhanicheskikh napryazheniy i elektricheskikh poley v p'ezoelektricheskoye preobrazovatele akselerometra metodom konechnykh elementov [The research of mechanical stresses and electric fields in piezosensor of accelerometer by the final elements method], *Sb. tez. VIII Vseros. nauch.-tekhn. konf. «Aktual'nye problemy p'ezoelektricheskogo priborostroeniya»* [Digest of theses of the VIII All-Russian scientific and technological conference «The piezoelectric instrumentation actual problems»]. Rostov-on-Don: Izd-vo YuFU, 2012, pp. 50-54.
13. Han-Chin Wu. Continuum Mechanics and Plasticity. Taylor & Francis, 2004, 704 p.
14. Yanchich V.V. P'ezoelektricheskoye vibrozmeritel'nyye preobrazovateli (akselerometri). (P'ezoelektricheskoye priborostroeniye. T. 7) [Piezoelectric vibration measuring transducers (accelerometers). (Piezoelectric instrument making. Vol. 7)]. Rostov-on-Don: Izd-vo YuFU, 2010, 304 p.
15. Liu B., Tcherniak D., Jacobson N., Olsen M. Smart Setup and Accelerometer Mounting Check for Vibration Measurements, *Sensors and Instrumentation. Proceedings of the 34th IMAC, A Conference and Exposition on Structural Dynamics*, 2016, Vol. 5, pp. 9-34.
16. Veber M. Piezoelektrische Beschleunigungsaufnehmer. MFF, 2012, 55 p.
17. Acoustic and vibration transducers. Reliability and precision based on superior technology. Denmark: Brüel & Kjær. 1996, 60 p.
18. Panich A.E., Topolov V.Yu. Fizika segneto- i p'ezoelektrikov [Physics of ferroelectrics and piezoelectrics]. Rostov-on-Don: Izd-vo YuFU, 2009, 171 p.

19. A.s. SU 1103161. Ustroystvo dlya izmereniya prodol'nogo p'ezomodulya [Patent SU 1103161. Ustroystvo dlya izmereniya prodol'nogo piezomodulya [Longitudinal piezomodule], *V.K. Dolya, O.P. Kramarov, Yu.A. Kramarov*. 1984, bull. No. 26.
20. *Robert M. Jones*. Deformation Theory of Plasticity. Bull Ridge Publ, 2009, 621 p.

Статью рекомендовал к опубликованию д.т.н. М.В. Богуш.

Янчик Владимир Владимирович – Южный федеральный университет; e-mail: vibro1@mail.ru; г. Ростов-на-Дону, Россия; тел.: +79885344996; институт высоких технологий и пьезотехники; ведущий инженер.

Yanchich Vladimir Vladimirovich – Southern Federal University; e-mail: vibro1@mail.ru; Rostov-on-Don, Russia; phone: +79885344996; Institute of hightechnologies and piezotechniques; leading engineer.

Раздел IV. Информационный анализ и распознавание образов

УДК 004.932.2

DOI 10.18522/2311-3103-2021-2-141-154

А.Н. Каркищенко, В.Б. Мнухин

ОЦЕНКА ПОГРЕШНОСТИ ПРИ МНОЖЕСТВЕННОМ СРАВНЕНИИ ЗАШУМЛЕННЫХ ИЗОБРАЖЕНИЙ*

Целью работы является исследование влияния зашумления на изображения на результат сравнения конечного множества изображений, одинаковых по форме и размеру. Данная задача неизбежно возникает при анализе сцен, детекции отдельных объектов, обнаружении симметрии и пр. Фактор зашумленности необходимо принимать во внимание, поскольку различие цифровых объектов может быть вызвано не только несопадением сравниваемых изображений реальных объектов, но и искажениями из-за шумов, что практически всегда имеет место. Это отличие оказывается пропорциональным уровню шумовой составляющей. Основным результатом данной статьи являются аналитическая оценка для вероятности заданного уровня погрешности, которая может возникать при множественном сравнении конечного числа соразмерных цифровых изображений. Эта оценка основана на низкоуровневом анализе, сводящемся к попиксельному вычислению различия изображений с помощью евклидовой метрики. При этом делается стандартное предположение о независимом нормальном зашумлении интенсивностей изображения с нулевым математическим ожиданием и априорно установленным среднеквадратическим отклонением в каждом пикселе. Приведенные в статье доказательства позволяют утверждать, что полученную оценку следует рассматривать как достаточно «осторожную» и можно ожидать, что в реальности разброс меры, вызванный шумами на изображении, будет существенно меньше, чем теоретически найденная граница. Полученные в данной работе оценки оказываются полезными также для обнаружения различных видов симметрии на изображениях, которое, как правило, приводит к необходимости вычислять различие произвольного количества соразмерных цифровых областей. Кроме того, найденные оценки могут использоваться как теоретически обоснованные пороговые значения в задачах, требующих принятия решения о совпадении или различии изображений. Такие пороговые значения неизбежно появляются на различных этапах обработки зашумленных изображений, и вопрос об их конкретных значениях, как правило, остается открытым, в лучшем случае предлагаются эвристические соображения для их выбора.

Сравнение цифровых изображений; симметрия изображений; нормальное зашумление; оценка погрешности; оценка вероятности ошибки зашумления.

A.N. Karkishchenko, V.B. Mnukhin

ERROR ESTIMATION FOR MULTIPLE COMPARISON OF NOISY IMAGES

The aim of this work is to study the effect of noise on the image on the quality of comparison of a finite set of images of the same shape and size. This task inevitably arises when analyzing scenes, detecting individual objects, detecting symmetry, etc. The noise factor must be taken into account, since the difference between digital objects can be caused not only by the mismatch of the compared images of real objects, but also by distortions due to noise, which is almost always takes place. This difference

* Работа выполнена при финансовой поддержке РФФИ, проект № 19-07-00873.

turns out to be proportional to the level of the noise component. The main result of this article is an analytical estimate for the probability of a given level of error, which may arise in the multiple comparison of a finite set of commensurate digital images. This estimate is based on a low-level comparison, which is a pixel-by-pixel calculation of image differences using the Euclidean metric. In this case, a standard assumption is made about the independent normal noise of image intensities with zero mathematical expectation and a priori established standard deviation in each pixel. The evidence presented in the article allows us to assert that the obtained estimate should be regarded as sufficiently "cautious" and it can be expected that in reality the scatter of the measure caused by noise in the image will be significantly less than the theoretically found boundary. The estimates obtained in this work are also useful for detecting various types of symmetry in images, which, as a rule, lead to the need to calculate the difference of an arbitrary number of commensurate digital areas. In addition, they can be used as theoretically grounded threshold values in tasks requiring a decision on the coincidence or difference of images. Such threshold values inevitably appear at various stages of processing noisy images, and the question of their specific values, as a rule, remains open; at best, heuristic considerations are proposed for their selection.

Comparison of digital images; image symmetry; normal noise; error estimate; noise error probability estimate.

Введение. В задачах анализа изображений, как правило, возникает необходимость детекции отдельных объектов, присутствующих на изображении. Обнаружение таких объектов является необходимым этапом при формировании общего «понимания» сцены, представленной на изображении [1]. Детекция объектов и дальнейшее их распознавание, как правило, сводится к сопоставлению части изображения с некоторым эталонным изображением, имеющимся в базе известных объектов [2]. Подобное сопоставление сводится к выявлению и измерению сходства или, в двойственной интерпретации, – различия определенных областей на изображении с изображением известного объекта. Для количественной оценки различия двух частей изображения X_1 и X_2 , имеющих в геометрическом смысле одинаковую форму и одинаковые размеры, обычно вводится некоторая метрика $d(X_1, X_2)$ [3, стр. 167-185], [4]. В наиболее употребительном случае $d(X_1, X_2)$ – евклидова метрика, применяемая попиксельно.

В более сложном случае требуется оценить разброс предположительно одинаковых объектов X_1, X_2, \dots, X_m . В этом случае, опираясь на введенную метрику $d(X_i, X_j)$, позволяющую вычислять попарное различие произвольных объектов X_i и X_j , необходимо ввести некоторую функцию, агрегирующую эти различия на всем множестве. Необходимость в этом возникает, например, в кластерном анализе при оценке параметров образующихся кластеров [5, стр. 22]. В частности, для характеристики кластера вводится так называемая мера общего рассеяния

$s_d = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m d(X_i, X_j)$, которая представляет собой сумму всех попарных расстояний между объектами X_1, X_2, \dots, X_m . С этой мерой связана мера среднего рассеяния $\bar{s}_d = \frac{2s_d}{m(m-1)}$, которая является просто усреднением меры общего рассеяния.

Введение мер рассеяния (или разнородности) существенно опирается на содержательный смысл решаемой задачи анализа совокупности объектов. В случае анализа разброса объектов, вызванного естественными зашумлениями, для получения гарантированных с заданной вероятностью характеристик разброса целесообразно характеризовать этот разброс *диаметром* M сферы, с заданной вероятностью содержащей внутри себя все регистрируемые объекты X_1, X_2, \dots, X_m :

$$M = \max_{1 \leq j < i \leq m} d(X_i, X_j).$$

Типичный случай, когда реально возникает такая задача характеристики множественного рассеяния – определение симметрии объектов на изображении. Обнаружение и анализ симметрии – важная задача в таких областях как машинное зрение, медицинская диагностика, классификация паттернов, визуальное восприятие, дизайн и проектирование и др. [6, 7]. В большинстве случаев установление наличия любого вида симметрии приводит к необходимости анализа всей совокупности попарных различий блоков изображений (см., напр. [8–12]).

Если, например, объект имеет *вращательную* симметрию порядка m , то можно считать, что он состоит из m одинаковых блоков изображений, совпадающих друг с другом при действии любого оператора φ_k поворота на угол $\frac{2\pi k}{m}$, $k = 0, 1, \dots, m-1$, вокруг центра симметрии. В идеальном случае для любой метрики d эти блоки должны удовлетворять условиям $d(X_i, \varphi_k X_j) = 0$, $k = 0, 1, \dots, m-1$, $i, j \in \{1, 2, \dots, m\}$. Установление степени нарушения этих равенств собственно и представляет собой задачу анализа вращательной симметрии в заданной точке.

Аналогичная задача возникает и при обнаружении *трансляционной* симметрии. В этом случае такой вид симметрии описывается бесконечной циклической абелевой группой. Следовательно, в идеальном случае это должно проявляться в наличии бесконечного числа одинаковых блоков на изображении X_1, X_2, \dots (фундаментальных областей этой группы). На реальном изображении этих блоков может быть зарегистрировано лишь конечное число X_1, X_2, \dots, X_m , поэтому приходится предполагать, что это лишь наблюдаемая конечная совокупность симметричных блоков трансляционно-симметричного изображения. Поэтому теоретически различие между любой парой таких блоков также должно быть равно нулю по любой метрике.

Анализ *зеркальной* симметрии приводит к необходимости вычислять различие лишь двух блоков X_1 и X_2 , т.е. должно выполняться $d(X_1, rX_2) = 0$, где r – оператор отражения относительно некоторой оси.

В идеальном случае симметричные части объекта или фигуры должны в точности совпадать при наложении. Таких симметричных частей может быть много – в случае отражательной симметрии – две, в случае вращательной – столько, каков порядок вращательной симметрии, в случае трансляционной симметрии – бесконечное количество (на практике, как отмечалось, конечное число наблюдаемых симметричных частей). В реальности меры различия никогда не обращаются в нуль по двум основным причинам: во-первых, объекты могут иметь существенные отличия (например, из-за реального несовпадения или, как в приведенном примере, из-за отсутствия полной симметрии), а во-вторых, из-за наличия естественных шумов, неизбежно присутствующих на изображении и поэтому исключающих возможность точного решения задачи. В работе [13], посвященной изучению влияния гауссовского зашумления на изображения, отмечается, что с помощью предварительной фильтрации можно снизить уровень шума на изображении, однако от него невозможно избавиться полностью, зашумление все равно сохраняется, независимо от того какие алгоритмы фильтрации применяются.

В связи со сказанным выше возникает основной вопрос – в каком случае различие между сравниваемыми объектами вызвано их существенным несовпадением, а в каком – это лишь следствие естественного зашумления.

Частично ответ на этот вопрос состоит в том, что если вычисленная мера различия объектов не превосходит уровень различия, которое вызвано шумами, то можно делать правдоподобный вывод о том, что сравниваемые цифровые объекты совпадают. Так, в случае анализа симметричности можно с очень большой вероятностью утверждать, что объект обладает заданным типом симметрии. Основная задача, которая при этом возникает и требует теоретического решения, – определить заранее максимальное значение различия объектов, которое обусловлено шумами.

В большинстве случаев под «зашумлением» понимается упрощенная и идеализированная модель возникающих искажений в цифровых изображениях реальных объектов, не учитывающая многих факторов, влияющих на яркость в каждом пикселе, в том числе и влияние пространственных характеристик окрестности пикселя. Тем не менее, в системах машинного зрения используются простые и достаточно эффективные модели зашумления. Для полутоновых изображений, как правило, рассматривают так называемую аддитивную модель шума, в которой считается, что зашумленное изображение $\text{Im}(x, y)$ в пикселе с координатами (x, y) образуется из незашумленного изображения $\text{Im}(x, y)$ по закону $\text{Im}(x, y) = \text{Im}(x, y) + \xi(x, y)$, где $\xi(x, y)$ – случайная шумовая составляющая. В большинстве приложений зависимость шума от координат пикселя (пространственный фактор) считается несущественной. Исходя из *закона больших чисел*, закон распределения аддитивной шумовой компоненты описывается однопараметрическим семейством *нормальных* распределений $\mathcal{N}(0, \sigma)$ с нулевым математическим ожиданием и среднеквадратическим отклонением σ . Таким образом, считается, что $\xi(x, y)$ – случайная величина, которая подчиняется нормальному распределению, т.е. $\xi(x, y) \sim \mathcal{N}(0, \sigma)$.

Теоретическая оценка погрешностей, связанных с анализом зашумленных изображений, представляет в большинстве случаев непростую задачу. По этой причине большинство известных публикаций, посвященных разработке методов и алгоритмов обработки цифровых изображений, содержат только экспериментальную оценку точности методов. В тех случаях, когда анализ погрешностей проводится, обычно прибегают к статистическим оценкам на основе этих экспериментальных исследований. Данный подход не позволяет получить обоснованные теоретические выводы при всевозможных значениях параметров, описывающих как изображение, так и саму процедуру распознавания.

Теоретические оценки и методы их получения существенно зависят от используемой меры близости изображений $d(X_i, X_j)$. Традиционной и наиболее употребительной является евклидова метрика. Вместе с тем, известен ряд более сложных в вычислительном плане метрик – нормализованная метрика Хаусдорфа, метрика на основе локального расстояния, метрика на глобальных признаках, метрика на основе симметрии [14]. Эти метрики учитывают также пространственные факторы окрестности, влияющие на яркость пикселей, поэтому иногда воспринимаются как в большей степени соответствующие человеческому восприятию различий.

В частном случае, для вращательной симметрии третьего порядка задача теоретического получения верхней оценки погрешности была решена в работе авторов [15]. В данной статье рассматривалась гексагональная пиксельная решетка, которая обладает рядом преимуществ перед традиционной квадратной решеткой [16]. Одним из свойств гексагональной решетки является ее естественная симметрия относительно группы вращений третьего порядка.

Обнаружение вращательной симметрии третьего порядка имеет важные практические приложения. Распознавание симметрии 3-го порядка востребовано в различных областях кристаллографии [17, 18], вирусологии [19], анализе изображений, полученных с электронного микроскопа [20] и др.

Данная работа обобщает полученный в [15] результат по оценке диаметра рассеяния на случай произвольного конечного числа анализируемых паттернов. Полученная в работе оценка погрешности применима, в том числе, для других типов симметрии, которые приводят к необходимости вычислять различие произвольного количества соразмерных цифровых областей одинакового размера.

Постановка задачи. Пусть требуется осуществить сравнение m одинаковых по форме и размерам цифровых блоков X_1, X_2, \dots, X_m . Введем в этих блоках произвольную, но фиксированную и единообразную одноиндексную нумерацию пикселей, например слева-направо и сверху-вниз. Это позволяет представить заданное полутоновое изображение в виде вектора интенсивностей яркости. Будем считать, что количество пикселей в каждом блоке, значит, и размерность вектора равны n . Таким образом, вычисление расстояния между парой цифровых изображений сводится к вычислению расстояния между представляющими их векторами $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m \in R^n$.

В силу сделанных выше предположений о характере аддитивного зашумления будем считать, что ε_i – независимые случайные векторы, компоненты которых зашумлены по нормальному закону $\mathcal{N}(0, \sigma^2)$ с нулевым математическим ожиданием и дисперсией компонент σ^2 . Требуется найти вероятностную оценку для диаметра сферы рассеяния, т.е. для величины

$$M = \max_{1 \leq j < i \leq m} \|\varepsilon_j - \varepsilon_i\|,$$

где $\|\cdot\|$ – эвклидова норма.

Оценка влияния шума на сравниваемые изображения. Рассмотрим случайные векторы $\mu_{ij} = \varepsilon_j - \varepsilon_i$, $1 \leq j < i \leq m$. Заметим, что они могут быть выражены как линейные комбинации векторов $\mu_{m1}, \mu_{m2}, \dots, \mu_{m,m-1}$, т.е. являются неслучайными функциями случайных векторов. Рассмотрим блочные векторы

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} \text{ и } \mu = \begin{pmatrix} \mu_{m1} \\ \mu_{m2} \\ \vdots \\ \mu_{m,m-1} \end{pmatrix}.$$

Введенные обозначения устанавливают простую связь между этими векторами, которая с помощью блочной матрицы может быть выражена следующим образом:

$$\mu = \begin{pmatrix} I_n & 0 & \dots & 0 & -I_n \\ 0 & I_n & \dots & 0 & -I_n \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & I_n & -I_n \end{pmatrix} \varepsilon = A\varepsilon,$$

где символ I_n обозначает единичную матрицу размера $n \times n$. Блочная структура матрицы A позволяет представить ее в виде кронекерова произведения:

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix} \otimes I_n. \quad (1)$$

В силу предположения о независимости ковариационные матрицы компонент случайных векторов $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ имеют вид $K_{\varepsilon_1} = K_{\varepsilon_2} = \dots = K_{\varepsilon_m} = \sigma^2 I_n$, значит, ковариационная матрица K_ε вектора ε равна $K_\varepsilon = \sigma^2 I_{mn}$, где I_{mn} – единичная матрица размера $mn \times mn$. Учитывая, что $\mu - M[\mu] = A(\varepsilon - M[\varepsilon])$, найдем ковариационную матрицу компонент вектора μ :

$$K_\mu = AM \left[(\varepsilon - M[\varepsilon])(\varepsilon - M[\varepsilon])^T \right] A^T = AK_\varepsilon A^T = \sigma^2 AA^T.$$

Далее будем пользоваться следующим свойством, связывающим кронекерово произведение матриц с обычным умножением матриц. Пусть A, B, C, D – матрицы такие, что произведения AC и BD определены. Тогда $(A \otimes B)(C \otimes D) = AC \otimes BD$.

Кроме того, справедливо очевидное равенство $(A \otimes B)^T = A^T \otimes B^T$ [21, стр. 19].

Применяя эти свойства и учитывая представление (1), получаем, что

$$\begin{aligned} AA^T &= \left(\begin{pmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix} \otimes I_n \right) \left(\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ -1 & -1 & \dots & -1 \end{pmatrix} \otimes I_n \right) = \\ &= \begin{pmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ -1 & -1 & \dots & -1 \end{pmatrix} \otimes I_n = \underbrace{\begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 2 \end{pmatrix}}_B \otimes I_n = B \otimes I_n. \end{aligned}$$

Заметим, что матрица B имеет, очевидно, размер $(m-1) \times (m-1)$. Таким образом, $K_\mu = \sigma^2 AA^T = \sigma^2 B \otimes I_n$.

Последнее выражение показывает, что векторы $\mu_{m1}, \mu_{m2}, \dots, \mu_{m,m-1}$ коррелированы (в случае нормального распределения это означает, что они также и статистически зависимы). Найдем декоррелирующее преобразование, позволяющее перейти к статистически независимым случайным векторам. Для этого перейдем к базису, состоящему из собственных векторов ковариационной матрицы K_μ . Поскольку матрица K_μ симметрична и положительно определена, то она имеет простую структуру, следовательно, указанный базис существует. Найдем его.

Для отыскания собственных векторов матрицы K_μ необходимо найти все ее собственные значения. Как известно [21, стр. 39], собственные значения кронекерова произведения $A \otimes B$ произвольных матриц A и B равны всевозможным произведениям собственных значений матриц A и B , а соответствующие собственные векторы получаются как кронекеровы произведения собственных векторов, отвечающих соответствующим собственным значениям. Непосредственными вычислениями нетрудно убедиться, что матрица $\sigma^2 B$ имеет собственное значение σ^2 кратности $m-2$ и однократное собственное значение $m\sigma^2$. Собственные значения удобно записать в виде диагональной матрицы $\Lambda = \text{diag}(\sigma^2, \dots, \sigma^2, m\sigma^2)$.

Найдем собственные векторы. Поскольку матрица $\sigma^2 B$ имеет простую структуру, то собственное значение σ^2 имеет геометрическую кратность $m-2$, следовательно, существует большой произвол в выборе ортонормированной системы собственных векторов, отвечающих данному собственному значению. Нетрудно убедиться, что ортонормированную систему собственных векторов матрицы $\sigma^2 B$ можно, в частности, выбрать в виде

$$T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3 \cdot 2}} & \frac{1}{\sqrt{4 \cdot 3}} & \cdots & \frac{1}{\sqrt{(m-1)(m-2)}} & \frac{1}{\sqrt{m-1}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3 \cdot 2}} & \frac{1}{\sqrt{4 \cdot 3}} & \cdots & \frac{1}{\sqrt{(m-1)(m-2)}} & \frac{1}{\sqrt{m-1}} \\ 0 & -\frac{2}{\sqrt{3 \cdot 2}} & \frac{1}{\sqrt{4 \cdot 3}} & \cdots & \frac{1}{\sqrt{(m-1)(m-2)}} & \frac{1}{\sqrt{m-1}} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sqrt{(m-1)(m-2)}} & \frac{1}{\sqrt{m-1}} \\ 0 & 0 & 0 & \cdots & -\frac{m-2}{\sqrt{(m-1)(m-2)}} & \frac{1}{\sqrt{m-1}} \end{pmatrix}.$$

В данной матрице первые $m-2$ столбцов являются собственными векторами, соответствующими собственному значению σ^2 , а последний столбец соответствует $m\sigma^2$. В силу ортонормированности матрица T удовлетворяет соотношениям $T^T T = I_{m-1}$. Таким образом, учитывая свойства кронекерова произведения, получаем, что все собственные значения матрицы K_μ могут быть записаны в виде диагональной матрицы $\Lambda \otimes I_n$, а соответствующие собственные векторы – в виде матрицы $T \otimes I_n$. Таким образом, $K_\mu(T \otimes I_n) = (\sigma^2 B \otimes I_n)(T \otimes I_n) = T \Lambda \otimes I_n$.

Рассмотрим теперь матрицу $\sqrt{\Lambda^{-1}} \otimes I_n = \text{diag}(\sigma^{-1}, \dots, \sigma^{-1}, (\sqrt{m})^{-1} \sigma^{-1}) \otimes I_n$ и введем преобразование $\nu = (\sqrt{\Lambda^{-1}} T^T \otimes I_n) \mu$. Тогда ковариационная матрица K_ν компонент вектора ν имеет вид:

$$\begin{aligned} K_\nu &= M[\nu \nu^T] = (\sqrt{\Lambda^{-1}} T^T \otimes I_n) M[\mu \mu^T] (T \sqrt{\Lambda^{-1}} \otimes I_n) = \\ &= (\sqrt{\Lambda^{-1}} T^T \otimes I_n) (\sigma^2 B \otimes I_n) (T \sqrt{\Lambda^{-1}} \otimes I_n) = \sqrt{\Lambda^{-1}} T^T (\sigma^2 B) T \sqrt{\Lambda^{-1}} \otimes I_n = I_{(m-1)n}. \end{aligned}$$

Матрица K_ν диагональная, следовательно, все компоненты вектора ν некоррелированы, т.е. $\nu = (\sqrt{\Lambda^{-1}} T^T \otimes I_n) \mu$ является искомым декоррелирующим преобразованием. Выразим вектор ν через исходный случайный вектор μ в явном виде:

$$\nu = \begin{pmatrix} \left(\begin{array}{cccccc} \frac{1}{\sigma} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma \sqrt{m}} \end{array} \right) \left(\begin{array}{cccccc} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{3 \cdot 2}} & \frac{1}{\sqrt{3 \cdot 2}} & \frac{-2}{\sqrt{3 \cdot 2}} & \cdots & 0 \\ \frac{1}{\sqrt{4 \cdot 3}} & \frac{1}{\sqrt{4 \cdot 3}} & \frac{1}{\sqrt{4 \cdot 3}} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{\sqrt{(m-1)(m-2)}} & \frac{1}{\sqrt{(m-1)(m-2)}} & \frac{1}{\sqrt{(m-1)(m-2)}} & \cdots & \frac{-(m-2)}{\sqrt{(m-1)(m-2)}} \\ \frac{1}{\sqrt{(m-1)}} & \frac{1}{\sqrt{(m-1)}} & \frac{1}{\sqrt{(m-1)}} & \cdots & \frac{1}{\sqrt{(m-1)}} \end{array} \right) \otimes I_n \right) \mu =$$

Докажем, что неравенство справедливо и при $m = k + 1$. С этой целью для произвольных $j, t, 1 \leq j, t \leq k$, запишем очевидное равенство:

$$\varepsilon_j - \varepsilon_{k+1} = (\varepsilon_t - \varepsilon_{k+1}) + (\varepsilon_j - \varepsilon_t) \text{ или } \mu_{k+1,j} = \mu_{k+1,t} + \mu_{ij}.$$

Просуммируем последнее выражение по t :

$$\mu_{k+1,j} = \frac{1}{k} \sum_{t=1}^k \mu_{k+1,t} + \frac{1}{k} \sum_{t=1}^k \mu_{ij}.$$

Применяя необходимое число раз неравенство треугольника, получаем:

$$\|\mu_{k+1,j}\| \leq \frac{1}{k} \left\| \sum_{t=1}^k \mu_{k+1,t} \right\| + \frac{1}{k} \left\| \sum_{t=1}^k \mu_{ij} \right\| \leq \frac{1}{k} \left\| \sum_{t=1}^k \mu_{k+1,t} \right\| + \frac{1}{k} \sum_{t=1}^k \|\mu_{ij}\|.$$

Заметим, что в правой части для любого $j = 1, 2, \dots, k$ норма $\|\mu_{ij}\| = 0$. Поэтому правая часть содержит не более k ненулевых слагаемых. Следовательно,

$$\|\mu_{k+1,j}\| \leq \max \left(\left\| \sum_{t=1}^k \mu_{k+1,t} \right\|, \|\mu_{1j}\|, \|\mu_{2j}\|, \dots, \|\mu_{kj}\| \right) = \max \left(\left\| \sum_{t=1}^k \mu_{k+1,t} \right\|, \max_{1 \leq i \leq k} \|\mu_{ij}\| \right).$$

Поскольку данное неравенство справедливо для любого $j = 1, 2, \dots, k$, то, учитывая, что $\|\mu_{ij}\| = \|\mu_{ji}\|$ и $\|\mu_{ii}\| = 0$ для всех i, j , получаем, что

$$\max_{1 \leq j \leq k} \|\mu_{k+1,j}\| \leq \max \left(\left\| \sum_{t=1}^k \mu_{k+1,t} \right\|, \max_{1 \leq j < i \leq k} \|\mu_{ij}\| \right).$$

Заметим теперь, что неравенство не нарушится, если переписать его следующим образом:

$$\max \left(\max_{1 \leq j \leq k} \|\mu_{k+1,j}\|, \max_{1 \leq j < i \leq k} \|\mu_{ij}\| \right) = \max_{1 \leq j < i \leq k+1} \|\mu_{ij}\| \leq \max \left(\left\| \sum_{t=1}^k \mu_{k+1,t} \right\|, \max_{1 \leq j < i \leq k} \|\mu_{ij}\| \right).$$

Если теперь воспользоваться предположением индукции и заменить $\max_{1 \leq j < i \leq k} \|\mu_{ij}\|$ в правой части последнего неравенства в соответствии с выражением (4), то получим доказываемое утверждение. ■

С учетом этого утверждения, вводя для удобства новую случайную величину L , можно записать:

$$M = \max_{1 \leq j < i \leq m} \|\varepsilon_j - \varepsilon_i\| = \max_{1 \leq j < i \leq m} \|\mu_{ij}\| \leq \max \left(\left\| \sum_{j=1}^{m-1} \mu_{mj} \right\|, \left\| \sum_{j=1}^{m-2} \mu_{m-1,j} \right\|, \dots, \|\mu_{21}\| \right) = L.$$

Вероятностная оценка сверху для диаметра сферы разброса. Основным результатом дает следующее

Утверждение 2. Для некоторого положительного δ имеет место следующая оценка вероятности P того, что диаметр M сферы разброса не превысит δ :

$$P(M \leq \delta) \geq P(L^2 \leq \delta^2) = \prod_{k=2}^m \Phi \left(\frac{\delta^2 - k(k-1)n\sigma^2}{\sqrt{2nk(k-1)\sigma^2}} \right),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$.

Доказательство. Как известно [22], случайные величины (2) имеют χ^2 -распределение с n степенями свободы. При этом χ^2 -распределение с ростом n асимптотически сходится к нормальному распределению $\mathcal{N}(n, 2n)$. Поэтому при достаточно большом значении n ($n > 50$, см., напр. [23]) можно считать, что $\frac{\gamma_i - n}{\sqrt{2n}} \sim \mathcal{N}(0, 1)$, $i = 1, 2, \dots, m-1$. Отсюда следует, что

$$\begin{aligned} \|\mu_{21}\|^2 &\sim \mathcal{N}(2n\sigma^2, 8n\sigma^4), \\ \|\mu_{31} + \mu_{32}\|^2 &\sim \mathcal{N}(6n\sigma^2, 72n\sigma^4), \\ &\dots \dots \dots \sim \dots \dots \dots \\ \|\mu_{m-1,1} + \mu_{m-1,2} + \dots + \mu_{m-1,m-2}\|^2 &\sim \mathcal{N}((m-1)(m-2)n\sigma^2, 2(m-1)^2(m-2)^2n\sigma^4), \\ \|\mu_{m1} + \mu_{m2} + \dots + \mu_{m,m-2} + \mu_{m,m-1}\|^2 &\sim \mathcal{N}(m(m-1)n\sigma^2, 2m^2(m-1)^2n\sigma^4). \end{aligned}$$

Но функция распределения максимума независимых случайных величин равна произведению функций распределения этих величин [24]. Поэтому

$$L^2 = \max \left(\left\| \sum_{j=1}^{m-1} \mu_{mj} \right\|^2, \left\| \sum_{j=1}^{m-2} \mu_{m-1,j} \right\|^2, \dots, \|\mu_{21}\|^2 \right) \sim \prod_{k=2}^m \mathcal{N}(k(k-1)n\sigma^2, 2k^2(k-1)^2n\sigma^4).$$

Тогда для некоторого положительного δ получаем следующую оценку вероятности P того, что M не превысит δ :

$$P(M \leq \delta) \geq P(L^2 \leq \delta^2) = \prod_{k=2}^m \Phi \left(\frac{\delta^2 - k(k-1)n\sigma^2}{\sqrt{2nk(k-1)}\sigma^2} \right),$$

где $\Phi(x)$ – функция Лапласа. ■

Следствие 1. *Имеет место неравенство*

$$P(M \leq \delta) > \Phi^{m-1} \left(\frac{\delta^2 - m(m-1)n\sigma^2}{\sqrt{2nm(m-1)}\sigma^2} \right).$$

Доказательство. Заметим, что при $k_1 < k_2$

$$\frac{\delta^2 - k_1(k_1-1)n\sigma^2}{\sqrt{2nk_1(k_1-1)}\sigma^2} > \frac{\delta^2 - k_2(k_2-1)n\sigma^2}{\sqrt{2nk_2(k_2-1)}\sigma^2}.$$

Поэтому в силу монотонности функции $\Phi(x)$ в рассматриваемой области

$$\prod_{k=2}^m \Phi \left(\frac{\delta^2 - k(k-1)n\sigma^2}{\sqrt{2nk(k-1)}\sigma^2} \right) > \Phi^{m-1} \left(\frac{\delta^2 - m(m-1)n\sigma^2}{\sqrt{2nm(m-1)}\sigma^2} \right). \blacksquare$$

Следствие 2. *Величина δ , которую M не превысит с вероятностью заданно большей α , т.е. $P(M \leq \delta) > \alpha$, равна*

$$\delta = \sigma \sqrt{m(m-1) \left(n + \sqrt{2n\Phi^{-1} \left(\sqrt[m-1]{\alpha} \right)} \right)}. \tag{5}$$

Доказательство. В силу Следствия 1 указанное условие можно записать в виде $\Phi^{m-1} \left(\frac{\delta^2 - m(m-1)n\sigma^2}{\sqrt{2nm(m-1)\sigma^2}} \right) = \alpha$, откуда следует доказываемое равенство. ■

Из формулы (5) видно, что верхняя оценка диаметра сферы, содержащей зашумленные значения линейно зависит от σ и почти линейно от числа m сравниваемых блоков изображений.

Пример. Рассчитаем с помощью Следствия 2 величину погрешности δ , которую M не превысит с вероятностью заведомо большей $\alpha = 0.99$ при условии, что $m = 5$ и $\sigma = 0.01$. По формуле (5) получаем

$$\begin{aligned} \delta &= 0.01 \sqrt{5 \cdot 4 \left(n + \sqrt{2n} \Phi^{-1} \left(\sqrt[4]{0.99} \right) \right)} \approx 0.045 \sqrt{n + 1.414 \cdot \Phi^{-1}(0.998) \cdot \sqrt{n}} \approx \\ &\approx 0.045 \sqrt{n + 1.414 \cdot 3.03 \cdot \sqrt{n}} = 0.045 \sqrt{n + 4.285 \sqrt{n}}. \end{aligned}$$

В таблице ниже приведены значения погрешности δ при различных значениях размера n квадратного окна со стороной в $l = 10, 20, 30, 40, 50, 60, 70, 80$ пикселей.

$n (= l^2)$	100	400	900	1 600	2 500	3 600	4 900	6 400
δ	0.538	0.992	1.443	1.984	2.344	2.795	3.245	3.695

Заключение. Основным новым результатом данной статьи является аналитическая оценка для вероятности погрешности, которая может возникать при множественном сравнении конечного числа соразмерных цифровых изображений. Эта оценка основана на низкоуровневом сравнении, сводящемся к попиксельному вычислению разностей изображений с помощью евклидовой метрики. При этом делается стандартное предположение о независимом нормальном зашумлении интенсивностей изображения с нулевым математическим ожиданием и априорно установленным среднеквадратическим отклонением в каждом пикселе.

Приведенные в статье доказательства позволяют утверждать, что полученная верхняя оценка для меры различия множества изображений является достаточно «осторожной», т.е. можно ожидать, что в реальности разброс меры, вызванный шумами на изображении, будет значительно меньше теоретически найденной границы.

Приведенные в работе оценки погрешностей могут использоваться как теоретически обоснованные числовые пороговые значения в задачах, требующих принятия решения о совпадении или различии изображений. Такие пороговые значения неизбежно вводятся на различных этапах обработки зашумленных изображений (например, сегментации), и вопрос об их конкретных значениях, как правило, остается открытым, в лучшем случае предлагаются эвристические соображения для их выбора.

Являясь обобщением результатов, полученных в предшествующей работе авторов [15], данная работа также допускает и предполагает обобщения, которые могут быть связаны с нижними оценками вероятности заданного различия для других мер различия (или сходства), основанных на низкоуровневом сравнении изображений. Основным мотивом для этого является то обстоятельство, что в отличие от рассматриваемой в данной работе евклидовой метрики, которая основана только на различиях интенсивностей, известны другие метрики, которые учитывают также и пространственные зависимости пикселей. Можно ожидать, что такие более тонкие меры будут в большей степени соответствовать восприятию сходства и различия множественных изображений человеком.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Li-Jia Li, Socher Richard, Li Fei-Fei.* Towards Total Scene Understanding: Classification, Annotation and Segmentation in an Automatic Framework // *Computer Vision and Pattern Recognition.* – 2009.
2. *Dornaika F., Chakik F.* Efficient Object Detection and Matching using Feature Classification // *2010 International Conference on Pattern Recognition, ICPR 2010.* – 2010. – P. 3073-3076.
3. *Mitchell H.B.* Image Fusion. Theories, Techniques and Applications. – Springer-Verlag, Berlin, Heidelberg, 2010. – 247 p.
4. *Di Gesù Vito, Starovoitov Valery.* Distance-based functions for image comparison // *Pattern Recognition Letters.* Elsevier. – 1999. – No. 20. – P. 207-214.
5. *Дюран Б., Оддел П.* Кластерный анализ. – М.: Статистика, 1977.
6. *Gool L., Moons T., Ungureanu D., Pauwels E.* Symmetry from Shape and Shape from Symmetry // *Int. J. Robotics Res.* – 1995. – 14 (5). – P. 407-424.
7. *Martinet A., Soler C., Holzhuch N., Sillion F.* Accurate Detection of Symmetries in 3D Shapes // *ACM Trans. Graph.* – 2006. – 25 (2). – P. 439-464.
8. *Каркищенко А.Н., Горбань А.С.* Моделирование и классификация поточечных мер сходства // *Тр. международной конференции AIS/CAD'08.* – Таганрог, 2008.
9. *Karkishchenko A.N., Gorban A.S.* Detection of Symmetry of Images Based on Similarity Measures of Sets // *9th International Conference "Pattern Recognition and Image Analysis: New Information Technologies" (PRIA-9-2008): Conf. Proceedings.* – Nizhniy Novgorod, 2008. – Vol. 1. – P. 261-264.
10. *Каркищенко А.Н., Горбань А.С.* К определению мер сходства полутоновых изображений // *Известия ЮФУ. Технические науки.* – 2008. – № 4 (81). – С. 98-103.
11. *Каркищенко А.Н., Мнухин В.Б.* Классификация изображений периодических структур на основе непрерывного преобразования симметрии // *Тр. 8-й Международной конференции «Интеллектуализация обработки информации – 2010».* – Пафос, Кипр, 2010. – С. 359-362.
12. *Karkishchenko A.N., Mnukhin V.B.* Fourfold Symmetry Detection in Digital Images Based on Finite Gaussian Fields // *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (ITI'16).* – Vol. 2. – Vol. 451 of the series *Advances in Intelligent Systems and Computing.* – Springer, 2016. – P. 153-163.
13. *Ameen Mohammed Abd-alsalam Selami, Ahmed Freidoon Fadhil.* A Study of the Effects of Gaussian Noise on Image Features // *Kirkuk University Journal / Scientific Studies (KUJSS).* – September, 2016. – Vol. 11. – Iss. 3. – P. 152-169.
14. *Di Gesù Vito, Starovoitov Valery.* Distance-based functions for image comparison // *Pattern Recognition Letters,* – Elsevier, 1999. – No. 20. – P. 207-214.
15. *Каркищенко А.Н., Мнухин В.Б.* О влиянии зашумления на распознавание симметрии 3-го порядка в гексагональных изображениях // *Известия ЮФУ. Технические науки.* – 2020. – № 5 (215). – С. 171-184.
16. *Karkishchenko A.N., Mnukhin V.B.* Threefold Symmetry Detection in Hexagonal Images Based on Finite Eisenstein Fields // *Analysis of Images, Social Networks, and Texts. 5th International Conference, AIST'2016. Selected Papers. Communications in Computer and Information Science 661.* – Springer, 2017. – P. 281-292.
17. *Hundt R., Schön J.C., Hannemann A., Jansen M.* Determination of Symmetries and Idealized Cell Parameters for Simulated Structures // *Journal of Applied Crystallography.* – 1999. – Vol. 32. – P. 413-416.
18. *Spek A.L.* Structure Validation in Chemical Crystallography // *Acta Crystallographica.* – 2009. – D65. – P. 148-155.
19. *Zeyun Yu, Bajaj C.* Automatic Ultrastructure Segmentation of Reconstructed CryoEM Maps of Icosahedral Viruses // *IEEE Transactions on Image Processing.* – 2005. – 14 (9). – P. 1324-1337.
20. *Seiichi Kondo, Mark Lutwyche, Yasuo Wada.* Observation of Threefold Symmetry Images due to a Point Defect on a Graphite Surface Using Scanning Tunneling Microscope (STM) // *Japanese Journal of Applied Physics.* – 1994. – 33 (9B). – P. 1342-1344.
21. *Маркус М., Минк Х.* Обзор по теории матриц и матричных неравенств. – М.: Наука, 1972. – 232 с.
22. *Крамер Г.* Математические методы статистики. – М.: Мир, 1975. – 648 с.

23. Кибзун А.И., Горяинова Е.Р., Наумов А.В. Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами. – М.: Физматлит, 2013. – 232 с.
24. Вентцель Е.С., Овчаров Л.А. Теория вероятностей. – М.: Наука, 1969. – 368 с.

REFERENCES

1. Li-Jia Li, Socher Richard, Li Fei-Fei. Towards Total Scene Understanding: Classification, Annotation and Segmentation in an Automatic Framework, *Computer Vision and Pattern Recognition*, 2009.
2. Dornaika F., Chakik F. Efficient Object Detection and Matching using Feature Classification, *2010 International Conference on Pattern Recognition, ICPR 2010*, 2010, pp. 3073-3076.
3. Mitchell H.B. Image Fusion. Theories, Techniques and Applications. Springer-Verlag, Berlin, Heidelberg, 2010, 247 p.
4. Di Gesù Vito, Starovoitov Valery. Distance-based functions for image comparison, *Pattern Recognition Letters. Elsevier*, 1999, No. 20, pp. 207-214.
5. Dyuran B., Odell P. Klasternyy analiz [Cluster analysis]. Moscow: Statistika, 1977.
6. Gool L., Moons T., Ungureanu D., Pauwels E. Symmetry from Shape and Shape from Symmetry, *Int. J. Robotics Res.*, 1995, 14 (5), pp. 407-424.
7. Martinet A., Soler C., Holzschuch N., Sillion F. Accurate Detection of Symmetries in 3D Shapes, *ACM Trans. Graph.*, 2006, 25 (2), pp. 439-464.
8. Karkishchenko A.N., Gorban' A.S. Modelirovanie i klassifikatsiya potocheynykh mer skhodstva [Modeling and Classification of Pointwise Similarity Measures], *Tr. mezhdunarodnoy konferentsiya AIS/CAD'08* [Proceedings of the International Conference AIS/CAD'08]. Taganrog, 2008.
9. Karkishchenko A.N., Gorban A.S. Detection of Symmetry of Images Based on Similarity Measures of Sets, *9th International Conference "Pattern Recognition and Image Analysis: New Information Technologies" (PRIA-9-2008): Conf. Proceedings*. Nizhniy Novgorod, 2008, Vol. 1, pp. 261-264.
10. Karkishchenko A.N., Gorban' A.S. K opredeleniyu mer skhodstva polutonovykh izobrazheniy [To the Definition of Measures of Similarity of Halftone Images], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2008, No. 4 (81), pp. 98-103.
11. Karkishchenko A.N., Mnukhin V.B. Klassifikatsiya izobrazheniy periodicheskikh struktur na osnove nepreryvnogo preobrazovaniya simmetrii [Classification of images of periodic structures based on continuous symmetry transform], *Tr. 8-y Mezhdunarodnoy konferentsii «Intelktualizatsiya obrabotki informatsii – 2010»* [Proceedings of the 8th International Conference "Intellectualization of Information Processing - 2010"]. Paphos, Cyprus, 2010, pp. 359-362.
12. Karkishchenko A.N., Mnukhin V.B. Fourfold Symmetry Detection in Digital Images Based on Finite Gaussian Fields, *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry" (ITI'16)*, Vol. 2, Vol. 451 of the series Advances in Intelligent Systems and Computing. Springer, 2016, pp. 153-163.
13. Ameen Mohammed Abd-alsalam Selami, Ahmed Freidoon Fadhil. A Study of the Effects of Gaussian Noise on Image Features, *Kirkuk University Journal / Scientific Studies (KUJSS)*. September, 2016, Vol. 11, Iss. 3, pp. 152-169.
14. Di Gesù Vito, Starovoitov Valery. Distance-based functions for image comparison, *Pattern Recognition Letters. Elsevier*, 1999, No. 20, pp. 207-214.
15. Karkishchenko A.N., Mnukhin V.B. O vliyani zashumleniya na raspoznavanie simmetrii 3-go poryadka v geksagonal'nykh izobrazheniyakh [On the influence of noise on the recognition of the third order symmetry in hexagonal images], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2020, No. 5 (215), pp. 171-184.
16. Karkishchenko A.N., Mnukhin V.B. Threefold Symmetry Detection in Hexagonal Images Based on Finite Eisenstein Fields, *Analysis of Images, Social Networks, and Texts. 5th International Conference, AIST'2016. Selected Papers. Communications in Computer and Information Science 661*. Springer, 2017, pp. 281-292.
17. Hundt R., Schön J.C., Hannemann A., Jansen M. Determination of Symmetries and Idealized Cell Parameters for Simulated Structures, *Journal of Applied Crystallography*, 1999, Vol. 32, pp. 413-416.

18. *Spek A.L.* Structure Validation in Chemical Crystallography, *Acta Crystallographica*, 2009, D65, pp. 148-155.
19. *Zeyun Yu, Bajaj C.* Automatic Ultrastructure Segmentation of Reconstructed CryoEM Maps of Icosahedral Viruses, *IEEE Transactions on Image Processing*, 2005, 14 (9), pp. 1324-1337.
20. *Seiichi Kondo, Mark Lutwyche, Yasuo Wada.* Observation of Threefold Symmetry Images due to a Point Defect on a Graphite Surface Using Scanning Tunneling Microscope (STM), *Japanese Journal of Applied Physics*, 1994, 33 (9B), pp. 1342-1344.
21. *Markus M., Mink Kh.* Obzor po teorii matrits i matrichnykh neravenstv [Overview of the theory of matrices and matrix inequalities]. Moscow: Nauka, 1972, 232 p.
22. *Kramer G.* Matematicheskie metody statistiki [Mathematical Methods of Statistics]. Moscow: Mir, 1975, 648 p.
23. *Kibzun A.I., Goryainova E.R., Naumov A.V.* Teoriya veroyatnostey i matematicheskaya statistika. Bazovyy kurs s primerami i zadachami [Theory of Probability and Mathematical Statistics. Basic course with examples and tasks]. Moscow: Fizmatlit, 2013, 232 p.
24. *Ventsel' E.S., Ovcharov L.A.* Teoriya veroyatnostey [Theory of Probability]. Moscow: Nauka, 1969, 368 p.

Статью рекомендовал к опубликованию д.ф.-м.н. Г.В. Куповых.

Каркищенко Александр Николаевич – Научно-исследовательский институт робототехники и процессов управления ЮФУ; e-mail: karkishalex@gmail.com; 347928 г. Таганрог, Россия; тел.: +78634371694; д.ф.-м.н.; профессор; в.н.с.

Мнухин Валерий Борисович – Южный федеральный университет; e-mail: mnukhin.valeriy@mail.ru; г. Таганрог, Россия; тел.: +78634371606; к.ф.-м.н.; доцент.

Karkishchenko Alexander Nikolaevich – Scientific Research Institute of Robotics and Control Processes of the Southern Federal University; e-mail: karkishalex@gmail.com; Taganrog, Russia; phone: +78634371694, dr. of math. sc.; professor; leading researcher.

Mnukhin Valeriy Borisovich – Southern Federal University; e-mail: mnukhin.valeriy@mail.ru; Taganrog, Russia; phone: +78634371606; cand. of math. sc.; associate professor.

УДК 004.89

DOI 10.18522/2311-3103-2021-2-154-167

Ю.А. Кравченко, А.М. Мансур, Ж.Х. Мохаммад

ВЕКТОРИЗАЦИЯ ТЕКСТА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ*

В задачах интеллектуального анализа текста текстовое представление должно быть не только эффективным, но и интерпретируемым, поскольку это позволяет понять операционную логику, лежащую в основе моделей интеллектуального анализа данных. Традиционные методы векторизации текста, такие как TF-IDF и Bag-of-words, эффективны и имеют интуитивно понятную интерпретируемость, но страдают от «проклятия размерности» и не могут понимать смысл слов. С другой стороны, современные распределенные методы эффективно определяют скрытую семантику, но требуют больших вычислительных ресурсов и времени, а также им не хватает интерпретируемости. В этой статье предлагается новый метод векторизации текстов под названием Bag of weighted Concepts WoWC, который представляет документ в соответствии с содержащейся в нем информацией о концептах. Предлагаемый метод создает концепты посредством кластеризации векторов слов (т.е. встраивания слов), и использует частоты этих кластеров кон-

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-29-22019.

цептов для представления векторов документов. Чтобы обогатить итоговое представление документа, предлагается модифицированная весовая функция для взвешивания концептов на основе статистики, извлеченной из информации вложений слов. Векторы, сгенерированные с помощью предложенного метода, характеризуются интерпретируемостью, низкой размерностью, высокой точностью, а также низкими вычислительными затратами при использовании в задачах классификации и кластеризации. Предлагаемый метод протестирован на пяти различных наборах эталонных данных для кластеризации и классификации текстовых документов и сравнивается с несколькими базовыми методами, включая Bag-of-words, TF-IDF, Averaged GloVe, Bag-of-Concepts и VLAC. Результаты показывают, что BoWC превосходит большинство базовых методов и дает в среднем на 7 % лучшую точность.

Векторизация текста; интеллектуальный анализ данных; классификация; кластеризация; машинное обучение; концепты; семантика.

Yu.A. Kravchenko, A.M. Mansour, J.H. Mohammad

TEXT VECTORIZATION USING DATA MINING METHODS

In the text mining tasks, textual representation should be not only efficient but also interpretable, as this enables an understanding of the operational logic underlying the data mining models. Traditional text vectorization methods such as TF-IDF and bag-of-words are effective and characterized by intuitive interpretability, but suffer from the «curse of dimensionality», and they are unable to capture the meanings of words. On the other hand, modern distributed methods effectively capture the hidden semantics, but they are computationally intensive, time-consuming, and uninterpretable. This article proposes a new text vectorization method called Bag of weighted Concepts BoWC that presents a document according to the concepts' information it contains. The proposed method creates concepts by clustering word vectors (i.e. word embedding) then uses the frequencies of these concept clusters to represent document vectors. To enrich the resulted document representation, a new modified weighting function is proposed for weighting concepts based on statistics extracted from word embedding information. The generated vectors are characterized by interpretability, low dimensionality, high accuracy, and low computational costs when used in data mining tasks. The proposed method has been tested on five different benchmark datasets in two data mining tasks; document clustering and classification, and compared with several baselines, including Bag-of-words, TF-IDF, Averaged GloVe, Bag-of-Concepts, and VLAC. The results indicate that BoWC outperforms most baselines and gives 7 % better accuracy on average.

Text vectorization; data mining; classification; clustering; machine learning; concepts; semantic.

Введение. Текстовое представление данных является наиболее широко используемой формой общения и выражения среди всех различных источников данных. Кроме того, текстовые данные создаются и из других источников данных, таких как аудио и видео, где при построении автоматизированных систем респондентов голос преобразуется в текстовые данные, а также добавляются видеообъяснения для задач поиска информации, рекомендаций и оценки. Эти данные растут с каждым днем, и потребность в поиске инновационных механизмов и методов обработки, хранения, понимания и извлечения из них скрытой информации возрастает.

Чтобы применить различные методы машинного обучения и интеллектуального анализа данных, необработанные документы необходимо преобразовать в формат понятный машине [1]. Первым шагом к тому, чтобы сделать текстовые документы машиночитаемыми, является векторизация, которая определяется как преобразование текстового документа в цифровой вектор, и представляет собой процесс извлечения признаков из текста для выполнения любых задач интеллектуального анализа текста и математического решения проблем [2].

Методы векторизации – это методы построения векторных представлений текстов на естественном языке [1, 3]. Они применяются во многих приложениях *Data mining*, обработки естественного языка (ОЕЯ) и поиска информации (ПИ), в таких задачах, как оценка семантического сходства и семантической близости для сопоставления текстов, классификация и кластеризация текстовых документов, обнаружение тем, генерация вопросов, ответы на вопросы, моделирование языков, машинный перевод, обобщение текста и др. Как правило, представление документа направлено на его преобразование в вектор фиксированной длины, который может описывать содержимое, чтобы уменьшить сложность документов и упростить их обработку [4].

Традиционные методы представления документов, такие как «мешок слов» (BoW, Bag-of-Words) и TF-IDF (term frequency invers document frequency), достигли многообещающих результатов во многих задачах классификации и кластеризации документов благодаря своей простоте, эффективности и точности [4–7]. Однако у этих методов есть две основные проблемы. Первая – проблема размерности, поскольку количество функций в результирующих векторах значительно увеличивается по мере увеличения размера корпуса (количества документов). Следовательно, размерность векторов документов может стать чрезвычайно большой и разреженной, и обычные метрики расстояния, такие как евклидово расстояние или косинусное расстояние, станут бессмысленными [8].

Вторая проблема заключается в том, что полученное представление не учитывает семантическое отношение между словами и игнорирует порядок слов, это означает, что разные документы будут иметь одинаковое представление, если используются одни и те же слова. Эти недостатки ограничивают способность моделей интеллектуального анализа текста фиксировать истинное сходство между документами, представленными этими методами.

Метод набора концептов VOC (Bag-of-Concepts) [8] был предложен как решение этой проблемы. Основываясь на корпусе коллекции документов, Bag-of-Concepts кодирует слова как встраиваемые векторы, а затем применяет кластеризацию для группировки похожих слов в кластеры, называемые концептами. Как и в случае метода «мешка слов», каждый вектор документа будет представлен частотами этих концептов в документе. Чтобы уменьшить влияние концептов, которые появляются в большинстве документов, применяется схема взвешивания, подобная TF-IDF, с заменой частоты термина TF на частоту концепта CF. Однако применение такой схемы взвешивания к концептам неэффективно, и снижает выразительную способность результирующих векторов. Это связано с тем, что концепт здесь представлен группой терминов, схожих по смыслу, и поэтому частота появления того или иного концепта на уровне корпуса будет намного больше, чем частота появления одного из его терминов. Математически соотношение между количеством документов, в которых присутствует концепт, к общему количеству документов будет незначительным для большинства концептов, что будет указывать на то, что все концепты являются распространенными на уровне корпуса и не содержат дискриминационной информации, а это ложный вывод.

Следуя тому же подходу, что и VOC, в работе [9] разработан метод «Векторы локально агрегированных концептов» (VLAC, Vectors of Locally Aggregated Concepts), который группирует вложения слов для генерации признаков. Однако вместо того, чтобы подсчитывать частоту кластеризованных вложения слов, VLAC берет сумму остатков каждого кластера относительно его центроида и объединяет их для создания вектора признаков. Результирующие векторы признаков содержат более ценную информацию, чем пакет концептов, благодаря дополнительному включению этих статистических данных первого порядка. Однако этот метод соз-

дает векторы относительно больших размеров с высокими вычислительными затратами. Если бы 10 концептов нужно было создать из вложений слов размером 300, то результирующий вектор документа содержал бы 10×300 значений.

Чтобы преодолеть вышеупомянутые недостатки, в данной работе предлагается новый метод векторизации, названный «Пакетом взвешенных концептов» (BoCW bag-of-weighted concepts), который применяет тот же подход, что и BOC, за исключением того, что он имеет несколько основных отличий:

1. Функция взвешивания, основанная на обратной частоте документа, была заменена функцией монотонного убывания, которая позволит лучше регулировать влияние частоты документа на уровне коллекции документов;
2. Представлена эвристическая функция для извлечения дополнительной дискриминирующей информации, используемая для взвешивания концептов. Эта функция вычисляет важность концепта в документе, измеряя сходство слов, образующих концепт со словами документа, которые ему принадлежат;
3. Для повышения эффективности предложенного метода при выполнении задачи классификации и кластеризации, в которых известны правильные классы документов, введен новый параметр для взвешивания вектора концептов по частоте класса документа.

Таким образом, предлагаемый метод сочетает семантику на уровне терминов с семантикой на уровне концептов. Полученные в результате векторы признаков документов содержат более ценную информацию, чем при использовании BOC и VLAC, и в то же время, эти векторы интерпретируемые и их размеры намного меньше.

1. Аналитический обзор методов векторизации текстов. Метод набора слов «Bag-of-Words» основан на предположении, что частота слов в документе может надлежащим образом отражать сходства и различия между документами. Следовательно, признаки векторов документов, сгенерированные методом «мешка слов», представляют вхождения каждого слова в документе. Этот метод имеет серьезный недостаток: часто встречающиеся слова могут доминировать в пространстве признаков, тогда как редкие слова могут нести больше ценной информации.

Для улучшения представления «мешка слов» предлагается метод TF-IDF, отражающий значимость слова для документа в корпусе, то есть механизм взвешивания, где *Term Frequency* (TF) – это количество раз, когда термин появляется в документе, а обратная частота документа *IDF* измеряет редкость термина во всем корпусе. Обозначая общее количество документов в коллекции $|D|$, понятия частоты термина и обратной частоты документа объединяются, чтобы получить общий вес для каждого термина в каждом документе следующим образом:

$$tf - idf(t, d, D) = tf(t, d) \times idf(t, D) = \frac{n_t}{\sum_k n_k} \times \log \frac{|D|}{|\{D_i \in D | t \in D_i\}|} \quad (1)$$

где n_t – количество вхождений слова t в документе d , а n_k – общее количество слов в этом документе. D_i – количество документов из коллекции D , в которых встречается t .

Метод набор концептов (BOC, Bag-of-concepts), описанный в работе [8], генерирует векторы слов документа с помощью совокупности моделей *word2vec*, которые встраивают семантически похожие слова в соседнюю область. Это позволяет сгруппировать соседние слова в один общий кластер концептов. *Bag-of-Concepts* генерирует кластеры слов, применяя сферические k -средства к вложениям слов. Полученные кластеры содержат слова с похожим значением и поэтому называются концептами. Подобно принципу метода «мешок слов», каждый вектор документа представляется частотой каждого кластера концептов в документе. Чтобы смягчить влияние концептов, которые появляются в большинстве докумен-

тов, используется схема взвешивания, аналогичная TF_IDF , с заменой термина частота TF на частоту концепта CF . Следовательно, он называется $CF-IDF$ (частота концепта с обратной частотой документа) и рассчитывается на основе следующего уравнения.

$$CF-IDF(c_i, d_j, D) = \frac{n_c}{\sum_k n_k} \times \log \frac{|D|}{|\{d \in D | c_i \in d\}|} \quad (2)$$

где $|D|$ – число документов в коллекции, а в знаменателе это число документов из коллекции D , в которых встречается концепт c ; n_c – число вхождений концепта c в документ d , а n_k – общее число концептов в данном документе.

Этот метод решает проблему больших размеров в методе набора слов, поскольку он нелинейно уменьшает размеры при преобразовании пространства слов в пространство концептов на основе семантического сходства. Кроме того, в [8] показано, что этот метод обеспечивает лучшее представление документа, чем *Bag-of-Words* и $TF-IDF$, в задаче классификации для поиска двух наиболее похожих документов среди троек документов. Однако в задаче прогнозирования правильной метки для каждого документа *BOC* не удалось превзойти $TF-IDF$ на двух из трех наборов данных.

2. Постановка задачи. Пусть дан D текстовый словарь, т.е. список уникальных слов, которые появляются в коллекции текстовых документов. Пусть $x_i \in R^D$ – вектор вложения i -го слова словаря D , где D – размерность вложения слова. Множество всех векторов вложения слов обозначается $\mathcal{E} = \{x_i, i = 1, \dots, |D|\}$.

Кроме того, пусть N – количество текстовых документов, которые должны быть закодированы с использованием предложенного метода. Каждый документ описывается векторами вложения своих слов $N_i: x_{ij} \in \mathcal{E}, (i = 1, \dots, N, j = 1, \dots, N_i)$, где N_i – количество слов i -го документа. Каждое слово в документе имеет вектор вложения, то есть x_{ij} – это вектор вложения j -го слова из i -го документа. Количество слов варьируется от одного документа к другому.

Таким образом, ставится задача найти преобразование $y = f(x): R^D \rightarrow R^C$, такое, что преобразованный вектор признаков $y_i \in R^C$ сохраняет (большую часть) преобразование или структуру в R^D . Оптимальное преобразование $y = f(x)$ будет таким, которое не приводит к увеличению минимальных ошибок вероятности.

Соответственно, требуется найти преобразование из пространства слов в пространство концептов, которое позволяет каждому документу быть представленным вектором фиксированной длины. $N_i: c_{ij} = (i = 1, \dots, K, j = 1, \dots, N_i)$, где k – количество извлеченных концептов, а c_{ij} – это признак j -го концепта i -го документа.

3. Разработка метода «Bag of weighted concepts». Для реализации поставленной задачи разрабатывается принципиально новый метод для представления текстовых документов в виде числовых векторов фиксированной длины. Предлагаемый метод преобразует документ в вектор в соответствии с содержащейся в нем информацией о концептах. Для этого, аналогично методу «пакета концептов», создаётся словарь концептов T , а затем векторы документа создаются на основе статистики частотности концептов в документе.

Все векторы вложения кластеризуются в N_k кластеров. Используется сферический алгоритм k -средних, в котором косинусное подобие применяется в качестве метрики расстояния. Для заранее определенного значения K алгоритм итеративно назначает каждую точку данных одному из k центроидов и обновляет каждый центроид с учетом принадлежности точек данных. Слово может быть перегруппировано более чем в один кластер, это фактически означает, что различные значения слова могут быть объединены с их синонимами.

Необходимо обеспечить минимальное количество слов в каждом кластере, представляющем концепт, для этого проверяется размер результирующего кластера и расширяется его. После этого слова располагаются в каждом кластере в соответствии с близостью к центру кластера, а затем выбираются M слова, наиболее близкие к центру кластера. В итоге получается группа концептов (понятий), каждое из которых представлено M словами принадлежащие к одному общему понятию или имеющие общий гипероним. Словарь концептов представляется следующим образом $T = (w_1^1, w_2^1, \dots, w_M^1, w_1^2, w_2^2, \dots, w_M^2, \dots, w_1^K, w_2^K, \dots, w_M^K)$, где w_i^j – i -е слово j -го кластера.

Алгоритм построения словаря концептов

Ввод	Набор $D = \{d_1, d_2, \dots, d_N\}$ из N документов $M_{desired}$ // минимальное количество слов кластера
Вывод	Словарь концептов $T = []$
1:	Отсканировать документы и создать словарь D
2:	Инициализировать вложение слова \mathcal{E} с помощью D
3:	Инициализировать словарь T , запустив k -means на \mathcal{E}
4:	$C = count(T)$ // количество концептов
5:	While $i < C$ do
6:	$M_{real} = count(clusters[i])$
7:	$L = M_{desired} - M_{real}$ // Количество слов, которыми следует расширить кластер
8:	if $(L > 0)$ then
9:	for $j = 0$ to L do
10:	$w := similarTo(T[i, j])$ // получить синоним слова j
11:	$w^* = embedding(w)$ // встраивание слова
12:	$T[i] \leftarrow w^*$ // Добавить этот слово в текущий кластер
	end
	else
13:	$sort(clusters[i])$
14:	$T[] \leftarrow clusters[M_{desired}]$
	end
15:	end Сохранить кластеры в виде словаря T .

Для рассматриваемого документа d создается вектор признаков размера k , равного количеству концептов $V^d = (c_1^d, c_2^d, \dots, c_k^d)$, где признак c_i^d выражает степень значимости i -го концепта (его вес) в документе. Значимость концепта в векторе документа рассчитывается аналогично методу *ВОС*, где слова документа сравниваются с концептами, т.е. измеряется степень косинусного сходства между вектором слова документа и вектором центра кластера и записывается появление концептов, превышающих определенный порог θ , определяемый экспериментально. Частота концепта задается следующей формулой:

$$CF(c_i, d_j, D) = \frac{n_c}{\sum_k n_k}, \quad (3)$$

где n_k – общее количество концептов в документе, а n_c – количество вхождений концепта c в документ, которое вычисляется следующей двоичной функцией $g(s)$:

$$g(s) = \begin{cases} 1, & s > \theta \\ 0, & otherwise \end{cases} \quad (4)$$

где сходство s между двумя векторами слов X и Y вычисляется следующим образом:

$$s = \text{similarity}(X, Y) = \cos(\theta) = \frac{x \cdot y}{\|x\| \cdot \|y\|} = \frac{\sum_{i=1}^n X_i Y_i}{\sqrt{\sum_{i=1}^n X_i^2} \sqrt{\sum_{i=1}^n Y_i^2}}, \quad (5)$$

где X_i и Y_i – компоненты вектора слов X и Y , соответственно.

Чтобы уменьшить влияние концептов, которые появляются в большинстве документов, в методе *BOC* применяется схема взвешивания, подобная *TF-IDF*, в которой частота концепта *CF* взвешивается по обратной частоте документа *IDF* (2). Однако использование этой формулы на уровне концепта неэффективно, потому что частота появления концепта на уровне коллекции документов намного выше, чем частота одного из его терминов. Математически это приводит к очень малым значениям логарифма – близкими к нулю, что указывает ложный вывод.

Для преодоления этой проблемы авторы предлагают новую функцию взвешивания, основанную на обратной частоте документа ранее установленной функцией монотонного убывания [10], которая позволяет повысить качество регулировки влияния частоты документа на уровне коллекции документов. Функция задается следующей формулой:

$$f(F) = e^{-\alpha \times F}, \quad (6)$$

где α – константа, а F – частота документа, вычисляемая по формуле:

$$F = \left(\frac{| \{d \in D | c_i \in d\} |}{|D|} \right). \quad (7)$$

Экспоненциальная функция была выбрана для гарантии, что значение f находится в диапазоне $[0,1]$. Тогда весовая функция, которая получила название *CF-EDF* (concept frequency – exponential document frequency), принимает следующий вид:

$$CF - EDF(c_i, d_j, D) = \frac{n_c}{\sum_k n_k} \times \exp \left(- \frac{| \{d \in D | c_i \in d\} |}{|D|} \right). \quad (8)$$

Чтобы полученный вектор документа содержал более значимую (ценную) информацию, авторы предлагают эвристическую функцию, которая извлекает статистику, характеризующую отношение каждого концепта c документом, путем вычисления сходства слов, образующих концепт, со словами документа, которые ему принадлежат. Для документа d с N словами, принадлежащими c -му концепту, сходство вычисляется следующим образом:

$$S_c = \frac{1}{N} \sum_{i=1}^N \max_{1 \leq j \leq M} \text{sim}(u_i, v_j), \quad (9)$$

где u_i – i -е слово документа, принадлежащее концепту; v_j – j -е слово концепта. Функция *Max* возвращает наивысшую оценку сходства, записанную для каждого слова документа, принадлежащего концепту c .

Итоговая формула взвешивания концепта выглядит следующим образом:

$$CF - EDF(c_i, d_j, D) = \frac{n_c}{\sum_k n_k} \times \exp \left(- \frac{| \{d \in D | c_i \in d\} |}{|D|} \right) \times \exp(S_{c_i}). \quad (10)$$

Для повышения эффективности предложенного метода при выполнении задачи классификации и кластеризации, в которых известны правильные классы документов, введен новый параметр для представления внутриклассовых характеристик, который называется частотой класса. Аналогично принципу весовой функции *TF-IDF-CF*, предложенному в [11] в данной работе вычисляется частота концептов в документах в пределах одного класса. Итоговая формула взвешивания концепта выглядит следующим образом:

$$f(c_i, d_j, D) = CF - EDF(c_i, d_j, D) \times \frac{\mu_{kij}}{N_{kj}}, \quad (11)$$

где μ_{kij} – количество документов, в которых концепт c_i появляется в том же классе k , которому принадлежит j -ый документ; N_{kj} – количество документов в том же классе k , к которому принадлежит j -ый документ.

4. Вычислительный эксперимент и анализ полученных результатов. Эффективность метода *BoWC* оценивалась по нескольким базовым критериям посредством их оценки при выполнении задачи кластеризации. Приведем описание использованных наборов данных и метрик оценки.

Для оценки предложенного метода использовались пять наборов данных представленных в табл. 1.

Таблица 1

Наборы использованных данных

Набор данных	<i>BBC</i>	<i>R8</i>	<i>OH</i>	<i>20Newsgroup</i>	<i>WebKB</i>
Количество документов	2225	8491	5380	18821	4199
Количество категорий	5	8	7	20	4
Среднее количество слов в документе	2262	742	1008	1902	909
Среднее количество словарных слов в одном документе	207	66	79	135	98

BBC содержит 2225 документов, принадлежащих к 5 различным классам. *Reuters (R8)* содержит статьи из новостной ленты *Reuters*. В этой работе используется разделение *R8* набора данных *Reuters*, которое содержит 8491 документ, где документы принадлежат 8 различным классам. *20Newsgroups*, содержит 18821 документ, где документы принадлежат 20 различным категориям групп новостей. *OHSUMED (OH)* содержит 5380 документов, принадлежащих 7 различным классам. *WebKB* содержит веб-страницы из различных разделов информатики, которые были разделены на 7 разных классов: студенты, преподаватели, сотрудники и т.д. В этой работе использовался предварительно обработанный набор данных *WebKB*, который содержит 4 различных класса и всего 4199 документов [12, 13].

Чем больше объём корпуса, на котором обучается модель вложений, тем выше репрезентативная способность полученных векторов слов. Следовательно, обучение модели вложений на данных разного размера было бы неправильным вариантом, поэтому используется предварительно обученная модель, а именно *GloVe* из-за ее преимуществ перед другими [14]. Модель *GloVe* была обучена на наборе данных *Common Crawl* и содержит векторы для 1,9 миллиона английских слов.

Такая же предварительная обработка была применена ко всем наборам данных путем перевода текста в нижний регистр и разбивки на самые длинные непустые последовательности буквенно-цифровых символов, которые содержат не менее трехбуквенных символов. Стоп-слова, редкие слова и слова без вложений были удалены. Поскольку используется предварительно обученное встраивание (*embeddings*), выделение корней слов (*stemming*) не выполнялось, чтобы уменьшить количество слов вне словарного запаса, за исключением набора данных *WebKB*, где исходные данные получить не удалось, поэтому использовалась предварительно обработанная версия.

В задаче кластеризации, зная присвоения эталонных данных классам образцов, определена интуитивно понятная метрика с помощью анализа условной энтропии, одной из которых является V -мера.

V -мера – мера, основанная на энтропии [15], которая явно измеряет, насколько успешно были удовлетворены критерии однородности (*homogeneity*), когда каждый кластер содержит только членов одного класса, и полноты (*completeness*), когда все члены данного класса относятся к одному кластеру. V -мера вычисляется как среднее гармоническое для различных оценок однородности и полноты, точно так же, как точность и полнота обычно объединяются в F -меру [16].

$$V - \text{мера} = \frac{(1+\beta) \times \text{Однородность} \times \text{Полнота}}{(\beta \times \text{Однородность} + \text{Полнота})} \quad (12)$$

Фактор β может быть откорректирован, чтобы обеспечить либо однородность, либо полноту алгоритма кластеризации.

В проведенном вычислительном эксперименте предложенный метод *BoWC* оценивался по результатам решения задачи кластеризации путем сравнения с пятью проанализированными ранее методами, а именно, *Bag-of-Words*, *TF-IDF*, *GloVe*, *Bag-of-Concepts* и *VLAC*. Результаты эксперимента показаны в табл. 2.

Таблица 2

Результаты кластеризации по V -мере

	<i>BBC</i>	<i>R8</i>	<i>OHSUMED</i>	<i>20NG</i>	<i>WebKB</i>
<i>TF-IDF</i>	0,663	0,513	0,122	0,362	0,313
<i>Количество признаков</i>	29821	29290	40179	173446	7632
<i>Bag of Words</i>	0,209	0,248	0,027	0,021	0,021
<i>Количество признаков</i>	17350	14446	17481	92718	7637
<i>Averaged GloVe</i>	0,774	0,481	0,109	0,381	0,219
<i>Количество признаков</i>	300	300	300	300	300
<i>Bag-of-Concepts</i>	0,638	0,131	0,1	0,394	0,084
<i>Количество признаков</i>	200	200	200	200	200
<i>VLAC</i>	0,808	0,456	0,118	–	0,286
<i>Количество признаков</i>	9000	9000	9000	–	9000
<i>BoWC</i>	0,889	0,534	0,151	0,511	0,158
<i>Количество признаков</i>	100	55	75	65	70

Эффективность метода *BoWC* была проанализирована с постоянно увеличивающимся количеством концептов от 5 до 100 концептов, чтобы обеспечить справедливое сравнение со всеми методами. Признаки усредненных вложений слов *GloVe*, *Bag-of-Words* и *TF-IDF* максимизированы по умолчанию. Результаты сравнения предложенного метода векторизации текста *BoWC* с рассмотренными каноническими при решении задачи кластеризации документов на наборах данных (*BBC*, *Reuters*, *20Newsgroups*, *OHSUMED*, *WebKB*) с использованием V -меры представлены на рис. 1.

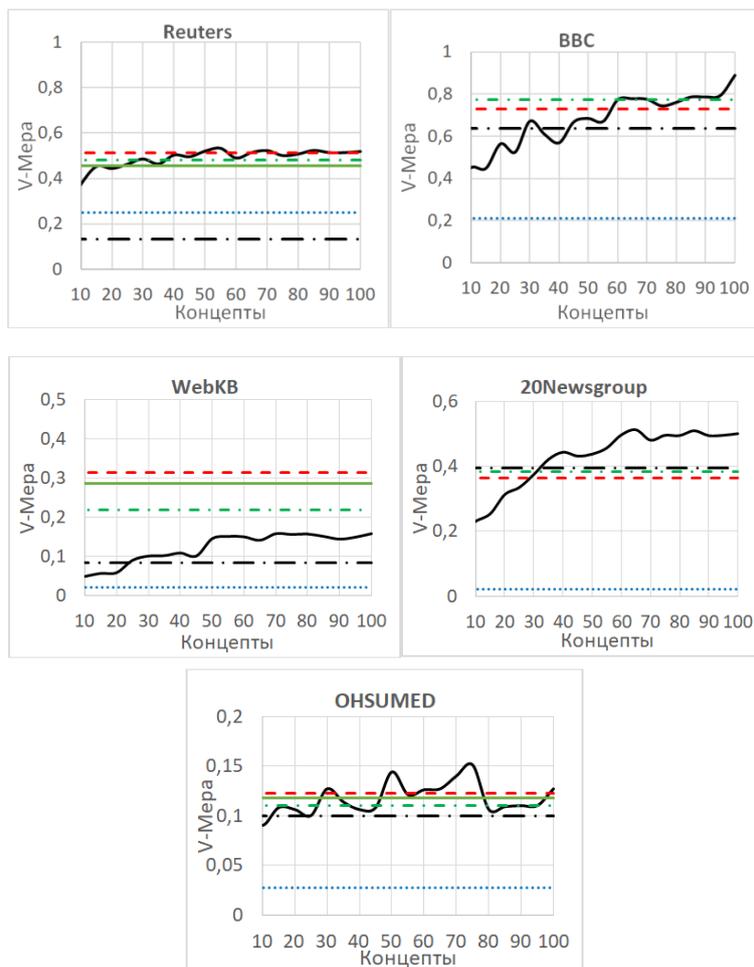


Рис. 1. Результаты сравнения предложенного метода векторизации текста с другими методами при выполнении задачи кластеризации документов на наборах данных (BBC, Reuters, 20Newsgroups, OHSUMED, WebKB) с применением V-меры

Для метода VLAC, прилагаемый к работе API был использован при повторной реализации экспериментов на наборах данных из-за разницы в источнике данных и критериях оценки. Было принято 30 концептов (9000 признаков), учитывая, что большее количество концептов требует больших вычислительных затрат [9]. Существует значительная разница в количестве признаков между *BoWC* и методом VLAC, однако они будут сравниваться с точки зрения информативности первых тридцати концептов, а не только с точки зрения количества признаков.

Учитывая, что *BOC* является частным случаем предложенного метода *BoWC*, он был реализован через код метода *BoWC*. В [8] показано, что *Bag-of-Concepts* по сравнению с *TF-IDF* требуется не менее 100 концептов для достижения конкурентоспособных результатов. Итак, чтобы максимизировать точность *Bag-of-Concepts*, количество концептов было установлено равным 200.

Чтобы сделать сравнение между реализациями *BoWC* возможным, все вложения слов имели размер 300. Метод *K-средних* использовался в этом эксперименте в качестве классификатора методов генерации признаков.

Табл. 1 показывает, что предложенный метод, превзошел большинство других методов, с уменьшением количества признаков в среднем на 50%, сохранив высокую точность классификации. Значения *V-меры* на наборе данных *OHSUMED* низкие для всех методов. Вероятно, это связано с природой этих данных, поскольку они содержат большое число медицинских терминов, связанных с точки зрения темы, и, следовательно, на уровне термина требуется значительное количество признаков для различения каждого документа, а на уровне концептов требуются дополнительные ресурсы для построения соответствующего концептуального словаря. Тем не менее, предложенный метод дал наилучшее значение для *V-меры* с 75 концептами.

Для набора данных *WebKB* методы *TF-IDF* и *VLAC* превзошли *BoWC*. Этот результат логичен, учитывая, что взята предварительно обработанная версия этого набора данных, что увеличивает количество слов вне словарного запаса и влияет на качество и количество результирующих концептов [17–21].

Эксперимент *VLAC* с набором данных *20Newsgroup* был исключен из-за его высоких вычислительных затрат, так как алгоритм кластеризации завершился неудачно из-за полной загрузки 12,72 ГБ ОЗУ. Причина этого в том, что этот набор данных огромен, а векторы документов, сгенерированные *VLAC*, длинные и, следовательно, требуют большого объема памяти для хранения и обработки.

Колебания кривых на рис. 1 обусловлены разным качеством концептов, извлекаемых при каждом запуске метода, и это весьма логично, поскольку алгоритм построения словаря концептов при каждом запуске дает разные концепты, это связано с типом используемой функции расстояния и порога схожести, по которому определяется появление того или иного концепта в документе. Соответственно, сложно определить конкретное количество признаков, обеспечивающих стабильную производительность метода, что является одним из недостатков предложенного метода, требующим дополнительных исследований.

Заключение. В данной работе представлена разработка метода векторизации текстов *BoWC*, который представляет документ в соответствии с содержащейся в нем информацией о концептах. Предлагаемый метод создает концепты посредством кластеризации векторов слов (т.е. встраивания слов), и использует частоты этих кластеров концептов для представления векторов документов. Чтобы обогатить итоговое представление документа, предлагается новая модифицированная весовая функция для взвешивания концептов на основе статистики, извлеченной из информации вложений слов.

Векторы, сгенерированные с помощью предложенного метода, характеризуются интерпретируемостью, низкой размерностью, высокой точностью, а также низкими вычислительными затратами при использовании в задачах кластеризации.

В целях контроля качества концептов, генерируемых разработанным методом, планируется провести дополнительные исследования для внесения улучшений в алгоритм формирования словаря концептов с целью создания более качественных концептов для преодоления неопределенности в результатах, которая проявлялась в колебаниях значений *V-меры*, отражающей точность процесса кластеризации. Кроме того, необходимо протестировать возможность применения процессов обрезки или слияния концептов, чтобы контролировать качество полученных концептов. С другой стороны, эксперименты будут расширены, чтобы включить анализ влияния других типов вложений на эффективность предложенного метода, включая создание вложений на основе самих данных.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Бенгфорт Б., Билбро Р., Охеда Т. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. – СПб.: Питер, 2019. – 368 с.
2. Лапшин С.В., Лебедев И.С., Сивак А.И. Классификация коротких сообщений с использованием векторизации на основе elmo // Известия ТулГУ. Технические науки. – 2019. – № 10. – С. 410-418.
3. Киреев В.С., Федоренко В.И. Использование методов векторизации текстов на естественном языке для повышения качества контентных рекомендаций фильмов // Современные наукоемкие технологии. – 2018. – № 3. – С. 102-106.
4. Lin Y., Liu Z., Sun M. Representation Learning for Natural Language Processing. – Singapore: Springer Nature, 2020. – 334 p.
5. Baeza-Yates R., Ribeiro-Neto B. Modern Information Retrieval. – New York: ACM Press, 1999. – 501 p.
6. Jones K.S. A Statistical Interpretation of Term Specificity and its Application in Retrieval // Journal of Documentation. – 1972. – Vol. 28, No. 1. – P. 11-21.
7. Hoi S., Wu L., Yu N. Semantics-Preserving Bag-of-Words Models and Applications // IEEE Transactions on Image Processing. – 2010. – Vol. 19, No. 7. – P. 1908-1920.
8. Kim, H.K., Kim H.-j. Bag-of-Concepts: Comprehending Document Representation through Clustering Words in Distributed Representation // Neurocomputing. – 2017. – Vol. 266. – P. 336-352.
9. Grootendorst M., Vanschoren J. Beyond Bag-of-Concepts: Vectors of Locally Aggregated Concepts // Joint European Conference on Machine Learning and Knowledge Discovery in Databases. – 2019. – P. 681-696.
10. Bandar Z., Crockett K., Li Y. et al. Sentence Similarity Based on Semantic Nets and Corpus Statistics // IEEE Transactions on Knowledge. – 2006. – Vol. 18. – P. 1138-1150.
11. Liu M., Yang J. An Improvement of TFIDF Weighting in Text Categorization // International Proceedings of Computer Science Information Technology. – 2012. – Vol. 47. – P. 44-47.
12. Cardoso-Cachopo, A.L., Oliveira A. Semi-Supervised Single-Label Text Categorization Using Centroid-Based Classifiers // Proceedings of the 2007 ACM Symposium on Applied Computing. – 2007. – P. 844-851.
13. Lang, K., Rennie J. The 20 Newsgroups Data Set. – 2008.
14. Manning C.D., Pennington J., Socher R. Glove: Global Vectors for Word Representation // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). – 2014. – P. 1532-1543.
15. Hirschberg J., Rosenberg A. V-measure: A Conditional Entropy-Based External Cluster Evaluation Measure // Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). – 2007. – P. 410-420.
16. Van Rijsbergen C.J. Information Retrieval. – Butterworth-Heinemann, 1979. – 224 p.
17. Bova V., Zaporozhets D., Kureichik V. Integration and Processing of Problem-Oriented Knowledge Based on Evolutionary Procedures // Advances in Intelligent Systems and Computing. – 2016. – Vol. 450. – P. 239-249.
18. Kureichik V.M., Semenova A.V. Ensemble of Classifiers for Ontology Enrichment // Journal of Physics: Conference Series. – 2018. – Vol. 1015. – Issue 3. – Article id. 032123.
19. Bova V.V., Nuzhnov E.V., Kureichik V.V. The Combined Method of Semantic Similarity Estimation of Problem Oriented Knowledge on the Basis of Evolutionary Procedures // Advances in Intelligent Systems and Computing. – 2017. – Vol. 573. – P. 74-83.
20. Pulyavina N., Taratukhin V. The Future of Project-Based Learning for Engineering and Management Students: Towards an Advanced Design Thinking Approach // ASEE Annual Conference and Exposition, Conference Proceedings. – 2018. – No. 125.
21. Becker J., Pulyavina N., Taratukhin V. Next-Gen Design Thinking. Using Project-Based and Game-Oriented Approaches to Support Creativity and Innovation // Proceedings of the 1st International Conference of Information Systems and Design. – 2020.

REFERENCES

1. Bengfort B., Bilbro R., Okheda T. Prikladnoy analiz tekstovyykh dannykh na Python. Mashinnoe obuchenie i sozдание prilozheniy obrabotki estestvennogo yazyka [Applied analysis of text data in Python. Machine learning and building natural language processing applications]. Saint Petersburg: Piter, 2019, 368 p.
2. Lapshin S.V., Lebedev I.S., Spivak A.I. Klassifikatsiya korotkikh soobshcheniy s ispol'zovaniem vektorizatsii na osnove elmo [Classification of short messages using elmo-based vectorization], *Izvestiya TulGU. Tekhnicheskie nauki* [News of TulSU. Technical sciences], 2019, No. 10, pp. 410-418.
3. Kireev V.S., Fedorenko V.I. Ispol'zovanie metodov vektorizatsii tekstov na estestvennom yazyke dlya povysheniya kachestva kontentnykh rekomendatsiy fil'mov [Using methods of vectorization of texts in natural language to improve the quality of content recommendations of films], *Sovremennye naukoemkie tekhnologii* [Modern science-intensive technologies], 2018, No. 3, pp. 102-106.
4. Lin Y., Liu Z., Sun M. Representation Learning for Natural Language Processing. Singapore: Springer Nature, 2020, 334 p.
5. Baeza-Yates R., Ribeiro-Neto B. Modern Information Retrieval. New York: ACM Press, 1999, 501 p.
6. Jones K.S. A Statistical Interpretation of Term Specificity and its Application in Retrieval, *Journal of Documentation*, 1972, Vol. 28, No. 1, pp. 11-21.
7. Hoi S., Wu L., Yu N. Semantics-Preserving Bag-of-Words Models and Applications, *IEEE Transactions on Image Processing*, 2010, Vol. 19, No. 7, pp. 1908-1920.
8. Kim H.K., Kim H.-j. Bag-of-Concepts: Comprehending Document Representation through Clustering Words in Distributed Representation, *Neurocomputing*, 2017, Vol. 266, pp. 336-352.
9. Grootendorst M., Vanschoren J. Beyond Bag-of-Concepts: Vectors of Locally Aggregated Concepts, *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2019, pp. 681-696.
10. Bandar Z., Crockett K., Li Y. et al. Sentence Similarity Based on Semantic Nets and Corpus Statistics, *IEEE Transactions on Knowledge*, 2006, Vol. 18, pp. 1138-1150.
11. Liu M., Yang J. An Improvement of TFIDF Weighting in Text Categorization, *International Proceedings of Computer Science Information Technology*, 2012, Vol. 47, pp. 44-47.
12. Cardoso-Cachopo, A.L., Oliveira A. Semi-Supervised Single-Label Text Categorization Using Centroid-Based Classifiers, *Proceedings of the 2007 ACM Symposium on Applied Computing*, 2007, pp. 844-851.
13. Lang, K., Rennie J. The 20 Newsgroups Data Set., 2008.
14. Manning C.D., Pennington J., Socher R. Glove: Global Vectors for Word Representation, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532-1543.
15. Hirschberg J., Rosenberg A. V-measure: A Conditional Entropy-Based External Cluster Evaluation Measure, *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 2007, pp. 410-420.
16. Van Rijsbergen C.J. Information Retrieval. Butterworth-Heinemann, 1979, 224 p.
17. Bova V., Zaporozhets D., Kureichik V. Integration and Processing of Problem-Oriented Knowledge Based on Evolutionary Procedures, *Advances in Intelligent Systems and Computing*, 2016, Vol. 450, pp. 239-249.
18. Kureichik V.M., Semenova A.V. Ensemble of Classifiers for Ontology Enrichment, *Journal of Physics: Conference Series*, 2018, Vol. 1015, Issue 3, Article id. 032123.
19. Bova V.V., Nuzhnov E.V., Kureichik V.V. The Combined Method of Semantic Similarity Estimation of Problem Oriented Knowledge on the Basis of Evolutionary Procedures, *Advances in Intelligent Systems and Computing*, 2017, Vol. 573, pp. 74-83.
20. Pulyavina N., Taratukhin V. The Future of Project-Based Learning for Engineering and Management Students: Towards an Advanced Design Thinking Approach, *ASEE Annual Conference and Exposition, Conference Proceedings*, 2018, No. 125.
21. Becker J., Pulyavina N., Taratukhin V. Next-Gen Design Thinking. Using Project-Based and Game-Oriented Approaches to Support Creativity and Innovation, *Proceedings of the 1st International Conference of Information Systems and Design*, 2020.

Статью рекомендовал к опубликованию к.т.н., доцент С.Г. Буланов.

Мансур Али Махмуд – Южный федеральный университет; e-mail: mansur@sfedu.com; г. Таганрог, Россия; тел.: 88634371651; кафедра систем автоматизированного проектирования; аспирант.

Мохаммад Жуман Хуссейн – e-mail: zmohammad@sfedu.ru; кафедра систем автоматизированного проектирования; аспирант.

Кравченко Юрий Алексеевич – e-mail: yakravchenko@sfedu.ru; тел.: +79289080151; кафедра систем автоматизированного проектирования, доцент.

Mansour Ali Mahmoud – Southern Federal University; e-mail: mansur@sfedu.com; Taganrog, Russia; phone: +78634371651; the department of computer aided design; graduate student.

Mohammad Juman Hussain – e-mail: zmohammad@sfedu.ru; the department of computer aided design, graduate student.

Kravchenko Yury Alekseevich – e-mail: yakravchenko@sfedu.ru; phone: +79289080151; the department of computer aided design, associate professor.

УДК 519.224.22

DOI 10.18522/2311-3103-2021-2-167-181

А.К. Мельников

**ОГРАНИЧЕНИЕ КОЛИЧЕСТВА РАЗЛИЧНЫХ ОПРОБУЕМЫХ
ВЕКТОРОВ ДЛЯ ПОЛУЧЕНИЯ ВСЕХ РЕШЕНИЙ СИСТЕМЫ
ЛИНЕЙНЫХ УРАВНЕНИЙ ВТОРОЙ КРАТНОСТИ
НА МНОГОПРОЦЕССОРНОЙ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЕ**

Статья посвящена нахождению всех целочисленных неотрицательных решений системы линейных уравнений второй кратности типов, далее с.л.у., методом последовательного опробования векторов на принадлежность к решениям системы. Рассматривается количество различных векторов, опробование которых на принадлежности к решениям с.л.у. приведет к получению всех решений с.л.у. Вектор опробований с.л.у. состоит из элементов определяющих число знаков алфавита, имеющих одинаковое число вхождений в выборку. С.л.у. связывает между собой число вхождений элементов всех типов в рассматриваемую выборку, мощность алфавита, объём выборки и ограничение на максимальное число вхождений знаков алфавита в выборку. Решение с.л.у. является основой расчета точных распределений вероятностей значений статистик и их точных приближений методом второй кратности, где в качестве точных приближений выступают Δ -точные распределения, отличающиеся от точных распределений не более чем на заранее заданную, сколь угодно малую величину Δ . Величина, выражающая количество опробуемых векторов, является одной из величин определяющих алгоритмическую сложность метода второй кратности, без знания значения которой нельзя определить параметры выборок, для которых при ограничениях на вычислительный ресурс могут быть рассчитаны точные распределения и их точные приближения. Количество различных опробуемых векторов рассматривается в условиях ограничения на максимальное значение числа вхождений элементов алфавита в выборку, так и без ограничений. Найдены аналитические выражения, позволяющие для любых значений мощности алфавита, объёма выборки и ограничения на значение максимального числа вхождений знаков алфавита в выборку вычислять количество опробований различных векторов для получения всех целочисленных неотрицательных решений системы линейных уравнений второй кратности типов. Вид полученного аналитического выражения для количества опробований векторов позволяет использовать его при изучении алгоритмической сложности расчетов точных распределений и их точных приближений с заранее указанной точностью Δ .

Вероятность; точное распределение; точное приближение; система линейных уравнений; алгоритмическая сложность; многопроцессорная вычислительная система.

A.K. Melnikov

LIMITING THE NUMBER OF DIFFERENT TEST VECTORS TO OBTAIN ALL SOLUTIONS OF A SYSTEM OF THE SECOND MULTIPLICITY LINEAR EQUATIONS ON MULTIPROCESSOR COMPUTER SYSTEM

In the paper we consider calculation of all integer nonnegative solutions of a linear equation system (LES) of the second types order by a method of sequential vector testing. The method checks whether a vector is a solution of the LES. We consider different vectors and test if they belong to the set of the LES solutions. As a result, after such testing we obtain all solutions of the LES. The LES testing vector consists of the elements which are the numbers of some alphabet signs with the same number of occurrences in the sample. The LES unites the number of occurrences of the elements of all types into the considering sample, the power of the alphabet, the size of the sample, and the limitation for the maximum number of occurrences of the alphabet signs into the sample. The LES solution is the base for calculation of exact statistics probability distributions and their exact approximations by the method of the second types order. Here, the exact approximations are Δ -exact distributions. The difference between the Δ -exact distributions and the exact distributions does not exceed the predefined arbitrary small value Δ . The number of test vectors is one of those which defines algorithmic complexity of the method of second types order. Without it, it is impossible to define the parameters of samples, and to calculate exact distributions and their exact approximations for limited hardware resource. We consider various test vectors for the limited maximum number of occurrences of the alphabet signs in the sample, and for the unlimited one. We have obtained formulas to calculate the number of tests for various vectors. Here, the values of the power of the alphabet, the size of the sample, and the limitations for the maximum number of occurrences of the alphabet signs into the sample can be arbitrary. Using the obtained formulas, we can get all integer nonnegative solutions of the LES of the second types order. We can use the obtained formula for analysis of algorithmic complexity of calculations of exact distributions and their exact approximations with the predefined accuracy Δ .

Probability; exact distribution; exact approximation; linear equations system; algorithmic complexity; multiprocessor computer system.

Введение. Информационные технологии позволяют рационально использовать современные достижения в области развития многопроцессорных вычислительных систем (МВС) [1, 2] и средств автоматизации [3] для решения задачи статистического анализа последовательностей. Данная задача возникает при распознавании текстов на естественных языках [4, 5], обработки информации в дискретных каналах связи [6] и при противодействии кибератакам на вычислительные ресурсы [7].

Статистический анализ последовательностей при необходимости разделения гипотез проводится, в частности при помощи критериев согласия. Для получения наибольшей относительной эффективности [8,9] критерия необходимо применять точные распределения [10]. Применение распределений отличных от точных например таких как предельные распределения ведет к потере эффективности обработки, что выражается в увеличении ложно принятых решений о справедливости проверяемых гипотез.

Расчет точных распределений является вычислительно трудоёмкой задачей [11], поэтому для расширения диапазона значений параметров выборок, для которых возможен расчет точных распределений, наряду с использованием возможностей МВС необходимо модернизировать методы их расчета. Ограничение числа опробований при поиске всех решений уравнения второй кратности типов позволит уменьшить алгоритмическую сложность метода расчета точных распределений и увеличить диапазон параметров выборок для проведения эффективной статистической обработки последовательностей.

Метод второй кратности [12] расчета распределений вероятностей значений статистик [13], основанный на решении системы линейных уравнений второй кратности типов [14], применяется как для расчета точных распределений вероят-

ностей значений статистик [15], далее точных распределений, так и для расчета их точных приближений, отличающихся от точных распределений не более чем на заранее заданную, сколь угодно малую величину Δ (Δ -точных распределений [16]). Для оценки алгоритмической сложности [17] метода второй кратности при расчете точных распределений и их точных приближений, требуется оценить количество различных опробуемых векторов, необходимое для нахождения всех решений системы линейных уравнений второй кратности типов [14].

Данная работа посвящена расчету количества различных опробуемых векторов, необходимых для получения всех решений системы линейных уравнений второй кратности.

Постановка задачи. Основой применения метода второй кратности для расчета точных распределений [15] является получение всех решений системы линейных уравнений (с.л.у.) второй кратности типов

$$\begin{cases} \mu_0 + \mu_1 + \dots + \mu_n = N, \\ 1\mu_1 + 2\mu_2 + \dots + n\mu_n = n \end{cases}, \quad (1)$$

где N мощность алфавита $A_N = \{a_1, a_2, \dots, a_N\}$, n длина последовательности (объема выборки [18]), а μ_j есть количество знаков алфавита A_N , встретившихся в последовательности (выборке) j раз. Для получения всех решений с.л.у (1) может использоваться метод последовательного перебора (м.п.п.), состоящий из генерации всех векторов $\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_n^{(i)})\}$ длины $(n+1)$

$$\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_n^{(i)}), i = 1, \dots, L_{\mu(N,n,n)}\},$$

с одновременной подстановкой каждого $\mu^{(i)}$ в с.л.у. (1) и их непосредственной проверкой, опробованием, на возможность быть его решением, а $L_{\mu(N,n,n)}$ есть количество опробуемых векторов $\mu^{(i)}$. Опробование может проводиться в лексикографическом порядке [19, 20], начиная с вектора $\mu^{(1)}$,

$$\mu^{(1)} = \underbrace{(0, 0, \dots, 0, 1)}_{n+1}$$

с учетом ограничений, накладываемых видом с.л.у. второй кратности типов.

Алгоритмическая сложность [17] получения всех решений с.л.у. вида (1) методом последовательного перебора напрямую зависит от количества опробуемых векторов поэтому получение точной оценки числа опробуемых векторов является важнейшей задачей при определении алгоритмической сложности расчета точных распределений.

Применение метода второй кратности [12] для расчета точных приближений распределений [16] состоит в нахождении решений с.л.у. с ограничениями вида

$$\begin{cases} \mu_0 + \mu_1 + \dots + \mu_r = N, \\ 1\mu_1 + 2\mu_2 + \dots + r\mu_r = n. \end{cases}, \quad (2)$$

где r выступает в качестве параметра ограничения и $r \leq n$.

Получение решений с.л.у. вида (2) также проводится методом последовательного перебора в лексикографическом порядке всех векторов $\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_r^{(i)})\}$ длины $(r+1)$

$$\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_r^{(i)}), i = 1, \dots, L_{\mu(N,n,r)}\}$$

с последующей подстановкой в с.л.у. (2) и непосредственной проверкой, опробованием, где $L_{\mu(N,n,r)}$ количество перебираемых векторов.

Особо подчеркнем, что ведется поиск только неотрицательных целочисленных решений с.л.у. (1) и (2), т.е. для с.л.у. (1) справедливо $\{\forall \mu_j^{(i)} \in \mathbb{N}, \mu_j^{(i)} \geq 0 \mid i = 1, \dots, L_{\mu(N,n,r)}; j = 1, \dots, n\}$, а для с.л.у. (2) справедливо $\{\forall \mu_j^{(i)} \in \mathbb{N}, \mu_j^{(i)} \geq 0 \mid i = 1, \dots, L_{\mu(N,n,r)}; j = 1, \dots, r\}$.

Алгоритмическая сложность получения всех $K_{\mu(N,n,r)}$ решений [21] с.л.у. вида (2) методом последовательного перебора тоже напрямую зависит от количества опробуемых векторов $L_{\mu(N,n,r)}$ поэтому получение точной оценки числа опробуемых векторов также является важнейшей задачей при определении алгоритмической сложности расчета точных приближений. Процедура последовательного опробования векторов и получения решений схематично показана на рис. 1.

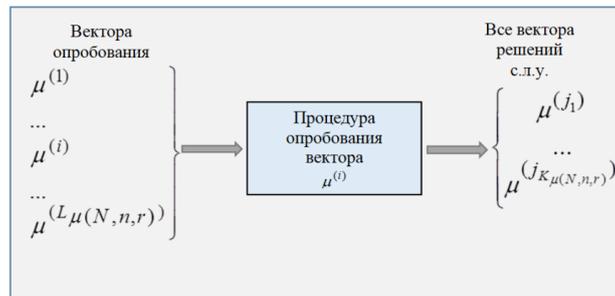


Рис. 1. Процедура последовательного опробования векторов для получения всех решений с.л.у.

Общая задача состоит в оценке значений $L_{\mu(N,n,n)}$ и $L_{\mu(N,n,r)}$ для дальнейшего использования этих оценок при вычислении алгоритмической сложности расчетов точных распределений и их точных приближений.

Оценка количества опробований. Оценим значение $L_{\mu(N,n,r)}$. Из построения с.л.у. второй кратности типов (2) координата $\mu_0^{(i)}$ есть число не встретившихся знаков $(a_{i_1}, a_{i_2}, \dots, a_{i_{\mu_0}})$ алфавита $A_N = \{a_1, a_2, \dots, a_N\}$ в исследуемой последовательности (выборке) длины n , составленной из знаков алфавита A_N мощности N . Следовательно координата $\mu_0^{(i)}$ может изменяться от 0 до N .

Отметим, что $\mu_0^{(i)} = N$ означает тот случай, когда в исследуемой последовательности не встретился ни один из знаков алфавита A_N , но это означает то, что последовательность не была реализована на практике, либо все её символы из другого алфавита. Но для общности будем рассматривать и этот случай, т.е. $0 \leq \mu_0^{(i)} \leq N$.

Каждая координата $\mu_j^{(i)}$, по определению является числом знаков алфавита A_N , встретившихся j раз, и, исходя из второго уравнения с.л.у. (2), не может превосходить минимума от N и $\lceil n/j \rceil$

$$\{0 \leq \mu_j^{(i)} \leq \min(N, \lceil n/j \rceil) \mid j = \overline{1, r}\}$$

и значит координаты $\{\mu_j^{(i)} \mid j = \overline{1, r}\}$ в случае $\{N \leq \lceil n/j \rceil \mid j = \overline{1, r}\}$ изменяются в диапазоне $0 \leq \mu_j^{(i)} \leq N$ и каждая из них может принимать $(N+1)$ различное значение.

Теперь, исходя из вышесказанного можно сделать вывод, что каждая из координат $\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \dots, \mu_r^{(i)})\}$ может изменять в диапазоне $\{0 \leq \mu_j^{(i)} \leq N \mid j = \overline{1, r}\}$ и принимать $(N+1)$ различное значение, всего координат $(r+1)$, следовательно в соответствии с [22] все вместе они могут принимать $(N+1)^{r+1}$ значений. Следовательно верхняя грубая оценка $P_1(L_{\mu(N,n,r)})$ значения $L_{\mu(N,n,r)}$ равна

$$L_{\mu(N,n,r)} \leq P_1(L_{\mu(N,n,r)}) = (N+1)^{r+1}. \quad (3)$$

Последовательно опробуя различные вектора $\mu^{(i)}$ мы можем и раньше найти все $K_{\mu}(N, n, r)$ [21] решения с.л.у. (2), так как не знаем как они распределены, но опробовав все $P_1(L_{\mu(N,n,r)})$ мы гарантировано получим все решения с.л.у. (2).

Далее, используя свойства координат $\mu_j^{(i)}$ векторов решений с.л.у. (2) $\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_r^{(i)})\}$ можно сформулировать и доказать утверждение, позволяющее получить более точную, ограниченную оценку $L_{\mu(N,n,r)}$ числа опробуемых векторов, при которой не будет потеряно ни одно из целочисленных неотрицательных решений с.л.у. вида (2).

Утверждение.

При опробовании векторов $\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_r^{(i)})\}$ для поиска решений с.л.у. (2), с ограничениями r на число переменных $\mu_j^{(i)}$, $j \leq r \leq n$, методом последовательного перебора количество опробуемых векторов $L_{\mu(N,n,r)}$ ограничивается оценкой $P_2(L_{\mu(N,n,r)})$ принимающей следующее значение:

$$\begin{aligned} L_{\mu(N,n,r)} &\leq P_2(L_{\mu(N,n,r)}) = \\ &= (N+1)^{\min(\lceil n/N \rceil, r)+1} \times \frac{(\min(\lceil n/N \rceil, r))!}{(n + \min(\lceil n/N \rceil, r))!} \cdot \frac{(n+r)!}{r!}, \end{aligned} \quad (4)$$

где N мощность алфавита $A_N = \{a_1, a_2, \dots, a_N\}$, n длина последовательности (объёма выборки) и $r \leq n$ ограничение на количество переменных с.л.у. (2).

Доказательство.

Попробуем найти для $L_{\mu(N,n,r)}$ более точную оценку $P_2(L_{\mu(N,n,r)})$. Для этого рассмотрим всевозможные соотношения между N , n и r .

В начале рассмотрим случай $N \leq n$.

При выполнении условия $N \leq n$ очевидно, что всегда найдется такое минимальное k , что

$$k \cdot N = n + \sigma,$$

где σ остаток от целочисленного деления n на N , а k есть антье от частного от деления n на N

$$k = [n / N].$$

Рассмотрим диапазон изменений координаты $\mu_1^{(i)}$, являющейся числом знаков алфавита A_N , встретившихся в исследуемой последовательности только один раз. Исходя из второго уравнения в с.л.у. (2)

$$1\mu_1 + 2\mu_2 + \dots + r\mu_r = n$$

координата $\mu_1^{(i)}$ может изменяться в следующем диапазоне $0 \leq \mu_1^{(i)} \leq \min(N, n)$, что эквивалентно выражению $0 \leq \mu_1^{(i)} \leq \min(N, [n/1])$.

Тогда при рассматриваемом случае $N \leq n$ координата $\mu_1^{(i)}$ может изменяться в диапазоне $0 \leq \mu_1^{(i)} \leq N$ и, следовательно, $\mu_1^{(i)}$ может принимать $(N+1)$ различных значений.

Напомним, что при ограничении на число координат $\{\mu_j^{(i)}, j = \overline{r+1, n}\}$, при которых все остальные $\{\mu_j^{(i)} | j = \overline{r+1, n}\}$ равняются 0 и следовательно могут принимать только одно значение.

Рассмотрим изменение координат $\{\mu_j^{(i)} | j = \overline{1, r}\}$.

Если $k < r \leq n$ то:

Координаты $(\mu_1^{(i)}, \dots, \mu_k^{(i)})$ могут изменяться в диапазоне $\{0 \leq \mu_j^{(i)} \leq N | j = \overline{1, k}\}$, так как для $\{\mu_j^{(i)} | j = \overline{1, k}\}$ в условиях $k = [n/N]$ и $\{j = \overline{1, k}\}$ всегда справедливо $\{\min(N, [n/j]) = N | j = \overline{1, k}\}$ и каждая координата может принимать $(N+1)$ различных значений, а все вместе они могут принимать значений $(N+1)^k$ значений.

Следующие $r - k$ координат $(\mu_{k+1}^{(i)}, \dots, \mu_r^{(i)})$ изменяются в рамках диапазона

$$\{0 \leq \mu_j^{(i)} \leq \min(N, [n/j]) | j = \overline{k+1, r}\},$$

а так как

$$\{\min(N, [n/j]) = [n/j] | j = \overline{k+1, r}\}$$

то значит координаты $\{\mu_j^{(i)} | j = \overline{k+1, r}\}$ могут изменяться в диапазоне $0 \leq \mu_j^{(i)} \leq [n/j]$ и каждая из них может принимать $[n/j] + 1$ различных значений.

Тогда все вместе координаты $\{\mu_j^{(i)} | j = \overline{k+1, r}\}$ могут принимать [23] не более

$$\prod_{j=k+1}^r \left(\binom{n}{j} + 1 \right) \leq \left(\binom{n}{k+1} + \frac{k+1}{k+1} \right) \cdot \left(\binom{n}{k+2} + \frac{k+2}{k+2} \right) \cdot \dots \cdot \left(\binom{n}{r} + \frac{r}{r} \right) =$$

$$= \left(\frac{n+k+1}{k+1} \right) \cdot \left(\frac{n+k+2}{k+2} \right) \cdot \dots \cdot \left(\frac{n+r}{r} \right) = \frac{(n+k+1) \cdot (n+k+2) \cdot \dots \cdot (n+r)}{(k+1) \cdot (k+2) \cdot \dots \cdot r} =$$

значений. Что умножением полученного значения на 1 равную

$$1 = \frac{k! \cdot (n+k)!}{k! \cdot (n+k)!}$$

и соответствующей группировкой сомножителей

$$= \frac{k! \cdot (n+k)! \cdot (n+k+1) \cdot (n+k+2) \cdot \dots \cdot (n+r)}{k! \cdot (n+k)! \cdot (k+1) \cdot (k+2) \cdot \dots \cdot r} =$$

$$= \frac{k! \cdot (n+k)! \cdot (n+k+1) \cdot (n+k+2) \cdot \dots \cdot (n+r)}{(n+k)! \cdot k! \cdot (k+1) \cdot (k+2) \cdot \dots \cdot r} =$$

$$= \frac{k! \cdot (n+k)! \cdot (n+k+1) \cdot (n+k+2) \cdot \dots \cdot (n+r)}{(n+k)! \cdot k! \cdot (k+1) \cdot (k+2) \cdot \dots \cdot r} =$$

приводится к виду

$$= \frac{k! \cdot (n+r)!}{(n+k)! \cdot r!}.$$

При выполнении условия $k < r \leq n$ координаты $\{\mu_j^{(i)} \mid j = \overline{r+1, n}\}$ вектора $\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \dots, \mu_k^{(i)}, \mu_{k+1}^{(i)}, \dots, \mu_r^{(i)}, \mu_{r+1}^{(i)}, \dots, \mu_n^{(i)})\}$ по определению равны 0, т.е. могут принимать единственное значение

$$\{\mu_j^{(i)} = 0 \mid j = \overline{r+1, n}\}.$$

Теперь выпишем число различных значений, которое могут принимать все координаты $\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_k^{(i)}, \mu_{k+1}^{(i)}, \dots, \mu_r^{(i)}, \mu_{r+1}^{(i)}, \dots, \mu_n^{(i)})\}$ при одновременном выполнении условий $N \leq n$ и $k < r \leq n$. Это число будет равно произведению числа значений [23], которое могут принимать координаты: $\mu_0^{(i)}$, $\{\mu_j^{(i)} \mid j = \overline{1, k}\}$, $\{\mu_j^{(i)} \mid j = \overline{k+1, r}\}$ и $\{\mu_j^{(i)} \mid j = \overline{r+1, n}\}$:

- ◆ координата $\mu_0^{(i)}$ – $(N+1)$ различных значений,
- ◆ координаты $\{\mu_j^{(i)} \mid j = \overline{1, k}\}$ – $(N+1)^k$ и числа значений,
- ◆ координаты $\{\mu_j^{(i)} \mid j = \overline{k+1, r}\}$ – $= \frac{k!}{(n+k)!} \cdot \frac{(n+r)!}{r!}$ различных значений,
- ◆ координаты $\{\mu_j^{(i)} \mid j = \overline{r+1, n}\}$ – единственное нулевое значение.

Таким образом

$$\{P_2(L_{\mu(N,n,r)}) / (N \leq n, k < r \leq n)\} = (N+1)^{k+1} \times \frac{k!}{(n+k)!} \cdot \frac{(n+r)!}{r!}. \quad (5).$$

Где в (5) под выражением $\{P_2(L_{\mu(N,n,r)}) / (N \leq n, k < r \leq n)\}$ понимаем значение оценки $P_2(L_{\mu(N,n,r)})$ при выполнении условия $(N \leq n, k < r \leq n)$.

При рассмотренных условиях $N \leq n$ и $k < r \leq n$ выполняется $\min([n/N], r) = [n/N] = k$ и выражение (4) из условия утверждения принимает вид,

$$\begin{aligned} P_2(L_{\mu(N,n,r)}) &= (N+1)^{\min([n/N], r)+1} \times \frac{(\min([n/N], r))!}{(n + \min([n/N], r))!} \cdot \frac{(n+r)!}{r!} = \\ &= (N+1)^{k+1} \times \frac{k!}{(n+k)!} \cdot \frac{(n+r)!}{r!} \end{aligned}$$

равный выражению (5).

В рассматриваемом случае $N \leq n$ при условии $r \leq k \leq n$ будет следующее:

Каждая координата $\mu_j^{(i)}$, по определению являющаяся числом знаков алфавита A_N , встретившихся j раз, исходя из второго уравнения с.л.у. (2), не может превосходить минимума от N и $[n/j]$

$$\{0 \leq \mu_j^{(i)} \leq \min(N, [n/j]) \mid j = \overline{1, r}\}$$

и значит координаты $\{\mu_j^{(i)} \mid j = \overline{1, r}\}$ изменяются в диапазоне $0 \leq \mu_j^{(i)} \leq N$ и каждая из них может принимать $(N+1)$ различное значение.

Теперь, исходя из вышесказанного можно сделать вывод, что каждая из координат $\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \dots, \mu_r^{(i)})\}$ может изменяться в диапазоне $\{0 \leq \mu_j^{(i)} \leq N \mid j = \overline{1, r}\}$ и принимать $(N+1)$ различное значение, а все вместе они могут принимать значений $(N+1)^{r+1}$, следовательно

$$\{P_2(L_{\mu(N,n,r)}) / (N \leq n, r \leq k \leq n)\} = (N+1)^{r+1}. \quad (6)$$

В условиях $N \leq n$ и $r \leq k \leq n$ выполняется $\min([n/N], r) = r$ и выражение (4) из утверждения принимает вид,

$$\begin{aligned} P_2(L_{\mu(N,n,r)}) &= (N+1)^{\min([n/N], r)+1} \times \frac{(\min([n/N], r))!}{(n + \min([n/N], r))!} \cdot \frac{(n+r)!}{r!} = \\ &= (N+1)^{r+1} \times \frac{(r)!}{(n+r)!} \cdot \frac{(n+r)!}{r!} = (N+1)^{r+1}. \end{aligned}$$

равный выражению (6).

Теперь рассмотрим случай, когда $N > n$.

В этом случае координата $\mu_0^{(i)}$ так же может изменяться от 0 до N

$$0 \leq \mu_0^{(i)} \leq N,$$

следовательно $\mu_0^{(i)}$ может принимать $(N+1)$ различных значений.

Число $k=[n/N]$, определяющее сколько мощностей алфавитов полностью укладывается в объёме выборки, равно 0 в нашем случае $N > n$.

Каждая координата $\mu_j^{(i)}$, по определению являющаяся числом знаков алфавита A_N , встретившихся j раз, исходя из второго уравнения с.л.у. (2), не может превосходить минимума от N и $[n/j]$

$$\{0 \leq \mu_j^{(i)} \leq \min(N, [n/j]) \mid j = \overline{1, r}\},$$

но

$$\{\min(N, [n/j]) = [n/j] \mid j = \overline{1, r}\}$$

и значит координаты $\{\mu_j^{(i)} \mid j = \overline{1, r}\}$ могут изменяться в диапазоне $0 \leq \mu_j^{(i)} \leq [n/j]$ и каждая из них может принимать $([n/j]+1)$ различное значение.

Теперь, исходя из вышесказанного можно сделать вывод, что общее количество значений, принимаемых первыми координатами $\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \dots, \mu_r^{(i)})\}$ при одновременном выполнении условий $N > n$ и $r \leq n$ равен произведению количества значений, принимаемых координатой $\mu_0^{(i)}$ и координатами $\{0 \leq \mu_j^{(i)} \leq N \mid j = \overline{1, r}\}$. Это значение равно

$$\begin{aligned} (N+1) \cdot \prod_{i=1}^r \left(\left[\frac{n}{i} \right] + 1 \right) &\leq \left(\frac{n}{1} + \frac{1}{1} \right) \cdot \left(\frac{n}{2} + \frac{2}{2} \right) \cdot \left(\frac{n}{3} + \frac{3}{3} \right) \cdot \dots \cdot \left(\frac{n}{r} + \frac{r}{r} \right) = \\ &= (N+1) \cdot \left(\frac{n+1}{1} \right) \cdot \left(\frac{n+2}{2} \right) \cdot \left(\frac{n+3}{3} \right) \cdot \dots \cdot \left(\frac{n+r}{r} \right) = \\ &= (N+1) \cdot \frac{(n+1) \cdot (n+2) \cdot \dots \cdot (n+r)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot r} = \end{aligned}$$

и далее умножая полученное значение на 1 равную $n!/n!$ и группируя члены получаем

$$= (N+1) \cdot \frac{n!}{n!} \cdot \frac{(n+1) \cdot (n+2) \cdot \dots \cdot (n+r)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot r} = (N+1) \cdot \frac{(n+r)!}{n! \cdot r!}.$$

Следовательно

$$\{P_2(L_{\mu(N,n,r)}) / (N > n)\} = (N+1) \cdot \frac{(n+r)!}{n! \cdot r!}. \quad (7)$$

В случае $N > n$ при условии $r \leq n$ всегда выполняется $[n/N] = 0$ и $\min([n/N], r) = \min(0, r) = 0$, и тогда доказываемое выражение (4) из утверждения принимает вид,

$$\begin{aligned} P_2(L_{\mu(N,n,r)}) &= (N+1)^{\min([n/N], r)+1} \times \frac{(\min([n/N], r))!}{(n + \min([n/N], r))!} \cdot \frac{(n+r)!}{r!} = \\ &= (N+1)^{0+1} \times \frac{(\min(0, r))!}{(n + \min(0, r))!} \cdot \frac{(n+r)!}{r!} = (N+1) \times \frac{(0)!}{(n+0)!} \cdot \frac{(n+r)!}{r!} = \\ &= (N+1) \times \frac{1}{n!} \cdot \frac{(n+r)!}{r!} = (N+1) \times \frac{(n+r)!}{n! \cdot r!} \end{aligned}$$

равный выражению (7).

Таким образом для всевозможные соотношений между N , n и r рассмотрены и значение $L_{\mu(N,n,r)}$ при всех соотношениях равно (4). Утверждение доказано.

Аналитическое выражение (4) определяет количество опробуемых векторов для получения всех решений с.л.у (2) и ограничивает грубую предварительную оценку (3). Выражение (4) учитывает зависимость необходимого количества опробований различных векторов для получения всех решений с.л.у. (2) от мощности алфавита N , длины последовательности n и значения ограничения r , напрямую учитывая соотношение длины последовательности n и мощности алфавита N .

Следствие из утверждения.

При опробовании векторов $\{\mu^{(i)} \mid (\mu_0^{(i)}, \mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_n^{(i)})\}$ для поиска решений с.л.у. (1), без ограничения на число координат $\{\mu_j^{(i)} \mid j = \overline{1, n}\}$, методом последовательного перебора, количество опробуемых векторов $L_{\mu(N, n, n)}$ принимает следующее значение:

$$P_2(L_{\mu(N, n, n)}) = (N + 1)^{[n/N]+1} \times \frac{[n/N]!}{(n + [n/N])!} \cdot \frac{(2n)!}{n!}. \quad (8)$$

Доказательство.

Как следует из утверждения

$$P_2(L_{\mu(N, n, n)}) = (N + 1)^{\min([n/N], r)+1} \times \frac{(\min([n/N], r))!}{(n + \min([n/N], r))!} \cdot \frac{(n+r)!}{r!},$$

но при выполнении условия $n=r$ всегда будет выполнено

$$\min([n/N], r) = \min([n/N], n) = [n/N]$$

и простой подстановкой в выражение (4) значений $r=n$ и $\min([n/N], r) = [n/N]$ получаем (8).

Следствие доказано.

Сравнение оценок количества опробований. Для сравнения оценок количества опробований $P_1(L_{\mu(N, n, n)})$ и $P_2(L_{\mu(N, n, n)})$ при нахождении решений с.л.у. (1), использующейся при расчете точных распределений, рассмотрим их отношение $d(N, n, n)$, назовём его коэффициентом ограничения опробования векторов,

$$d(N, n, n) = P_1(L_{\mu(N, n, n)}) / P_2(L_{\mu(N, n, n)}).$$

Используя результаты грубой оценки (3) для $r=n$ и результаты следствия (8) получаем

$$d(N, n, n) = \frac{(N + 1)^{n+1}}{(N + 1)^{k+1}} \cdot \frac{(n+k)!}{k!} \cdot \frac{n!}{(2n)!} = (N + 1)^{n-k} \cdot \prod_{i=1}^n \frac{(k+i)}{(n+i)}, \quad (9)$$

где $k=[n/N]$ определяет полное число алфавитов мощности N , помещающихся в последовательности длиной n . Рассчитанные согласно (9) коэффициенты ограничения для различных мощностей алфавита частично показаны в табл. 1. Графики изменения коэффициентов ограничения для разных значений мощности алфавита $N=2, 10, 26$ и 64 приведены на рис. 2.

Анализ аналитического выражения (9) для коэффициента ограничения $d(N, n, n)$ показывает, что величина $d(N, n, n)$ является возрастающей функцией по N и n .

Анализ значений коэффициента ограничения количества опробуемых векторов при поиске решений с.л.у. второй кратности, приведенных в табл. 1 и на рис. 2, показывает, что найденная уточненная оценка числа опробований уменьшает количество опробуемых векторов от $d(2, 1, 1) = 1,5$ раз до $d(64, 85, 85) = 1,81 \times 10^{104}$ и более раз по сравнению с грубой оценкой (3) и подтвердил вывод о возрастающем характере величины $d(N, n, n)$.

Таблица 1

Коэффициент ограничения количества опробований векторов СЛУ второй кратности для алфавитов разной мощности N

№ п/п	Длина текста n	Коэффициент ограничения количества опробований векторов для алфавитов разной мощности N $d(N,n,n)$			
		Мощность алфавита $N=2$	Мощность алфавита $N=10$	Мощность алфавита $N=26$	Мощность алфавита $N=64$
1	1	1,50E+00	5,50E+00	1,35E+01	3,25E+01
2	5	2,25E+00	6,39E+02	5,69E+04	4,60E+06
3	10	3,95E+00	1,40E+05	1,11E+09	7,29E+12
4	15	7,21E+00	3,92E+07	1,90E+13	1,01E+19
5	20	1,29E+01	9,32E+09	3,08E+17	1,31E+25
6	25	2,34E+01	2,49E+12	4,81E+21	1,66E+31
7	30	4,18E+01	6,05E+14	8,47E+25	2,06E+37
8	35	7,58E+01	1,59E+17	1,49E+30	2,52E+43
9	40	1,36E+02	3,90E+19	2,54E+34	3,06E+49
10	45	2,46E+02	1,02E+22	4,23E+38	3,67E+55
11	50	4,42E+02	2,51E+24	6,93E+42	4,38E+61
12	55	7,98E+02	6,51E+26	1,18E+47	5,21E+67
13	60	1,43E+03	1,62E+29	2,05E+51	6,16E+73
14	65	2,59E+03	4,17E+31	3,49E+55	7,38E+79
15	70	4,66E+03	1,04E+34	5,86E+59	9,33E+85
16	75	8,41E+03	2,67E+36	9,73E+63	1,17E+92
17	80	1,51E+04	6,67E+38	1,64E+68	1,46E+98
18	85	2,73E+04	1,71E+41	2,82E+72	1,81E+104

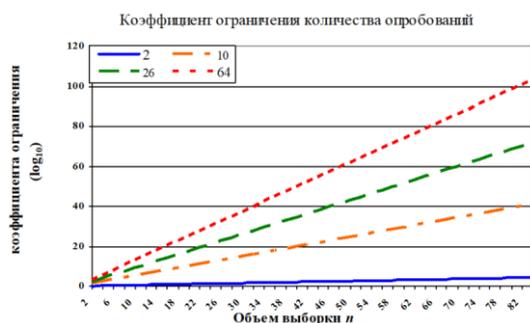


Рис. 2. Значение коэффициентов ограничения количества опробуемых векторов для получения всех решений с.л.у. второй кратности

Заключение и выводы. Рассмотрена задача нахождения всех целочисленных неотрицательных решений системы линейных уравнений второй кратности типов методом последовательного опробования векторов на принадлежность к решениям системы. Учитывая особенности системы линейных уравнений второй кратности типов удалось ограничить число опробуемых векторов, позволяющих получить все решения системы.

Получены аналитические выражения ограниченного числа опробуемых векторов для решения системы линейных уравнений второй кратности с ограничениями на максимальное число вхождений знаков алфавита в выборку и без ограничений.

Вид полученного выражения для ограниченного числа векторов опробования решений позволяет использовать его при изучении алгоритмической сложности расчета точных распределений вероятностей значений статистик и их точных приближений, вычисляемых с заранее заданной точностью Δ .

Применение ограниченного числа опробуемых векторов снижает алгоритмическую сложность получения всех решений системы линейных уравнений второй кратности и как следствие снижает алгоритмическую сложность вычисления точных и предельных распределений. Снижение алгоритмической сложности вычисления распределений позволяет увеличить значения параметров выборки, для которых при ограниченном вычислительном ресурсе могут быть рассчитаны распределения.

Анализ коэффициента ограничения, показывающего количество раз в которое, благодаря уточненной оценке, уменьшается сложность получения всех решений системы линейных уравнений, показывает положительность его значений и рост при увеличении значений мощности алфавита и объема выборки. Коэффициент ограничения растет от 1,5 до 10 в степени 50 и более.

Применение многопроцессорных вычислительных систем для проведения расчетов точных распределений позволяет увеличить вычислительный ресурс для расчета распределений, что вместе со снижением алгоритмической сложности расчета распределений позволит увеличить параметры выборок, для которых эти распределения могут быть рассчитаны.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Гузик В.Ф., Каляев И.А., Левин И.И.* Реконфигурируемые вычислительные системы / под общ. ред. И.А. Каляева. – Таганрог: Изд-во ЮФУ, 2016. – 472 с. – ISBN 978-5-9275-1918-7.
2. *Корнеев В.В.* Вычислительные системы. – М.: Гелиос АРВ, 2004. – 512 с. – ISBN 5-85438-117-6.
3. *Мельников А.К.* Автоматизация процедуры анализа сообщений на гибридных высокопроизводительных вычислительных комплексах // Суперкомпьютерные технологии (СКТ-2016): Матер. 4-й Всероссийской научно-технической конференции: в 2 т. – Ростов-на-Дону: Изд-во ЮФУ, 2016. – С. 69-73. – ISBN 978-5-9275-2039-8 (Т. 1).
4. *Чеповский А.М.* Информационные модели в задачах обработки текстов на естественных языках. – М.: Национальный открытый университет «ИНТУИТ», 2015. – 228 с. – ISBN 978-5-9556-0176-2.
5. *Двойникова А.А., Карпов А.А.* Аналитический обзор подходов к распознаванию тональности русскоязычных текстовых данных // Информационно-управляющие системы. – 2020. – № 4. – С. 20-30. – Doi: 10.31799/1684-8853-2020-4-20-30.
6. *Мальцев Г.Н., Джумков В.В.* Аддитивная граница вероятности ошибки в дискретном канале передачи информации с помехоустойчивым кодированием и группированием ошибок // Информационно-управляющие системы. – 2020. – № 4. – С. 78-86. – Doi: 10.31799/1684-8853-2020-4-78-86.

7. Дикий Д.И. Метод обнаружения DoS-атак на прикладном уровне в сетях «издатель-подписчик» // Информационно-управляющие системы. – 2020. – № 4. – С. 50-60. – Doi: 10.31799/1684-8853-2020-4-50-60.
8. Ронжин А.Ф. Асимптотическая локальная относительная эффективность (АЛОЭ) критериев согласия // Тез. докладов Всесоюзной конференции «Вероятностные методы в дискретной математике». – Петрозаводск, 1983. – С. 70-71.
9. Мельников А.К. Сравнение эффективности обработки текстов при применении в статистических критериях точных и предельных приближений базовых распределений вероятностей значений тестовых статистик // Обзорение прикладной и промышленной математики. – 2018. – Т. 25. – Вып. 4. – С. 375-378. – ISSN 0869-8325. – <https://tvp.ru/conferen/vsppmXIX/repso114.pdf> (дата обращения: 14.01.2019).
10. Мельников А.К. Применение точных распределений в процедуре двухэтапной обработки текстов // Обзорение прикладной и промышленной математики. – 2018. – Т. 25. – Вып. 2. – С. 89–95. – ISSN 0869-8325. – <https://tvp.ru/conferen/vsppmXIX/repso051.pdf> (дата обращения: 24.01.2019).
11. Мельников А.К. Сложность расчета точных распределений вероятности симметричных аддитивно разделяемых статистик и область применения предельных распределений // Доклады ТУСУР. – Томск, 2017. – Т. 20, № 4. – С. 126-130. – ISSN 1818-0442.
12. Мельников А.К. Методика расчета распределения вероятностей значений симметричных аддитивно разделяемых статистик, приближенных к их точному распределению // Научный вестник НГТУ. – 2018. – № 1 (70). – С. 153-166. – ISBN 1814-1196. – Doi: 10.17212/1814-1196-2018-1-153-166.
13. Кендалл М.Г., Стьюарт А. Теория распределений. – М.: Наука, 1966. – 302 с.
14. Зуев Ю.А. Современная дискретная математика: От перечислительной комбинаторики до криптографии XXI века. – М.: ЛЕНАРД, 2019. – 720 с. (Основы защиты информации № 17). – ISBN 978-5-9710-5660-7.
15. Крамер Г. Математические методы статистики: пер. с англ. – М.: Мир, 1975. – 648 с.
16. Мельников А.К., Ронжин А.Ф. Обобщенный статистический метод анализа текстов, основанный на расчете распределений вероятности значений статистик // Информатика и её применения. – 2016. – Т. 10. – Вып. 4. – С. 91-97. – ISSN 1992-2264.
17. Шурыгин В.А. Сложностный метод теории алгоритмов. – М.: Книжный дом "ЛИБРОКОМ", 2009. – 200 с.
18. Фишер Р.А. Статистические методы для исследователей: пер. с англ. – М.: Госстатиздат, 1958. – 73 с.
19. Алексеев В.Е. Дискретная математика: учеб. пособие. – Н. Новгород: Нижегородский госуниверситет, 2017. – 139 с.
20. Кнут Д.Э. Искусство программирования для ЭВМ. Т. 3. Сортировка и поиск. – М.: Мир, 1978. – 844 с.
21. Свидетельство о государственной регистрации программы для ЭВМ № 2020664143. Расчет числа решений системы линейных уравнений второй кратности типов. Правообладатель Колодзей А.В. Автор: Колодзей А.В. и Мельников А.К. Заявка № 2020662473. Дата поступления 19 октября 2020 г. Дата государственной регистрации в Реестре программ для ЭВМ 9 ноября 2020 г.
22. Холл М. Комбинаторика. – М.: Мир, 1970. – 424 с.
23. Риордан Дж. Введение в комбинаторный анализ. – М.: Изд-во иностранной лит., 1963. – 288 с.

REFERENCES

1. Guzik V.F., Kalyaev I.A., Levin I.I. Rekonfiguriruyemye vychislitel'nye sistemy [Reconfigurable computing systems], under the general ed. of I.A. Kalyaeva. Taganrog: Izd-vo YuFU, 2016, 472 p. ISBN 978-5-9275-1918-7.
2. Korneev V.V. Vychislitel'nye sistemy [Computing systems]. Moscow: Gelios ARV, 2004, 512 p. ISBN 5-85438-117-6.

3. *Mel'nikov A.K.* Avtomatizatsiya protsedury analiza soobshcheniy na gibridnykh vysokoproizvoditel'nykh vychislitel'nykh kompleksakh [Automation of message analysis procedure on hybrid high performance complexes], *Superkomp'yuternye tekhnologii (SKT-2016): Mater. 4-y Vserossiyskoy nauchno-tekhnicheskoy konferentsii* [Proceedings of the 5 th All-Russia Conference on Supercomputer Technologies (SCT-2016)]: in 2 vol. Rostov-on-Don: Izd-vo YuFU, 2016, pp. 69-73. ISBN 978-5-9275-2039-8 (T. 1).
4. *Chepovskiy A.M.* Informatsionnye modeli v zadachakh obrabotki tekstov na estestvennykh yazykakh [Information models in tasks of processing of natural language texts]. Moscow: Natsional'nyy otkrytyy universitet «INTUIT», 2015, 228 p. ISBN 978-5-9556-0176-2.
5. *Dvoynikova A.A., Karpov A.A.* Analiticheskiy obzor podkhodov k raspoznavaniyu tonal'nosti russkoyazychnykh tekstovykh dannykh [Analytical review of approaches to Russian text sentiment recognition], *Informatsionno-upravlyayushchie sistemy* [Information and Control Systems], 2020, No. 4, pp. 20-30. Doi: 10.31799/1684-8853-2020-4-20-30.
6. *Mal'tsev G.N., Dzhumkov V.V.* Additivnaya granitsa veroyatnosti oshibki v diskretnom kanale peredachi informatsii s pomexoustoychivym kodirovaniem i gruppировaniem oshibok [Additive boundary of error probability in a discrete data transmission channel with noise-immune coding and grouping of errors. Informatsionno-upravlyayushchie sistemy], *Informatsionno-upravlyayushchie sistemy* [Information and Control Systems], 2020, No. 4, pp. 78-86. Doi: 10.31799/1684-8853-2020-4-78-86.
7. *Dikiy D.I.* Metod obnaruzheniya DoS-atak na prikladnom urovne v setyakh «izdatel'-podpischik» [DoS attack detection at application level in publish-subscribe networks], *Informatsionno-upravlyayushchie sistemy* [Information and Control Systems], 2020, No. 4, pp. 50-60. Doi: 10.31799/1684-8853-2020-4-50-60.
8. *Ronzhin A.F.* Asimptoticheskaya lokal'naya odnositel'naya effektivnost' (ALOE) kriteriev soglasiya [Asymptotic local relative efficiency (ALRE) of fitting criteria], *Tez. dokladov Vsesoyuznoy konferentsii «Veroyatnostnye metody v diskretnoy matematike»* [Reports of All-USSR conference "Probabilistic methods in discrete mathematics"], Petrozavodsk, 1983, pp. 70-71.
9. *Mel'nikov A.K.* Sravnenie effektivnosti obrabotki tekstov pri primeneni v statisticheskikh kriteriyakh tochnykh i predel'nykh priblizheniy bazovykh raspredeleniy veroyatnostey znacheniy testovykh statistik [Comparison of the efficiency of text processing when applying accurate and limiting approximations of basic probability distributions values of test statistics in statistical criteria], *Obozrenie prikladnoy i promyshlennoy matematiki* [Review of applied and industrial mathematics], 2018, Vol. 25, Issue 4, pp. 375-378. ISSN 0869-8325. Available at: <https://tvp.ru/conferen/vsppmXIX/repso114.pdf> (accessed 14 October 2019).
10. *Mel'nikov A.K.* Primenenie tochnykh raspredeleniy v protsedure dvukhetapnoy obrabotki tekstov [Application of exact distributions in the procedure of two-step text processing], *Obozrenie prikladnoy i promyshlennoy matematiki* [Review of applied and industrial mathematics], 2018, Vol. 25, Issue 2, pp. 89-95. ISSN 0869-8325. Available at: <https://tvp.ru/conferen/vsppmXIX/repso051.pdf> (accessed 24 January 2019).
11. *Mel'nikov A.K.* Slozhnost' rascheta tochnykh raspredeleniy veroyatnosti simmetrichnykh additivno razdelyaemykh statistik i oblast' primeneniya predel'nykh raspredeleniy [The complexity of calculating the exact probability distributions of symmetric additive-separated statistics and the application of limit distributions], *Doklady TUSUR* [Proceedings of Tomsk State University of Control Systems and Radioelectronics]. Tomsk, 2017, Vol. 20, No. 4, pp. 126-130. ISSN 1818-0442.
12. *Mel'nikov A.K.* Metodika rascheta raspredeleniya veroyatnostey znacheniy simmetrichnykh additivno razdelyaemykh statistik, priblizhennykh k ikh tochnomu raspredeleniyu [A method for calculating the probability distribution of values of symmetric additively separated statistics that are close to their exact distribution], *Nauchnyy vestnik NGTU* [Science bulletin of the Novosibirsk state technical university], 2018, No. 1 (70), pp. 153-166. ISBN 1814-1196. Doi: 10.17212/1814-1196-2018-1-153-166.
13. *Kendall M.G., St'yuart A.* Teoriya raspredeleniy [Distribution theory]. Moscow: Nauka, 1966, 302 p.
14. *Zuev Yu.A.* Sovremennaya diskretnaya matematika: Ot perechislitel'noy kombinatoriki do kriptografii XXI veka [Modern Discrete Mathematics: From Enumerative Combinatorics to XXI century Cryptography]. Moscow: LENARD, 2019, 720 p. (Fundamentals of information security No. 17). ISBN 978-5-9710-5660-7.

15. *Kramer G.* Matematicheskie metody statistiki [Mathematical methods of Statistics]: trans. from engl. Moscow: Mir, 1975, 648 p.
16. *Mel'nikov A.K., Ronzhin A.F.* Obobshchennyi statisticheskiy metod analiza tekstov, osnovannyi na raschete raspredeleniy veroyatnosti znacheniy statistik [Generalized statistical method of text analysis based on the calculation of probability distributions of statistical values], *Informatika i ee primeneniya* [Informatics and its Applications], 2016, Vol. 10, Issue 4, pp. 91-97. ISSN 1992-2264.
17. *Shurygin V.A.* Slozhnostnyy metod teorii algoritmov [The complexity method of the theory of algorithms]. Moscow: Knizhnyy dom "LIBROKOM", 2009, 200 p.
18. *Fisher R.A.* Statisticheskie metody dlya issledovateley [Statistical methods for researchers]: trans. from engl. Moscow: Gosstatizdat, 1958, 73 p.
19. *Alekseev V.E.* Diskretnaya matematika: ucheb. posobie [Discrete mathematics. Training manual]. Nizhniy Novgorod: Nizhegorodskiy gosuniversitet, 2017, 139 p.
20. *Knut D.E.* Iskusstvo programmirovaniya dlya EVM. T. 3. Sortirovka i poisk [The art of computer programming. Vol. 3. Sorting and search]. Moscow: Mir, 1978, 844 p.
21. Svidetel'stvo o gosudarstvennoy registratsii programmy dlya EVM № 2020664143. Raschet chisla resheniy sistemy lineynykh uravneniy vtoroy kratnosti tipov. Pravoobladatel' Kolodzey A.V. Avtor: Kolodzey A.V. i Mel'nikov A.K. Zayavka № 2020662473. Data postupleniya 19 oktyabrya 2020 g. Data gosudarstvennoy registratsii v Reestre pro-gramm dlya EVM 9 noyabrya 2020 g. [Certificate of state registration of computer software № 2020664143. Calculation of the number of solutions to a system of linear equations of the second multiplicity of types. Proprietor Kolodzey A.V. Authors: Kolodzey A.V. and Melnikov A.K. Application № 2020662473. Date of filing – October 19, 2020. Date of state registration in the Register of computer software – November 9, 2020].
22. *Kholl M.* Kombinatorika [Combinatorics]. Moscow: Mir, 1970, 424 p.
23. *Riordan Dzh.* Vvedenie v kombinatornyy analiz [Introduction to combinatorial analysis]. Moscow: Izd-vo inostrannoy lit., 1963, 288 p.

Статью рекомендовал к опубликованию д.т.н., профессор И.И. Левин.

Мельников Андрей Кимович – АО «Вычислительные решения»; e-mail: ak@comp-sol.ru, anta-mak@umail.ru; г. Москва, Россия; тел.: +74952870035, +79037660462; к.т.н.; доцент ВАК; г.н.с.

Melnikov Andrey Kimovitch – SC «Computing solutions»; e-mail: ak@comp-sol.ru, anta-mak@umail.ru; Moscow, Russia; phone: +784952870035, +79037660462; cand. of eng. sc.; associate professor of SAC; chief research officer.

Раздел V. Системы управления

УДК 004.9: 007.5

DOI 10.18522/2311-3103-2021-2-182-189

З.Р. Майрансаев, А.Б. Чернышев

ОБОБЩЕННЫЙ КРУГОВОЙ КРИТЕРИЙ АБСОЛЮТНОЙ УСТОЙЧИВОСТИ РАСПРЕДЕЛЕННЫХ СИСТЕМ

Управление системами с распределенными параметрами является одним из сложных и важных разделов кибернетики, как науки об управлении, информации и системах. Необходимость изучения и развития данной научной дисциплины связана с тем, что для управления многими объектами приходится учитывать их геометрические параметры, то есть их пространственную протяженность. К настоящему времени в области теории систем с распределенными параметрами получено много результатов, однако по большей части эти результаты направлены на исследование линейных систем. В процессе исследования нелинейных автоматических систем, в качестве одной из основных задач, решается задача поиска возможных состояний равновесия исследуемой системы. Важнейшими задачами являются также исследование устойчивости таких систем. Используя прием разложения функций, описывающих распределенные сигналы в ряды, согласно общей теории рядов Фурье, выделен класс распределенных систем, в которых допустимо разложение по собственным вектор-функциям. Благодаря такой возможности, передаточная функция, описывающая объект с распределенными параметрами представляется в виде совокупности передаточных функций по отдельным пространственным модам. Для учета пространственных координат вводится понятие «обобщенная координата». Применительно к системам с распределенными параметрами коэффициент усиления пространственно-усилительного звена принят как угловой коэффициент прямой ограничивающей нелинейную характеристику. Разработан и сформулирован цилиндрический критерий абсолютной устойчивости нелинейных распределенных систем, на базе обобщения кругового критерия. Приведена иллюстрация пространственного сектора нелинейности. Впервые разработан обобщенный круговой критерий устойчивости распределенных систем, учитывающий зависимость нелинейной характеристики от пространственных координат. Представлена графическая иллюстрация этого критерия.

Распределенные системы; нелинейная характеристика; устойчивость; пространственный годограф.

Z.R. Mayransaev, A.B. Chernyshev

GENERALIZED CIRCULAR CRITERION FOR THE ABSOLUTE SUSTAINABILITY OF DISTRIBUTED SYSTEMS

Systems control with distributed parameters is one of the complex and important sections of cybernetics, like the science of control, information and systems. The need to study and develop this scientific discipline is due to the fact that to control many objects you have to take into account their geometric parameters, that is, their spatial length. So far, many results have been achieved in the field of distributed system theory, but for the most part these results are aimed at the study of linear systems. In the course of researching non-linear automatic systems, as one of the main tasks, the task of finding possible states of equilibrium of the system under study is solved. Research into the sustainability of such systems is also a major challenge. Using the technique of decomposition of functions that describe distributed signals in rows, according to the general theory of the Fourier series, a class of distributed systems is allocated, in which decompo-

sition by its own vector functions is permissible. Due to this capability, the transmission function that describes an object with distributed parameters appears as a combination of transmission functions in separate spatial mods. The concept of "generalized coordinates" is introduced to take into account the spatial coordinates. For systems with distributed parameters, the spatial-amplifier gain factor is adopted as a direct non-linear angular angular ratio. A cylindrical criterion for the absolute stability of non-linear distributed systems has been developed and formulated, based on the generalization of the circular criterion. An illustration of the spatial sector of nonlinearity is given. For the first time, a generalized circular criterion for the stability of distributed systems has been developed, taking into account the dependence of non-linear characteristics on spatial coordinates. A graphic illustration of this criterion is presented.

Distributed systems; non-linear characteristics; stability; spatial godograph.

Введение. Анализ систем управления, в частности анализ их устойчивости – одна из самых важных задач исследования систем. С технической точки зрения, устойчивость систем управления является свойством, при котором система работает стабильно в нормальном режиме функционирования [1, 2]. Рассмотрим нелинейные системы, в которых возможно выделение нелинейной части в качестве отдельного звена, последовательно соединенного с линейным устойчивым звеном [3, 4]. В этом случае имеется возможность использовать достаточно развитый в теории аппарат передаточных функций, который практически невозможно использовать для нелинейных систем общего вида. Частотные методы анализа и синтеза позволяют обобщить некоторые положения, известные из теории сосредоточенных систем для систем с распределенными параметрами [5, 6].

Круговой критерий для сосредоточенных систем. В теории систем с сосредоточенными параметрами известен круговой частотный критерий абсолютной устойчивости систем управления. Этот критерий отличается тем, что нелинейная характеристика может быть нестационарной. Пусть нелинейная характеристика $z=\varphi(\sigma,t)$ принадлежит сектору $[r,k]$ для всех значений t . Должно выполняться условие:

$$\varphi(0,t)=0; rx\leq\varphi(\sigma,t)\leq kx.$$

Положение равновесия нелинейной системы с нестационарным нелинейным элементом абсолютно устойчиво, если амплитудно-фазовая характеристика устойчивой линейной части не охватывает точек с центром на действительной оси (рис. 1). Для кругового критерия используется обычная, не модифицированная частотная характеристика линейной части, в отличие от критерия Попова [7, 8].

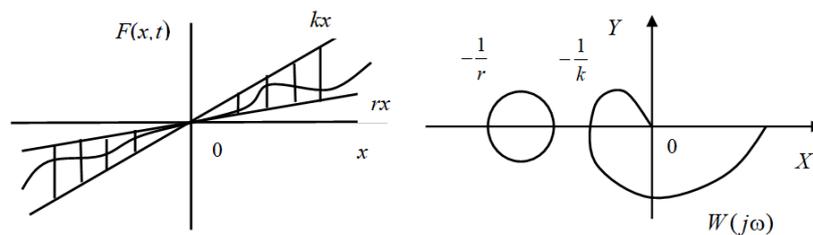


Рис. 1. Иллюстрация кругового критерия абсолютной устойчивости для сосредоточенных систем

Определим параметры круга: диаметр, радиус и координаты центра:

$$D = \frac{1}{r} - \frac{1}{k}, R = \frac{1}{2} \left(\frac{1}{r} - \frac{1}{k} \right), \left(-\frac{1}{2} \left(\frac{1}{r} + \frac{1}{k} \right); 0 \right).$$

Уравнение окружности, образующей круг будет иметь вид:

$$\left(X + \frac{1}{2} \left(\frac{1}{r} + \frac{1}{k} \right) \right)^2 + Y^2 = \frac{1}{4} \left(\frac{1}{r} - \frac{1}{k} \right)^2.$$

Цилиндрический критерий. Используя прием разложения функций, описывающих распределенные сигналы в ряды, согласно общей теории рядов Фурье, рассмотрим класс систем с распределенными параметрами, в которых допустимо разложение по собственным функциям [9–12]. Благодаря такой возможности, передаточная функция, описывающая объект может быть представлена в виде некоторой совокупности передаточных функций по отдельным пространственным модам. Для учета пространственных координат вводится понятие «обобщенная координата» [13]. Рассмотрим коэффициенты k и r как коэффициенты усиления пространственно-усилительного звена:

$$K(G) = E_1 \left[\frac{n_1 - 1}{n_1} + \frac{1}{n_1} G \right], \quad 0 \leq G \leq \infty.$$

где E_1 – общий коэффициент усиления; n_1 – весовой коэффициент; G – обобщенная координата. Уравнения прямых, ограничивающих сектор нелинейной характеристики, для каждого контура можно записать в виде:

$$z_{kn} = E_k \left[\frac{n_1 - 1}{n_1} + \frac{1}{n_1} G_n \right] \sigma_n, \quad z_{rn} = E_r \left[\frac{n_1 - 1}{n_1} + \frac{1}{n_1} G_n \right] \sigma_n. \quad (1)$$

Угловые коэффициенты прямых, образующих сектор, примут вид:

$$k_n = E_k \left[\frac{n_1 - 1}{n_1} + \frac{1}{n_1} G_n \right]; \quad r_n = E_r \left[\frac{n_1 - 1}{n_1} + \frac{1}{n_1} G_n \right].$$

Поверхности, ограничивающие пространственный сектор нелинейности будут иметь вид, изображенный на рис. 2. Множество прямых, ограничивающих сектор нелинейной характеристики, образуют гиперболические поверхности, форма которых зависит от весового коэффициента n_1 , общих коэффициентов усиления E_k и E_r , от обобщенной координаты G [14, 15].

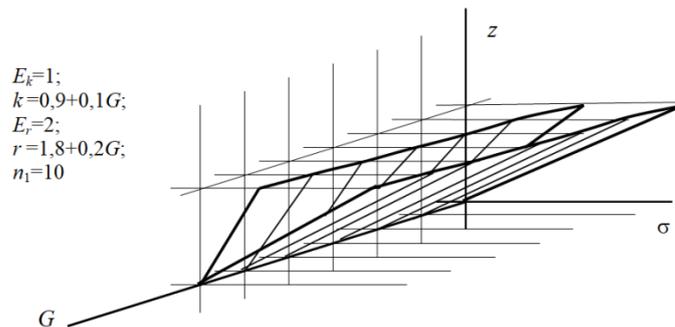


Рис. 2. Пространственный сектор нелинейности

При возрастании весового коэффициента n_1 гиперболические поверхности выпрямляются, при $n_1 \rightarrow \infty$ представляют собой плоскости. При увеличении общих коэффициентов усиления произойдет увеличение угловых коэффициентов k и r для каждого из значений G . Рассмотрим коэффициенты прямых, ограничивающих

сектор нелинейности, для каждого контура при значении весового коэффициента $n_1 \rightarrow \infty$. Это ситуация, когда входное воздействие нелинейного элемента не зависит от пространственных координат, а, следовательно, и от обобщенной координаты G . Из уравнений (1) получим: $k=E_k$; $r=E_r$. В этом случае для каждого контура угловые коэффициенты прямых будут оставаться неизменными и зависят только от постоянных общих коэффициентов усиления. Тогда для каждого контура параметры круга останутся неизменными.

Частотный круговой критерий для нелинейных систем с распределенными параметрами, при нелинейной характеристике, не зависящей от пространственных координат может быть сформулирован следующим образом [16, 17]: Пусть нелинейная характеристика $z=\varphi(\sigma, t)$ принадлежит сектору $[r, k]$ для всех значений t , т.е. выполняется условие:

$$\varphi(0, t)=0; \quad rx \leq \varphi(\sigma, t) \leq kx.$$

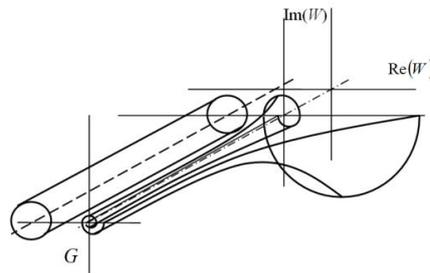


Рис. 3. Графическая иллюстрация цилиндрического критерия

Положение равновесия нелинейной распределенной системы с нестационарным нелинейным элементом, не зависящим от пространственных координат, абсолютно устойчиво, если пространственный годограф линейной части, распределенной системы не охватывает точек цилиндра с осью на действительной оси, при $k=E_k$; $r=E_r$, проходящей через точки:

$$\left(-\frac{1}{2} \left(\frac{1}{r} + \frac{1}{k} \right); \quad 0 \right),$$

радиусом:

$$R = \frac{1}{2} \left(\frac{1}{r} - \frac{1}{k} \right),$$

для всех независимых контуров (рис. 3).

Обобщенный круговой критерий. В реальных системах с распределенными параметрами, чаще всего входное воздействие зависит от пространственных координат [18–20].

Рассмотрим ситуацию при значении весового коэффициента $n_1=1$. Угловые коэффициенты прямых, образующих сектор нелинейности, примут вид:

$$k = E_k G; \quad r = E_r G.$$

Координата центра окружности по действительной оси и радиус примут вид:

$$X_0 = -\frac{1}{2} \left(\frac{1}{E_r G} + \frac{1}{E_k G} \right) = -\frac{1}{2} \left(\frac{E_k + E_r}{E_r E_k G} \right), \quad R = \frac{1}{2} \left(\frac{1}{E_r G} - \frac{1}{E_k G} \right) = \frac{1}{2} \left(\frac{E_k - E_r}{E_r E_k G} \right).$$

Найдем пределы функций, выражающих координаты центра окружности по действительной оси и радиуса окружности при стремлении обобщенной координаты к бесконечности. Получим:

$$\lim_{G \rightarrow \infty} \left[-\frac{1}{2} \left(\frac{E_k + E_r}{E_r E_k G} \right) \right] = 0; \quad \lim_{G \rightarrow \infty} \left[\frac{1}{2} \left(\frac{E_k - E_r}{E_r E_k G} \right) \right] = 0.$$

Таким образом, при увеличении значения обобщенной координаты G , окружности каждого последующего контура будут приближаться к началу координат и уменьшаться по радиусу. Частотный круговой критерий для нелинейных систем с распределенными параметрами, при произвольной нестационарной нелинейной характеристике, может быть сформулирован следующим образом: Пусть нелинейная характеристика $z=\varphi(\sigma, t)$ принадлежит сектору $[r, k]$ для всех значений t , т.е. выполняется условие:

$$\varphi(0, t) = 0; \quad rx \leq \varphi(\sigma, t) \leq kx.$$

Положение равновесия нелинейной распределенной системы с нестационарным нелинейным элементом, абсолютно устойчиво, если пространственный годограф линейной части, распределенной системы не охватывает точек круга с центром на действительной оси:

$$\left(-\frac{1}{2} \left(\frac{1}{r} + \frac{1}{k} \right); 0 \right),$$

радиусом:

$$R = \frac{1}{2} \left(\frac{1}{r} - \frac{1}{k} \right),$$

где для всех независимых контуров:

$$k_n = E_k \left[\frac{n_1 - 1}{n_1} + \frac{1}{n_1} G_n \right]; \quad r_n = E_r \left[\frac{n_1 - 1}{n_1} + \frac{1}{n_1} G_n \right].$$

На рис. 4 представлена графическая иллюстрация этого критерия.

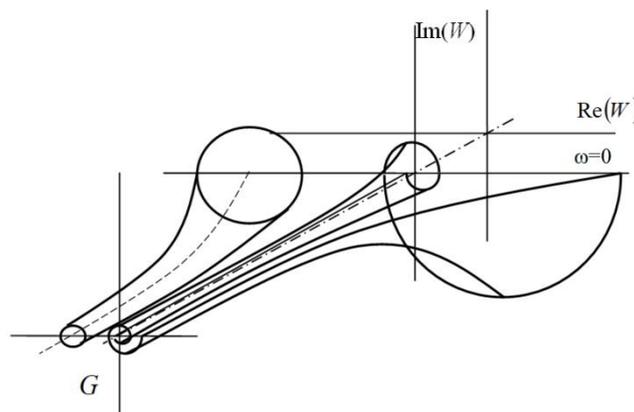


Рис. 4. Графическая иллюстрация обобщенного кругового критерия

Заключение. Разработан и сформулирован цилиндрический критерий абсолютной устойчивости нелинейных распределенных систем, на базе обобщения кругового критерия. Приведена иллюстрация пространственного сектора нелинейности. Разработан обобщенный круговой критерий устойчивости распределенных систем, учитывающий зависимость нелинейной характеристики от пространственных координат. Представлена графическая иллюстрация этого критерия. Анализ устойчивости при помощи обобщенного кругового критерия предполагает рассмотрение для каждого контура взаимного расположения в комплексной плоскости пространственного годографа соответствующего линейной части и круга, определяемого сектором, в котором заключена нелинейная характеристика. Взаимное расположение годографа и круга зависит от номера пространственной моды, от значения весового коэффициента, от величины и соотношения общих коэффициентов усиления.

Полученные результаты позволяют расширить теоретическую базу применительно к нелинейным распределенным системам. Результаты работы могут найти применение в различных отраслях промышленности при решении задач автоматизации процессов управления объектами с распределенными параметрами.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Сиразетдинов Т.К.* Устойчивость систем с распределенными параметрами. – Новосибирск: Наука, 1987. – 232 с.
2. *Душин С.Е., Зотов Н.С., Имаев Д.Х. и др.* Теория автоматического управления / под ред. В.Б. Яковлева. – М.: Высшая школа, 2003.
3. *Майрансаев З.Р., Чернышев А.Б.* Обзор методов исследования различных классов систем управления // Современные фундаментальные и прикладные исследования. – 2018. – № 4 (31). – С. 23-26.
4. *Майрансаев З.Р., Чернышев А.Б.* Проблемы исследования устойчивости нелинейных систем. // Евразийское Научное Объединение. – 2020. – № 12-2 (70). – С. 104-107.
5. *Чернышев А.Б.* Обобщение методов анализа устойчивости для распределенных систем // Современная наука и инновации. – 2015. – № 3 (11). – С. 16-22.
6. *Чернышев А.Б., Антонов В.Ф., Ильюшин Ю.В.* Моделирование релейно-импульсных распределенных систем. – Пятигорск: Изд-во ПГГТУ, 2012. – 248 с.
7. *Чернышев А.Б.* Интерпретация критерия абсолютной устойчивости для нелинейных распределенных систем // Автоматизация и современные технологии. – 2010. – № 2. – С. 28-32.
8. *Чернышев А.Б.* Исследование абсолютной устойчивости нелинейных распределенных систем // Автоматизация и современные технологии. – 2010. – № 4. – С. 21-26.
9. *Рапопорт Э.Я.* Структурное моделирование объектов и систем управления с распределенными параметрами. – М.: Высшая школа, 2003. – 299 с.
10. *Чернышев А.Б.* Модифицированный годограф пространственно-аперидического звена // Вестник Нижегородского университета им. Н.И. Лобачевского. – 2010. – № 2 (1). – С. 159-163.
11. *Майрансаев З.Р., Чернышев А.Б.* Представление распределенных систем в виде совокупности независимых контуров // Тенденции развития науки и образования. – 2021. – № 69-2. – С. 105-108.
12. *Martirosyan A.A., Martirosyan K.V., Chernyshev A.B.* Application of Fourier series in distributed control systems simulation // Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2019. – 2019. – P. 609-613.
13. *Першин И.М.* Синтез систем с распределенными параметрами. – Пятигорск: Изд-во РИА-КМВ, 2002. – 212 с.
14. *Чернышев А.Б., Майрансаев З.Р.* Проблемы гармонической линеаризации систем с распределенными параметрами // Обозрение прикладной и промышленной математики. – 2019. – Т. 26, № 1. – С. 31-42.

15. Чернышев А.Б., Назарцев М.С., Майрансаев З.Р. Гармоническая линеаризация распределенных систем // Современная наука и инновации. – 2018. – № 4 (24). – С. 94-101.
16. Майрансаев З.Р., Чернышев А.Б. Формирование температурного поля в результате точечных импульсных воздействий // Евразийское Научное Объединение. – 2018. – № 12-2 (46). – С. 82-84.
17. Chernyshev A.B., Tkachenko I.V., Mayransaev Z.R. Cylindrical criterion of absolute stability of distributed systems // Современная наука и инновации. – 2020. – № 4 (32). – С. 34-40.
18. Martirosyan A.A., Martirosyan K.V., Chernyshev A.B. Methods of distributed systems' structured modelling // Proceedings of the 2016 IEEE North West Russia Section Young Researchers in Electrical and Electronic Engineering Conference, EIConRusNW 2016. – 2016. – P. 283-289.
19. Martirosyan A.V., Chernyshev A.B., Martirosyan K.V. Calculation of the first switch-on time of distributed objects control action // Proceedings of the 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, EIConRus 2020. – 2020. – P. 750-754.
20. Martirosyan K.V., Chernyshev A.B., Martirosyan A.V., Tatyana K.V. Formation of the anterior heating function under the action of uniformly distributed sources // Proceedings of the 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, EIConRus 2020. – 2020. – P. 755-760.

REFERENCES

1. Sirazetdinov T.K. Ustoychivost' sistem s raspredelennymi parametrami [Stability of systems with distributed parameters.]. Novosibirsk: Nauka, 1987, 232 p.
2. Dushin S.E., Zotov N.S., Imaev D.Kh. i dr. Teoriya avtomaticheskogo upravleniya [Theory of automatic control], ed by V.B. Yakovleva. Moscow: Vysshaya shkola, 2003.
3. Mayransaev Z.R., Chernyshev A.B. Obzor metodov issledovaniya razlichnykh klassov sistem upravleniya [Overview of research methods for various classes of control systems], *Sovremennye fundamental'nye i prikladnye issledovaniya* [Modern fundamental and applied research], 2018, No. 4 (31), pp. 23-26.
4. Mayransaev Z.R., Chernyshev A.B. Problemy issledovaniya ustoychivosti nelineynykh sistem. [Problems of studying the stability of nonlinear systems], *Evraziyskoe Nauchnoe Ob"edinenie* [Eurasian Scientific Association], 2020, No. 12-2 (70), pp. 104-107.
5. Chernyshev A.B. Obobshchenie metodov analiza ustoychivosti dlya raspredelennykh sistem [Generalization of stability analysis methods for distributed systems], *Sovremennaya nauka i innovatsii* [Modern science and Innovation], 2015, No. 3 (11), pp. 16-22.
6. Chernyshev A.B., Antonov V.F., Il'yushin Yu.V. Modelirovanie releyno-impul'snykh raspredelennykh sistem [Simulation of relay-pulse distributed systems]. Pyatigorsk: Izd-vo PGGTU, 2012, 248 p.
7. Chernyshev A.B. Interpretatsiya kriteriya absolyutnoy ustoychivosti dlya nelineynykh raspredelennykh sistem [Interpretation of the absolute stability criterion for nonlinear distributed systems], *Avtomatizatsiya i sovremennye tekhnologii* [Automation and Modern Technologies], 2010, No. 2, pp. 28-32.
8. Chernyshev A.B. Issledovanie absolyutnoy ustoychivosti nelineynykh raspredelennykh sistem [Investigation of absolute stability of nonlinear distributed systems], *Avtomatizatsiya i sovremennye tekhnologii* [Automation and Modern Technologies], 2010, No. 4, pp. 21-26.
9. Rapoport E.Ya. Strukturnoe modelirovanie ob"ektov i sistem upravleniya s raspredelennymi parametrami [Structural modeling of objects and control systems with distributed parameters]. Moscow: Vysshaya shkola, 2003, 299 p.
10. Chernyshev A.B. Modifitsirovannyi godograf prostranstvenno-aperiodicheskogo zvena [Modified space-aperiodic link hodograph], *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo* [Bulletin of the Lobachevsky University of Nizhny Novgorod], 2010, No. 2 (1), pp. 159-163.
11. Mayransaev Z.R., Chernyshev A.B. Predstavlenie raspredelennykh sistem v vide sovokupnosti nezavisimykh konturov [Representation of distributed systems as a set of independent contours], *Tendentsii razvitiya nauki i obrazovaniya* [Trends in the development of science and education.], 2021, No. 69-2, pp. 105-108.

12. Martirosyan A.A., Martirosyan K.V., Chernyshev A.B. Application of Fourier series in distributed control systems simulation, *Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2019*, 2019, pp. 609-613.
13. Pershin I.M. Sintez sistem s raspredelennymi parametrami [Synthesis of systems with distributed parameters]. Pyatigorsk: Izd-vo RIA-KMV, 2002, 212 p.
14. Chernyshev A.B., Mayransaev Z.R. Problemy garmonicheskoy linearizatsii sistem s raspredelennymi parametrami [Problems of harmonic linearization of systems with distributed parameters], *Obozrenie prikladnoy i promyshlennoy matematiki* [Review of Applied and Industrial Mathematics], 2019, Vol. 26, No. 1, pp. 31-42.
15. Chernyshev A.B., Nazartsev M.S., Mayransaev Z.R. Garmonicheskaya linearizatsiya raspredelennykh sistem [Harmonic linearization of distributed systems], *Sovremennaya nauka i innovatsii* [Modern science and innovation], 2018, No. 4 (24), pp. 94-101.
16. Mayransaev Z.R., Chernyshev A.B. Formirovanie temperaturnogo polya v rezul'tate tochechnykh impul'snykh vozdeystviy [Formation of the temperature field as a result of point pulse effects], *Evrasiyskoe Nauchnoe Ob'edinenie* [Eurasian Scientific Association], 2018, No. 12-2 (46), pp. 82-84.
17. Chernyshev A.B., Tkachenko I.V., Mayransaev Z.R. Cylindrical criterion of absolute stability of distributed systems, *Современная наука и инновации* [Modern science and innovation], 2020, No. 4 (32), pp. 34-40.
18. Martirosyan A.A., Martirosyan K.V., Chernyshev A.B. Methods of distributed systems' structured modeling, *Proceedings of the 2016 IEEE North West Russia Section Young Researchers in Electrical and Electronic Engineering Conference, ElConRusNW 2016*, 2016, pp. 283-289.
19. Martirosyan A.V., Chernyshev A.B., Martirosyan K.V. Calculation of the first switch-on time of distributed objects control action, *Proceedings of the 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2020*, 2020, pp. 750-754.
20. Martirosyan K.V., Chernyshev A.B., Martirosyan A.V., Tatyana K.V. Formation of the anterior heating function under the action of uniformly distributed sources, *Proceedings of the 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2020*, 2020, pp. 755-760.

Статью рекомендовал к опубликованию к.т.н. А.Н. Попов.

Майрансаев Зураб Русланович – Северо-Кавказский федеральный университет, институт сервиса, туризма и дизайна (филиал); e-mail: sarmat007@inbox.ru; г. Пятигорск, Россия; тел.: +78793973927; кафедра систем управления и информационных технологий; аспирант.

Чернышев Александр Борисович – e-mail: chalbor@rambler.ru; кафедра систем управления и информационных технологий; профессор.

Mayransaev Zurab Ruslanovich – North Caucasus Federal University, Institute of Service, Tourism and Design (branch); e-mail: sarmat007@inbox.ru; Pyatigorsk, Russia; phone: +78793973927; the department of management systems and information technologies; postgraduate student.

Chernyshev Alexandr Borisovich – e-mail: chalbor@rambler.ru; the department of management systems and information technologies; professor.

УДК 519.711

DOI 10.18522/2311-3103-2021-2-189-199

Д.Е. Чикрин, А.А. Егорчев

СРАВНЕНИЕ МЕТОДОЛОГИЙ ПРОЕКТИРОВАНИЯ СВЕРХУ–ВНИЗ И СНИЗУ–ВВЕРХ ПРИ РАЗРАБОТКЕ СИСТЕМ ADAS

Выбор типа основной методологии проектирования оказывает значительное влияние на качество итогового продукта, в том числе и на его способность к дальнейшему развитию и масштабированию. В статье рассматриваются особенности стандартных мето-

долгий проектирования снизу–вверх и сверху–вниз применительно к системам ADAS (системам автоматизированного (беспилотного) управления автомобилем), показывается, что использование "чистых" методологий неприемлемо при проектировании указанных систем и требуется создание новой совмещённой методологии проектирования. Для этого рассмотрены особенности и ограничения подхода сверху–вниз: ориентация подхода на максимальное соответствие разрабатываемой системы предъявляемым к ней требованиям; методологическая строгость подхода; трудность тестирования системы в процессе разработки; чувствительность к изменениям требований к разрабатываемой системе. Рассмотрены особенности и ограничения подхода снизу–вверх: возможность итеративной разработки с получением промежуточного результата; возможность использования стандартных компонентов; масштабируемость и гибкость системы разрабатываемой системы; возможность несоответствия функций подсистем требованиям, которое может проявляться только на поздних этапах разработки; возможная несогласованность при разработке отдельных подсистем и элементов. Рассмотрены особенности и факторы разработки систем ADAS: повышенные требования по надёжности и безопасности работы системы; разнородность используемых компонентов. Выделены два этапа развития ADAS-систем: этап интенсивной разработки и этап экстенсивной эволюции. Рассмотрена применимость той или иной методологии относительно различных аспектов разработки и эволюции систем ADAS, таких как: определение требований; композиционный морфизм; масштабируемость и расширяемость; стабильность и устойчивость и время разработки; способность к развитию. В результате сравнения методологий делается вывод о том, что существуют аспекты разработки и развития технической системы, в которых наблюдается значительное преимущество одной или другой из методологий. В должной степени эволюция системы может быть обеспечена только при использовании подхода снизу–вверх. Однако, для сложных систем критически важным является определение изначальных требований к системе, что может быть достигнуто только с применением методологии сверху–вниз.

Методология проектирования; сверху–вниз; снизу–вверх; система ADAS; беспилотное транспортное средство; проектирование сложных систем.

D.E. Chickrin, A.A. Egorchev

TOP-DOWN VS BOTTOM-UP METHODOLOGIES FOR ADAS SYSTEM DESIGN

Selection of the principal design methodology has a significant impact on final product quality, including its evolvability and scalability. The article discusses the features of traditional bottom-up and top-down design methodologies in the context of ADAS (driver assistance and automated driving systems). Necessity of the combined design methodology is shown due to unacceptability of "pure" methodologies for design of this kind of systems. For this purpose, the features and limitations of the top-down approach are considered: commitment to maximum compliance of the developed system with its requirements; methodological rigor of the approach; difficulty of system testing in the process of the development; sensitivity to changes in requirements. The features and limitations of the bottom-up approach are considered: possibility of iterative development with obtaining intermediate results; possibility of using standard components; scalability and flexibility of the developed system; possibility of discrepancy of functions of subsystems to requirements, which may appear only at later stages of development; possible inconsistency in development of separate subsystems and elements. The features and factors of ADAS system development are considered: increased requirements for reliability and safety of the system; heterogeneity of used components. Two stages of ADAS-systems development are distinguished: the stage of intensive development and the stage of extensive evolution. The applicability of one or another methodology to various aspects of ADAS system development and evolution (such as: requirements definition; compositional morphism; scalability and extensibility; stability and sustainability; cost and development time; development capability) is considered. A comparison of the

methodologies concludes that there are aspects of technical system design and development in which there is a significant advantage of one or the other of the methodologies. Only the bottom-up approach can ensure the proper evolution of the system. However, for complex systems, it is critical to define the initial requirements for the system, which can only be achieved using the top-down methodology.

Design methodology; bottom-up; top-down; ADAS; self-driving vehicle; complex systems design.

Введение. Корреляция между особенностями методологии проектирования и качеством итоговой системы достаточно подробно показана применительно к программному обеспечению [1]. Современные же системы ADAS представляют собой сложную смесь программных и аппаратных компонент, но в то же время строятся по тем же принципам, что и сложные программные комплексы различного назначения, так как представляют собой автоматизированную систему управления исполнительными механизмами автомобиля [2–4]. Методология сверху–вниз является классической при осознанном (запланированном ранее) создании практически любой крупной аппаратно–программной или программной системы. Действительно, при заказе и планировании создания сложных систем любого рода возникает желание максимально детализировать требования к ним – чтобы получить именно то, что требуется на текущий момент времени. В свою очередь, для создания прототипов и получения быстрых результатов практически всегда используется методология снизу–вверх, использующая блоки – «чёрные ящики» с фиксированной, но уже реализованной функциональностью, позволяющие быстро достичь работающей в первом приближении системы и получить первые результаты и обратную связь от испытателей системы и заказчика работ.

1. Методологии сверху–вниз и снизу–вверх. При разработке программно–аппаратных систем активно применяется как методология сверху–вниз (top–down), так и снизу–вверх (bottom–up).

Методология сверху–вниз подразумевает проведение процесса разработки начиная с определения требований к разрабатываемой системе. Разрабатываемая система описывается с точки зрения того, что необходимо получить в результате, а не каким образом. При этом полное соответствие требованиям достигается на высоких уровнях абстракции, а при продвижении вниз могут возникать проблемы, связанные с технической невозможностью или сложностью конечной реализации запланированной функциональности составных частей системы. Такие проблемы могут быть исправлены двумя способами: 1) внедрением на уровне реализации обходных решений, которые, как правило, снижают эффективность системы и/или повышают стоимость разработки; 2) возвратом на более высокие уровни абстракции системы и переработкой декомпозиции [1]. Так как проблемы в реализации низкоуровневых компонент возникают на поздних этапах разработки, то и изменение становится очень дорогим – потенциально придётся реализовывать заново часть функциональности, и это не говоря о самом процессе повторного дизайна системы [5].

Подход сверху–вниз подразумевает реализацию подсистем, максимально соответствующих поставленным требованиям, что обычно приводит к полной реализации всей функциональности системы с нуля – повторное использование ранее разработанных и/или универсальных компонентов обычно невозможно, так как технические и функциональные возможности уже разработанных систем как правило не учитываются при декомпозиции («что получить, а не как») и могут не соответствовать поставленным требованиям.

Таким образом, можно выделить следующие особенности подхода сверху–вниз:

1. (+) система в полной мере соответствует поставленным изначально требованиям;
2. (+) подход более эффективен с методологической точки зрения;
3. (–) сколько-нибудь законченная система может быть получена только в конце процесса разработки – затрудняется тестирование и оценка получаемого результата;
4. (–) разрабатываемая система нетолерантна к изменению требований в процессе разработки – в худшем случае при изменениях весь процесс разработки необходимо начинать заново.

Методология снизу–вверх широко применяется в робототехнике, встраиваемых системах и сенсорных сетях [6]. При использовании подхода снизу–вверх на начальном этапе разработки определяется набор базовых систем и технологий. С каждой новой итерацией разработки низкоуровневые компоненты связываются и объединяются в более сложные системы, которые начинают проявлять новые свойства, превосходящие свойства простой совокупности отдельных компонентов – проявляется свойство эмерджентности.

Основная проблема методологии снизу–вверх – возможное несоответствие разрабатываемого продукта требованиям. Но при этом, как показано в [1, 7–9], требования к продукту неизбежно меняются, причём начинается этот процесс ещё во время разработки. С другой стороны, использование методологии снизу–вверх позволяет быстро получать прототипы системы, которые могут использоваться для постепенной верификации системы на соответствие требованиям, и, что более важно, для уточнения и обновления самих требований.

Особенности методологии снизу–вверх:

1. (+) возможна итеративная разработка с получением промежуточного результата, который можно использовать для тестирования, уточнения требований и особенностей реализации;
2. (+) активное использование стандартных компонентов позволяет снизить временные и финансовые издержки при разработке;
3. (+) высокая масштабируемость и гибкость системы: базовые элементы системы независимы, что позволяет заменять элементы и добавлять новые;
4. (–) возможные несоответствия функций подсистем высокоуровневым требованиям (недостаточная производительность вычислительных устройств, недостаточная точность сенсоров и т.д.), обнаружить которые возможно только на поздних этапах разработки при реализации взаимодействия высокоуровневых подсистем;
5. (–) при независимой разработке подсистем возможны ситуации, когда некоторые элементы, имеющие схожую функциональность, реализуются независимо и различными способами. При этом усложняется их обслуживание и увеличивается стоимость их разработки.

2. Особенности ADAS–систем. Системы автоматизированного управления транспортным средством – сложные гетерогенные системы управления автомобилем, которые могут быть как вспомогательными для водителя (например, круиз–контроль, автоматическая парковка и т.д.), так и полностью заменять его [10]. Базовыми высокоуровневыми функциями таких систем, как и любых технических систем, являются восприятие окружающей обстановки, принятие решений (планирование) и управление исполнительными механизмами автомобиля. Так как системы беспилотного вождения в качестве базы используют автомобили, к ним предъявляются повышенные требования по надёжности и безопасности работы. В связи с этим, помимо реализации трёх базовых функций, системы такого рода должны иметь возможность контролировать работу системы и обеспечивать отка-

зоустойчивость (путём резервирования подсистем и аварийных режимов), фиксацию системных событий и возможность наблюдения и контроля со стороны человека [11].

Развитие систем ADAS происходит на стыке автомобильной индустрии и робототехники, целесообразно рассмотреть подходы к проектированию, принятые в этих сферах.

Автомобильную промышленность отличают высокие требования к безопасности, устойчивости работы и надёжности всех компонентов системы – такие свойства системы возможно получить при проектировании системы по методологии сверху–вниз. Но в последние десятилетия в автомобилестроении наблюдается стремление к подходу снизу–вверх – активно развиваются модульные платформы, подразумевающие повторное использование базовых компонентов в различных моделях автомобилей. При этом сама платформа разрабатывается в соответствии со всеми требованиями к безопасности и после выпуска не меняется – изменение платформы потребует перестройки процессов производства и возможной потери совместимости с компонентами. Таким образом, процесс разработки в современной автомобильной промышленности строится в соответствии с обеими методологиями – платформа, спроектированная в строгом соответствии с требованиями и актуальная для нескольких поколений автомобилей строится по методологии сверху–вниз, а при проектировании самих автомобилей используется эта базовая платформа и набор дополнительных взаимозаменяемых компонентов в различных комбинациях и вариациях, что соответствует подходу снизу–вверх.

В современной автомобильной промышленности при разработке автомобиля используется множество готовых компонентов от сторонних производителей. При этом соблюдение стандартов в автомобильной промышленности (AUTOSAR, OSEK/VDX, FlexRay, CAN) абсолютно критично, т.к. данные стандарты ориентированы как раз на обеспечение совместной работы оборудования различных производителей, что решает одну из основных проблем подхода снизу–вверх – обеспечение корректного взаимодействия между подсистемами [11–14].

В ADAS–системах компоненты значительно сложнее и разнороднее, а значит и проблема стоит острее, чем для традиционных автомобильных (неинтеллектуальных) систем и узлов. В этой части ADAS–системы близки к робототехническим системам. Робототехнические системы представляют собой комплексы гетерогенных систем с совершенно разным функциональным назначением – различные сенсоры, актуаторы, вычислительные модули. Процессы разработки в робототехнике тяготеют к методологии снизу–вверх: как правило, синтез систем производится из уже готовых компонентов, которые сами по себе достаточно сложны; полная разработка таких подсистем в соответствии с требованиями при применении подхода сверху–вниз радикально увеличивает продолжительность и стоимость разработки.

В качестве примера подобных гетерогенных подсистем можно привести подсистему восприятия беспилотного автомобиля, которая представляет собой совокупность разнородных сенсоров: лидаров, сонаров, радаров, камер, спутниковых и инерциальных систем навигации и т.д. (см., например, [15, 16]). На рис. 1 приведён пример сенсорной подсистемы автомобиля, оснащённого системой ADAS.

Большая часть компонентов сенсорной подсистемы предлагается сторонними производителями, которые как правило, специализируются на разработке и производстве какого-то конкретного типа сенсоров, обладают значительной экспертизой в предметной области и инвестируют в активное развитие этих технологий. Самостоятельная разработка систем такого рода непозволительно затратна даже для крупных автоконцернов, в связи с чем можно сделать вывод, что использование сторонних компонентов в ADAS–системах неизбежно.

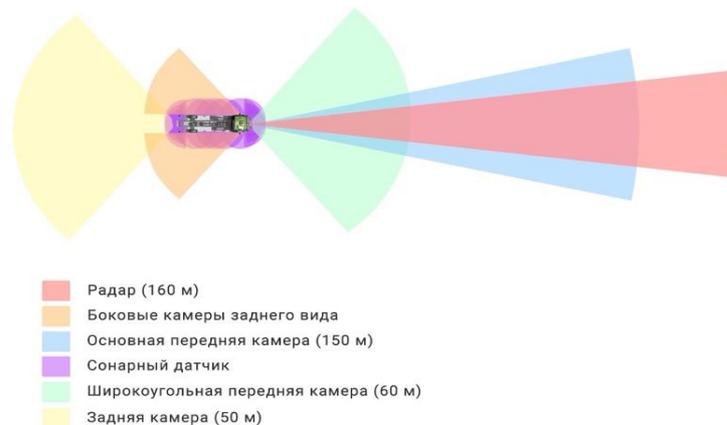


Рис. 1. Сенсорные подсистемы беспилотного автомобиля

Получение требуемой функциональности при минимизации стоимости – одна из ключевых задач при проектировании ADAS–систем. Активное развитие ADAS и робототехнических систем привело к значительному прогрессу в области как аппаратных компонентов, в частности, сенсоров, так и программных решений. Разнообразие решений позволяет соблюсти баланс между стоимостью и функциональностью систем в широком диапазоне степеней автоматизации. Постоянный прогресс в сфере компонентов также приводит к улучшению технических характеристик предлагаемых устройств. В связи с этим ADAS–системы должны иметь возможность адаптироваться и масштабироваться для сохранения актуальности системы – при этом с обязательным обеспечением поддержки стандартных интерфейсов.

3. Влияние выбора методологии на разрабатываемую ADAS– систему

3.1. Определение аспектов процессов разработки и эволюции системы.

Если рассмотреть систему ADAS как систему с фиксированным (монолитным) аппаратным ядром, модифицируемым программным ядром и атомарными периферийными компонентами, можно выделить два этапа развития системы – интенсивной разработки с реализацией/модификацией ядра и экстенсивной эволюции (с перестроением связей между компонентами и/или изменением состава периферийных компонентов). Для выбора оптимальной методологии разработки ADAS–систем необходимо провести сравнение влияния рассмотренных методологий на различные аспекты как процесса разработки системы, так и процесса её эволюции. В первую очередь необходимо определить аспекты, по которым будет производиться сравнение.

Необходимость корректного определения требований при проектировании сложных систем рассматривается в [17, 18]. Описание композиционного морфизма систем (возможности системы по изменению своей внутренней структуры) и его влияния на функциональный и эволюционный потенциал системы приводится в [1, 19]. Также в [19] как важные параметры систем рассматриваются стоимость системы (включающая стоимость разработки, производства и обслуживания), стабильность. Аспекты эволюции системы рассматриваются в статье [20], выделяют следующие критерии эволюционной способности системы:

- 1) общность системы – заведомое соответствие системы не только поставленным требованиям, но и потенциально возможным;
- 2) адаптируемость – способность системы к адаптации к изменяющимся требованиям посредством перестроения связей между компонентами (при неизменности состава системы);

3) масштабируемость – возможность вертикального масштабирования системы (увеличения производительности используемых компонентов) при увеличении нагрузки;

4) расширяемость – возможность адаптации системы к изменившимся требованиям посредством добавления новых компонентов, не входивших ранее в состав системы.

В работе [9] авторы выделяют два пути эволюционного развития системы – статический (без изменения структуры системы, адаптация только посредством заложенной избыточности) и динамический (изменение структуры, добавление либо замена компонентов) на основе рассмотренных в [20] аспектов. К статическому пути авторы относят общность систем, к динамическому – адаптируемость, масштабируемость и расширяемость. Вполне очевидно, что возможность статической эволюции системы всецело зависит от целеполагания и не может рассматриваться в контексте методологий проектирования. В связи с этим рассмотрим только аспекты динамической эволюции систем.

3.2. Влияние выбора методологии на различные аспекты процессов разработки и эволюции системы

3.2.1. Постановка задачи и определение требований. Использование методологии сверху–вниз является неизбежным при постановке задачи на начальной стадии проектирования (утверждения эскизного проекта) любых сложных технических систем – корректная формулировка начального технического задания является важнейшим этапом проектирования.

3.2.2. Композиционный морфизм. Композиционный морфизм возможно рассматривать как расширение фактора адаптируемости системы – адаптируемость подразумевает возможность модификации системы под влиянием внешних факторов, тогда как композиционный морфизм характеризует структурную гибкость системы в целом.

Подход снизу–вверх по своей концепции позволяет получить систему, состоящую из набора ортогональных (независимых и непересекающихся по функциональности) компонентов, и функциональность системы определяется системой связей между компонентами. При этом выбор низкоуровневых ортогональных компонентов в методологии снизу–вверх производится на ранних этапах разработки, и решение задачи адаптации системы для соответствия требованиям производится в большинстве случаев посредством корректировки связей между ними. Следовательно, системы, разработанные в соответствии с подходом снизу–вверх, по определению полиморфны – при неизменности их состава (набора подсистем и компонентов) такие системы обладают большим многообразием внутренних структур (систем связей). Очевидно, что подобная внутренняя гибкость структуры системы повышает адаптационные и эволюционные возможности системы. Согласно [9], высокая способность системы к адаптации посредством модификации структуры связей реализуема при использовании модульной архитектуры, открытых (непроприетарных) стандартов и стандартных интерфейсов, что соответствует методологии снизу–вверх.

В соответствии с методологией сверху–вниз системы разрабатываются с ориентацией на требования и сценарии использования [1], и фиксированная система связей (единственная, наиболее оптимально соответствующая изначально заданному набору требований) является равноправным и неотъемлемым компонентом системы. Системы такого рода мономорфны – возможен только один способ объединения подсистем, причём зачастую связи описываются раньше формализации

требований к функциональности подсистем при декомпозиции. Эволюция таких систем происходит при изменении требований, что приводит к проведению полного цикла проектирования заново – в результате формируется новая система, также имеющая монолитную структуру.

Для систем ADAS мономорфная структура слабо применима, т.к. активное развитие сферы требует постоянной эволюции разрабатываемых систем.

3.2.3. Масштабируемость и расширяемость. Помимо эволюции с сохранением состава системы и редактированием системы связей, в процессе развития системы может возникнуть потребность в замене каких-либо периферийных компонентов (например, при выходе на рынок устройств с улучшенными характеристиками). Движущим фактором эволюции такого рода являются подсистемы и компоненты, а конкретнее – их развитие (независимое от ADAS-системы, в которой они используются). Так как независимые подсистемы могут существовать только в рамках методологии снизу-вверх, то обеспечить масштабируемость ADAS-систем возможно только с использованием этого подхода.

Согласно определению из работы [20] расширяемость представляет собой модификацию системы путём добавления новых компонентов, что подразумевает, аналогично масштабируемости, развитие системы, движимое компонентами, и может быть реализовано только с применением методологии снизу-вверх. В контексте систем ADAS может быть вызвано выходом на рынок нового типа сенсоров или новых программных решений.

Масштабирование и расширение системы в рамках подхода сверху-вниз выражается в формировании обновлённых требований к системе и полной переработке системы по полному циклу проектирования, что подразумевает глобальные изменения в системе, так как подсистемы в концепции методологии сверху-вниз не являются независимыми и тесно связаны с другими компонентами. Изменение характеристик отдельных подсистем, не контролируемое целеполаганием ко всей системе в целом, может оказать непредсказуемое влияние на внутрисистемные взаимодействия и функционирование системы [8, 21].

3.2.4. Стабильность, устойчивость. Стабильность системы здесь корректно рассматривать с точки зрения необходимости внесения изменений. Требования к системам меняются, особенно это актуально для ADAS-систем. В условиях активного развития их программной и аппаратной базы, когда растущая функциональность потенциальных компонентов позволяет постоянно повышать требования к системе, системы, спроектированные по методологии сверху-вниз, обладает меньшим адаптационным ресурсом, а вносимые изменения более деструктивны для работающей системы. В соответствии с подходом снизу-вверх системы разрабатываются с учётом необходимости корректировки реализации системы с целью удовлетворения требований и обладают достаточной гибкостью внутренней структуры для неструктурной модификации, в том числе и в процессе эволюции системы. Таким образом, при использовании подхода снизу-вверх можно добиться более устойчивого функционирования системы в условиях изменяющихся требований, а подход сверху-вниз позволяет получить высокостабильную систему, но лишь при условии, что требования будут оставаться неизменными – в этом случае ни о какой эволюции системы говорить не приходится.

3.2.5. Стоимость и время разработки. В случае разработки достаточно сложных систем подход снизу-вверх обладает значительными преимуществами в части стоимости и времени разработки, так как позволяет (и даже подразумевает) использование сторонних и разработанных ранее компонентов. Как было показано ранее, разработка ADAS-систем невозможна без использования сторонних компонентов в виду их высокой сложности, соответственно применение методологии

сверху–вниз в ADAS–системах с точки зрения стоимости и времени разработки нецелесообразно. При этом процесс разработки в методологии снизу–вверх итеративен, так как подразумевает поэтапную адаптацию получаемого результата для соответствия требованиям, тогда как подход снизу–вверх предполагает полное проектирование системы, а затем полную её реализацию оптимальным способом. Таким образом, если рассматривать стоимость разработки не с практической, а с методологической точки зрения, предпочтительнее использование методологии сверху–вниз.

Однако, вышесказанное правомерно только в случае оптимального течения процесса разработки. Отклонения неизбежны при проектировании системы по обеим рассматриваемым методологиям. В методологии снизу–вверх невозможность конечной реализации запланированной функциональности, как было показано, проявляется на поздних этапах разработки. Также нельзя исключать вероятность неточного/некорректного определения изначальных требований [5], что может быть обнаружено при тестировании на поздних этапах разработки или даже конечными пользователями при использовании уже готовой системы. В обоих случаях процесс переработки системы потребует значительных финансовых и временных затрат, сопоставимых с разработкой системы заново. При использовании методологии снизу–вверх неизбежно несоответствие системы требованиям, но гибкость структуры системы и возможность частых итеративных уточнений позволяет адаптировать систему с минимальными издержками.

3.2.6. Способность к развитию и процесс развития (эволюции). По рассмотренным выше аспектам развития системы можно сделать обобщённые выводы по влиянию методологий сверху–вниз и снизу–вверх на эволюционный потенциал разрабатываемой системы. Для процесса эволюции ADAS–систем характерны постоянные, но незначительные по масштабам системы в целом обновления существующих подсистем и периодическое внедрение новых типов периферийных компонентов (как программных, так и аппаратных). В рамках подхода снизу–вверх изменения такого рода могут быть произведены с минимальными временными и финансовыми затратами при условии либо: 1) сохранения обратной совместимости интерфейсов заменяемых/обновляемых компонентов; либо 2) добавления компонентов в пределах заложенной коммуникационной и вычислительной избыточности монолитного ядра. В противном случае эволюция системы становится интенсивной и требуется обновление ядра. В случае же подхода сверху–вниз система по определению оптимальна и максимально соответствует поставленным требованиям, следовательно, в такой системе не заложена избыточность для масштабирования. Помимо этого, в таких системах эволюция может происходить только исходя из изменений требований, что приводит к необходимости перепроектирования системы. Сценарий покомпонентного развития функциональности ADAS–системы в рамках подхода сверху–вниз невозможен.

3.3. Результаты. Сравнительная таблица. В табл. 1 обобщены выводы по влиянию выбора методологии на процесс разработки и эволюции системы. Здесь символ «+» показывает, что методология полностью удовлетворяет приведённому аспекту, «–» – совершенно не удовлетворяет, «+/-» – в большей степени удовлетворяет, «-/+» – в некоторой степени удовлетворяет. Как видно из результатов, существуют аспекты процесса развития системы, в которых наблюдается значительное преимущество только одной из методологий.

Таблица 1

Сравнение влияния методологий сверху–вниз и снизу–вверх на различные аспекты системы

	сверху–вниз	снизу–вверх
Постановка задачи, определение требований	+	–
Композиционный морфизм	–	+
Масштабируемость, расширяемость	–	+
Стабильность, устойчивость	–/+	+/-
Стоимость, время разработки	–/+	+/-
Способность к развитию (эволюции)	–	+

Выводы. В работе исследуются достоинства и недостатки подходов проектирования снизу–вверх и сверху–вниз применительно к системам ADAS. Рассматривается влияние выбора методологии проектирования на процессы как проектирования, так и эволюции системы.

Оба традиционных подхода проектирования (сверху–вниз и снизу–вверх) обладают недостатками, отклоняющими течение процесса разработки от оптимального направления при возникновении ошибок проектирования, которые в обоих случаях проявляются лишь на поздних этапах разработки: в частности, для методологии сверху–вниз возможно несоответствие требований техническим возможностям на низких уровнях; в методологии снизу–вверх точное соответствие требованиям практически невозможно. При этом существуют аспекты процессов разработки и эволюции системы, в которых наблюдается значительное преимущество каждой из методологий.

При проектировании систем автоматизированного управления транспортным средством требуется обеспечить масштабируемость и модифицируемость системы в процессе её эволюции. Особенно это актуально в настоящее время – наблюдается постоянный прогресс в развитии сенсорных систем, робототехники, алгоритмов автоматизации, и недостаточная гибкость требований и самих процессов разработки приводит к неизбежному и быстрому устареванию систем. В должной степени эволюция системы может быть обеспечена только при использовании подхода снизу–вверх. Однако, для сложных систем критически важным является определение изначальных требований к системе, что может быть достигнуто только с применением методологии сверху–вниз. Формирование обобщённого подхода к проектированию с учётом проведённого анализа планируется как продолжение исследований на тему.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Pizka M., Bauer A.* A brief top-down and bottom-up philosophy on software evolution, *Proceedings. 7th International Workshop on Principles of Software Evolution, 2004*, [S.l.]: IEEE, 2004, pp. 131-136.
2. *Li L., Wen D., Zheng N., Shen L.* Cognitive Cars: A New Frontier for ADAS Research, *IEEE Transactions on Intelligent Transportation Systems*, 2012, Vol. 13, No. 1, pp. 395-407. Doi: 10.1109/TITS.2011.2159493.
3. *Ziebinski A., Cupek R., Grzechca D., Chruszczyk L.* Review of advanced driver assistance systems (ADAS), *Proceedings of the International Conference of Computational Methods in Sciences and Engineering 2017 (ICCMSE-2017). AIP Conference Proceedings*, Vol. 1906, [S.l.]: [AIP], 2017, Art. no. 120002, 4 p. Doi: 10.1063/1.5012394.
4. *Jumaa B.A., Abdulhassan A.M., Abdulhassan A.M.* Advanced Driver Assistance System (ADAS): A Review of Systems and Technologies, *International Journal of Advanced Research in Computer Engineering & Technology*, 2019, Vol. 8, No. 6, pp. 231-234.
5. *Sangiovanni-Vincenelli A., Damm W., Passerone R.* Taming Dr. Frankenstein: Contract-based design for cyber-physical systems, *European journal of control*, 2012, Vol. 18, No. 3, pp. 217-238.

6. Crespi V., Galstyan A., Lerman K. Top-down vs bottom-up methodologies in multi-agent system design, *Autonomous Robots*, 2008, Vol. 24, No. 3, pp. 303-313.
7. Wan-Kadir W.M.N., Loucopoulos P. Relating evolving business rules to software design, *Journal of Systems architecture*, 2004, Vol. 50, No. 7, pp. 367-382.
8. Borches P.D., Bonnema G.M. Coping with System Evolution-Experiences in Reverse Architecting as a Means to Ease the Evolution of Complex Systems, *INCOSE International Symposium*, 2009, Vol. 19, No. 1, pp. 955-969.
9. Christian J., Olds J. A quantitative methodology for identifying evolvable space systems, *1st Space Exploration Conference: Continuing the Voyage of Discovery*. [S.l.]: [s.n], 2005, pp. 2543.
10. Bernhart W., Winterhoff M. Autonomous Driving: Disruptive Innovation that Promises to Change the Automotive Industry as We Know It, *Energy Consumption and Autonomous Driving*. Cham: Springer, 2016, pp. 3-10.
11. Jo K. et al. Development of autonomous car – Part I: Distributed system architecture and development process, *IEEE Transactions on Industrial Electronics*, 2014, Vol. 61, No. 12, pp. 7131-7140.
12. Heinecke H. Automotive system design-challenges and potential, *Proceedings of the conference on Design, Automation and Test in Europe-Volume 1*. [S.l.]: IEEE Computer Society, 2005, pp. 656-657.
13. Benvenuti L. et al. Contract-based design for computation and verification of a closed-loop hybrid system, *International Workshop on Hybrid Systems: Computation and Control*. Berlin: Springer, 2008, pp. 58-71.
14. Sangiovanni-Vincentelli A. Quo vadis, SLD? Reasoning about the trends and challenges of system level design, *Proceedings of the IEEE*, 2007, Vol. 95, No. 3, pp. 467-506.
15. Thakurdesai H.M., Aghav J.V. Autonomous cars: technical challenges and a solution to blind spot, *Advances in Computational Intelligence and Communication Technology. Proceedings of CICT 2019*. Advances in Intelligent Systems and Computing, Vol. 1086. Singapore: Springer, 2021, pp. 533-547. Doi: 10.1007/978-981-15-1275-9_44.
16. Ziebinski A., Cupek R., Erdogan H., Waechter S. A Survey of ADAS Technologies for the Future Perspective of Sensor Fusion, *Computational Collective Intelligence. ICCCI 2016. Lecture Notes in Computer Science*, Vol. 9876. Cham: Springer, 2016, pp. 135-146. Doi: 10.1007/978-3-319-45246-3_13.
17. Boehm B.W. Verifying and validating software requirements and design specifications, *IEEE software*, 1984, Vol. 1, No. 1, pp. 75.
18. Arnaut B.M., Ferrari D.B., e Souza M.L.O. A requirements engineering and management process in concept phase of complex systems, *2016 IEEE International Symposium on Systems Engineering (ISSE)*. [S.l.]: IEEE, 2016, pp. 1-6.
19. Sztipanovits J. et al. Toward a science of cyber-physical system integration, *Proceedings of the IEEE*, 2012, Vol. 100, No. 1, pp. 29-44.
20. Rowe D., Leaney J., Lowe D. Defining systems evolvability-a taxonomy of change, *Change*, 1994, Vol. 94, pp. 541-545.
21. Watson J.D. et al. Optimization of excess system capability for increased evolvability, *Structural and Multidisciplinary Optimization*, 2016, Vol. 53, No. 6, pp. 1277-1294.

Статью рекомендовал к опубликованию к.т.н. П.А. Кокунин.

Чикрин Дмитрий Евгеньевич – Казанский федеральный университет; e-mail: dmitry.kfu@gmail.com; г. Казань, Россия; тел.: +79172727100; к.т.н.; доцент; директор Института вычислительной математики и информационных технологий.

Егорчев Антон Александрович – e-mail: eanton090@gmail.com; тел.: +78432337609; научный сотрудник Института фундаментальной медицины и биологии.

Chickrin Dmitry Evgenevich – Kazan Federal University; e-mail: dmitry.kfu@gmail.com; Kazan, Russia; phone: +79172727100; cand. of eng. sc.; associate professor; director of the Institute of Computer Mathematics and Information Technologies.

Egorchev Anton Aleksandrovich – e-mail: eanton090@gmail.com; phone: +78432337609; research fellow of the Institute of Fundamental Medicine and Biology.

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 объемом **не менее 250-300 слов**. В тексте аннотации указывается цель, задачи исследования и краткие выводы. В аннотации **не следует** давать ссылку 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 взимается.

Адрес журнала в Интернете: <http://izv-tn.tti.sfedu.ru/>.