

Программные продукты и системы

DOI 10.15827/0236-235X

/ Ускорение расчета
критериальной
функции

/ Об алгоритме
поиска кратчайших
путей ГН-графа

/ Хранение данных
функционирования
объектов
киберфизических
систем

/ О создании
прототипов
баз знаний



№ 3

ТОМ 37

2024

Научно-исследовательский институт
«Центрпрограммсистем»

Программные продукты и системы

НАУЧНЫЙ ЖУРНАЛ

2024, том 37, № 3
(год издания тридцать седьмой)

Главный редактор
Г.И. САВИН, академик РАН

SOFTWARE & SYSTEMS

Research Journal

2024, vol. 37, no. 3

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

Research Institute CENTERPROGRAMSYSTEM

© ПРОГРАММНЫЕ ПРОДУКТЫ И СИСТЕМЫ

Научный журнал

2024. Т. 37. № 3

DOI: 10.15827/0236-235X.142

Главный редактор

Г.И. САВИН, академик РАН

Научные редакторы номера:

Н.А. СЕМЕНОВ, д.т.н., профессор

Т.М. ТАТАРНИКОВА, д.т.н., профессор

Издатель НИИ «Центрпрограммсистем»

(г. Тверь, Россия)

Учредитель **В.П. Куприянов**

Журнал зарегистрирован в Роскомнадзоре

3 марта 2020 г.

Регистрационное свидетельство ПИ № ФС 77-77843

Подписной индекс в каталоге

Урал-Пресс **70799**

ISSN **0236-235X** (печатн.)

ISSN **2311-2735** (онлайн)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

АССОЦИИРОВАННЫЕ ЧЛЕНЫ РЕДАКЦИИ

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

Технологический институт Южного федерального университета, г. Таганрог, Россия

Тверской государственный технический университет, г. Тверь, Россия

АДРЕС ИЗДАТЕЛЯ И РЕДАКЦИИ

Россия, 170024,

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

Телефон (482-2) 39-91-49

Факс (482-2) 39-91-00

E-mail: red@cps.tver.ru

Сайт: www.swsys.ru

Дата выхода в свет 16.09.2024 г.

Отпечатано ИПП «Фактор и К»

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

Выпускается один раз в квартал

Год издания тридцать седьмой

Формат 60×84 1/8. Объем 160 стр.

Заказ № 7. Тираж 1000 экз. Цена 550,00 руб.

© SOFTWARE & SYSTEMS

Research Journal

2024, vol. 37, no. 3

DOI: 10.15827/0236-235X.142

Editor-in-chief

G.I. SAVIN, Academician of RAS

Science editors of the issue:

N.A. SEMENOV, Dr.Sc. (Engineering), Professor

T.M. TATARNIKOVA, Dr.Sc. (Engineering),
Professor

Publisher Research Institute
CENTERPROGRAMSYSTEM (Tver, Russian Federation)

Founder V.P. Kupriyanov

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

Registration certificate ПИИ № ФС 77-77843

ISSN 0236-235X (print)

ISSN 2311-2735 (online)

EDITORIAL BOARD

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ASSOCIATED EDITORIAL BOARD MEMBERS

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

EDITORIAL BOARD AND PUBLISHER OFFICE ADDRESS

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

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

E-mail: red@cps.tver.ru

Website: www.swsys.ru

Release date 16.09.2024

Printed in printing-office "Faktor i K"

Krylova St. 26, Tver, 170100, Russian Federation

Published quarterly. 37th year of publication

Format 60×84 1/8. Wordage 160 pages

Prod. order № 7. Circulation 1000 copies. Price 550,00 rub.

Вниманию авторов

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

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

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

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

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

Условия публикации

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

1.2. Компьютерные науки и информатика

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

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

2.3. Информационные технологии и телекоммуникации

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

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

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

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

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

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

2.3.8. Информатика и информационные процессы (технические науки).

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

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

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

Порядок рецензирования

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

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

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

Оптимизация процессов интеллектуального управления в реальном времени на физической модели робота-манипулятора с помощью генетического алгоритма

М.С. Катулин¹✉, А.Г. Решетников¹, А.Р. Рябов¹, С.В. Ульянов^{1,2}

¹ Объединенный институт ядерных исследований,

Лаборатория информационных технологий им. М.Г. Мещерякова, г. Дубна, 141980, Россия

² Университет «Дубна», Институт системного анализа и управления, г. Дубна, 141980, Россия

Ссылка для цитирования

Катулин М.С., Решетников А.Г., Рябов А.Р., Ульянов С.В. Оптимизация процессов интеллектуального управления в реальном времени на физической модели робота-манипулятора с помощью генетического алгоритма // Программные продукты и системы. 2024. Т. 37. № 3. С. 301–309. doi: 10.15827/0236-235X.142.301-309

Информация о статье

Группа специальностей ВАК: 1.2.1

Поступила в редакцию: 18.10.2023

После доработки: 23.03.2024

Принята к публикации: 02.04.2024

Аннотация. При разработке интеллектуальной системы управления возникает необходимость подбора оптимальных параметров, при которых эта система будет достигать цели управления. Для подбора параметров используют математическую модель или аналогичные методы. В данной работе рассматривается проблема настройки системы интеллектуального управления с недоопределенными параметрами. В качестве примера использована система управления четырехзвенным роботом-манипулятором в контуре обратной связи на основе системы машинного зрения. Наглядно продемонстрирована работа системы управления, для которой характерно наличие внешних факторов, влияющих на ее поведение и не всегда подлежащих учету. В рассматриваемом примере это люфты приводов, ошибки/неточности работы системы машинного зрения или внезапное изменение положения цели. В работе подробно описаны устройство робота-манипулятора, система машинного зрения и используемые регуляторы. На первом шаге настройки системы искомые параметры подобраны эмпирически, затем выполнена оптимизация с применением генетического алгоритма. Представлен способ применения генетического алгоритма для оптимизации параметров системы управления на физическом объекте в реальном времени. Рассмотрены алгоритмические особенности применяемого генетического алгоритма, в том числе реализации функции пригодности. Приведены результаты оптимизации и сделан сравнительный анализ работы системы с разными параметрами. Продемонстрирована возможность реализации подобных систем и применения генетических алгоритмов в реальном времени. Представлены методы технологии проектирования интеллектуальных систем, которые могут быть перенесены и на другие объекты управления.

Ключевые слова: генетический алгоритм, манипулятор, машинное зрение, глобальная отрицательная обратная связь

Введение. Решение задачи интеллектуального управления подразумевает наличие свойств адаптации и обучения в системе автоматического управления, например, подбор определенных параметров регуляторов, описывающих оптимальное поведение системы в условиях внешних и внутренних возмущений. Проверка и прогнозирование поведения системы в таких условиях осуществляются с помощью математической модели, в том числе в системе виртуальной реальности [1]. Однако поиск оптимальных решений усложняется внешними факторами, влияющими на поведение системы, например, температурой окружающей среды [2]. Существуют примеры систем, поведение которых зависит и от свойств объекта, с которым они взаимодействуют, как в случае летательного манипулятора, открывающего дверь [3]. Такие внешние факторы трудно учитывать в математической модели, и они

остаются в большей мере недоопределенными. Для обучения системы автоматического управления, содержащей недоопределенные (не учитываемые в математической модели объекта управления) параметры, нужен новый подход, позволяющий получать результат без применения точной математической модели.

Обзор существующих подходов

Применение теории интеллектуального управления рассмотрим на примере физической модели робота-манипулятора. В этой модели возмущениями выступают, в частности, износ деталей, их брак или заводские люфты приводов и редукторов, трения в местах крепления осей вращения робота или дополнительная нагрузка, которой оперирует робот [4]. Данные особенности объекта управления вносят множество недоопределенных параметров,

влияющих на поведение системы управления. Задачи адаптации и обучения для подобных систем могут быть решены путем подбора значений коэффициентов усиления традиционных ПИД-регуляторов в обратной связи, например, с помощью моделирования поведения манипулятора в среде Matlab Simulink [5]. Подбор оптимальных коэффициентов при этом может быть выполнен с помощью технологии интеллектуальных вычислений и *генетического алгоритма* (ГА) [6]. Главный недостаток данного подхода состоит в необходимости иметь математическую модель для настройки параметров системы управления.

Другой подход к управлению роботом-манипулятором состоит в применении алгоритма инверсной (обратной) кинематики [7]. Обратная кинематика позволяет работать с математическими и компьютерными моделями объектов управления, что при должной верификации с реальным объектом позволяет извлекать обучающие траектории системы с различными начальными условиями. Однако такой подход имеет и свои ограничения, в частности, не учитываются физические особенности системы, канала обратной связи, зашумленность и точность позиционирования объекта управления и др.

Принцип глобальной отрицательной обратной связи

Задача управления объектом, как правило, подразумевает перевод объекта управления в целевое положение в фазовом пространстве состояний. Чтобы определить силу управляющего воздействия, используют обратную связь. Обратная связь может быть и положительной, когда изменение управляющего воздействия совпадает по знаку с изменением выходного сигнала, что, как правило, приводит к экспоненциальному росту выходного сигнала, и отрицательной – в этом случае система стремится к стабильному целевому значению.

Целевые значения достигаются с помощью разных типов регуляторов. Широкое распространение получили *пропорциональные* (П), *пропорциональные интегральные* (ПИ), *пропорциональные дифференциальные* (ПД) и *пропорциональные интегрально-дифференциальные* (ПИД) регуляторы [8]. Интегральная составляющая регулятора позволяет устранить в системе статическую ошибку, дифференциальная – улучшить динамические показатели, форсируя переходный процесс, а пропорциональная формирует точность позиционного управления.

В качестве обратной связи используются показания датчиков. В данной работе сенсорная часть была реализована средствами машинного зрения. На изображении, полученном стандартной камерой в 2D-формате, определяются отклонения текущей позиции манипулятора от целевых координат захватываемого объекта.

Описание аппаратной реализации объекта управления

На робототехническом полигоне Лаборатории информационных технологий им. М.Г. Мещерякова Объединенного института ядерных исследований (г. Дубна) была поставлена задача разработать систему с наглядной демонстрацией управления объектом со множеством недоопределенных параметров. В качестве объекта управления выбран стационарный четырехосный робот-манипулятор. Манипулятор собран из общедоступных компонентов и имеет множество люфтов (рис. 1).

Задача манипулятора – захватить заданный объект, произвольно расположенный перед ним, при этом захват объекта осуществляется по принципу глобальной отрицательной обратной связи. Для определения положения цели и, соответственно, вычисления ошибки используется простейший детектор на основе веб-камеры и методов машинного зрения.

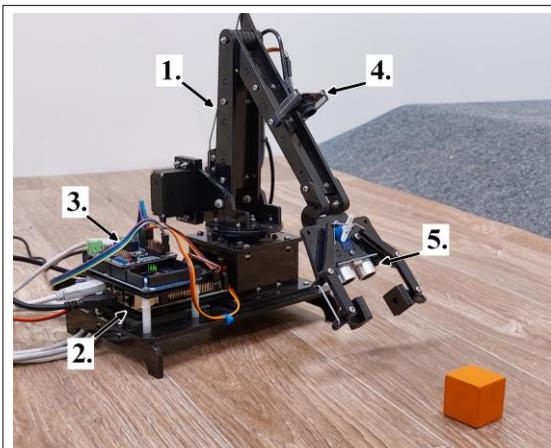


Рис. 1. Общий вид манипулятора с обратной связью: 1 – четырехосный манипулятор с захватом; 2 – одноплатный компьютер; 3 – контроллер на базе Arduino Mega; 4 – USB-камера; 5 – сонар

Fig. 1. General layout of the feedback manipulator: 1 – 4-axis manipulator with a grip; 2 – single-board computer; 3 – Arduino Mega controller; 4 – USB camera; 5 – sonar

Управление манипулятором осуществляется с помощью одноплатного компьютера Raspberry PI4, который взаимодействует с контроллером манипулятора по COM-порту. Манипулятор входит в состав экспериментального полигона и управляется по локальной сети с помощью REST API [9]. Видеосигнал поступает с web-камеры, подключенной непосредственно к одноплатному компьютеру. Сонар подключен к контроллеру манипулятора. Анализ кадров, полученных с USB-камеры, выполняется полностью на одноплатном компьютере. Для этих целей был разработан простейший алгоритм обнаружения, реализованный с использованием библиотеки OpenCV. Алгоритм обнаружения представляет собой следующую последовательность шагов:

- задание диапазонов допустимых значений цвета по трем каналам;
- получение изображения с камеры;
- предобработка изображения: перевод его в цветовое пространство HSV и размытие фильтром Гаусса с радиусом 11;
- отбрасывание пикселей, значения которых не подходят под заданные допустимые значения цветового канала;
- выделение контуров на основе оставшихся пикселей;
- выбор контура с максимальной площадью;
- если площадь выбранного контура больше 100 (защита от ложного срабатывания), то возврат прямоугольника, обрамляющего контур, в противном случае возврат прямоугольника нулевого размера.

Как уже упоминалось, рассматриваемый манипулятор имеет четыре сервопривода: первый расположен в основании робота, последний в захвате. Минимальные и максимальные значения заложены в виде констант в программе контроллера, а стартовые хранятся в энергонезависимой памяти.

Процесс захвата объекта происходит итеративно, с помощью машинного зрения манипулятор определяет положение объекта и перемещается в его сторону. Процесс останавливается на основании показаний сонара, когда информация о размере и положении искомого объекта подтверждает, что он находится в зоне работы захвата.

Этапы решения задачи

Вычисление смещения манипулятора.
Для вычисления обновленных углов приводов

манипулятора необходимо рассчитать, насколько нужно переместить захват манипулятора с камерой, чтобы объект оказался в рабочей зоне захвата. Есть несколько путей для достижения этой цели.

Один из них – использовать знание геометрии манипулятора, измерить положение камеры, исправить дисторсию объектива камеры и вычислить коэффициенты пересчета координат из детектора в физическое положение объекта относительно камеры. Однако, во-первых, это крайне трудоемкий процесс, поскольку измерения и калибровка требуют больших временных затрат и уникальны для конкретного устройства. Во-вторых, качество деталей манипулятора очень низкое, учесть все неточности печати и сборки крайне затруднительно. В-третьих, поворот сервоприводов осуществляется неточно с люфтом местами до трех градусов и с шагом порядка одного градуса, что значительно сказывается на точности наведения.

Другой путь заключается в использовании итеративного процесса корректировки положения осей манипулятора с маленьким шагом таким образом, чтобы система постепенно сходилась к положению, когда объект находится в зоне работы захвата. Для этого нужно сменить стандартную систему координат изображения, при которой центр координат находится в верхнем левом углу, на систему координат с центром, совпадающим с целевым положением объекта в кадре. После данного преобразования координат положение объекта можно представить как ошибку, которую требуется минимизировать. При успешном решении этой задачи камера будет наведена точно на объект. Но для захвата нужно не просто повернуться в сторону объекта, но и приблизиться к нему. В качестве характеристики расстояния до объекта была использована разница между целевой шириной объекта в кадре и шириной объекта из результатов работы детектора.

Таким образом, была получена некая характеристика, которую также нужно минимизировать.

За движения робота по направлению к объекту отвечают три сервопривода, при этом на ошибку положения по осям Y и Z влияют сразу два из них. В первом приближении задача вычисления обновленного положения на каждом шаге может быть решена методом подбора коэффициентов для системы уравнений:

$$p_1 = p_1 + w_{11}err_x + w_{12}err_y + w_{13}err_z,$$

$$p_2 = p_2 + w_{21}err_x + w_{22}err_y + w_{23}err_z,$$

$$p_3 = p_3 + w_{31}err_x + w_{32}err_y + w_{33}err_z,$$

где p_i – значение угла i -го сервопривода; err_x , err_y – ошибки положения центра объекта; err_z – характеристика расстояния до объекта; w_{ij} – веса влияния соответствующей ошибки на положение сервопривода.

Первый привод выполняет поворот манипулятора относительно основания и не может влиять на величины err_y и err_z . По этой причине w_{12} и w_{13} можно приравнять к нулю и опустить, как и пренебречь членами w_{21} и w_{31} .

Поскольку шаг работы сервоприводов довольно большой, а конструкция манипулятора имеет множество люфтов, такой подход работает только в случае небольших значений весов, а это значит, что робот будет достигать цели достаточно медленно. Исправить это и сгладить формы движения можно с помощью ПИД-регулятора. Чтобы не перегружать уравнения дополнительными коэффициентами, можно использовать ПД-регулятор, который показал свою эффективность при стабилизации подобных процессов [10, 11].

Эмпирически было установлено, что применение ПД-регуляторов вместо пяти весовых коэффициентов приводит к непредсказуемым резким движениям манипулятора при приближении к искомому объекту. Предположительно, при некоторых положениях объекта два ПД-регулятора входят в резонанс, что приводит к экстремально большому шагу сервоприводов, и робот теряет объект из поля зрения телекамеры.

Чтобы ПД-регуляторы не оказывали нежелательного воздействия друг на друга, их количество должно соответствовать количеству управляемых сервоприводов. В представленной модели ошибки err_y и err_z имеют очевидную связь при влиянии на сервоприводы 2 и 3. Влияние данных ошибок на каждый привод необязательно равнозначное, поэтому справедливо будет ввести весовые коэффициенты влияния каждой ошибки на соответствующий привод. Пусть нумерация приводов начинается с 1, при этом первый привод – карусель робота. Тогда для всех приводов можно будет вычислить ошибки по следующим уравнениям:

$$\begin{aligned} err_1 &= err_x, \\ err_2 &= aerr_y + berr_z, \\ err_3 &= cerr_y + derr_z, \end{aligned}$$

или

$$\begin{aligned} err_1 &= err_x, \\ err_2 &= a(err_y + b/a err_z), \\ err_3 &= c(err_y + d/c err_z), \end{aligned}$$

где a , b , c и d – пропорциональные коэффициенты линейной зависимости ошибок err_y и err_z .

Формула ПД-регулятора:

$$p_i = p_i + P_i + D_i,$$

где i – номер сервопривода; p_i – значение угла сервопривода; P_i и D_i – пропорциональная и дифференциальная части ПД-регулятора сервопривода.

Вычисление пропорциональной части ПД-регулятора первого сервопривода осуществляется по формуле

$$P_1 = k_{p1}err_1 = k_{p1}err_x,$$

где k_{p1} – пропорциональный коэффициент для ПД-регулятора первого сервопривода.

Пропорциональная часть второго и третьего приводов вычисляется аналогично:

$$P_2 = k_{p2}err_2 = k_{p2}a(err_y + b/a err_z) = k_{p2}'(err_y + a'err_z),$$

$$P_3 = k_{p3}err_3 = k_{p3}c(err_y + d/c err_z) = k_{p3}'(err_y + c'err_z),$$

где $k_{p2}' = k_{p2}a$; $a' = b/a$; $k_{p3}' = k_{p3}c$; $c' = d/c$.

Дифференциальная составляющая для первого сервопривода:

$$D_1 = k_{d1}(err_1 - err_1^{(t-1)}) = k_{d1}(err_x - err_x^{(t-1)}),$$

где err_1 – ошибка положения первого сервопривода; $err_1^{(t-1)}$ – ошибка положения сервопривода на предыдущем шаге; k_{d1} – дифференциальный коэффициент для регулятора первого сервопривода.

Дифференциальная часть второго и третьего сервоприводов включает две компоненты ошибки. Для второго сервопривода это будет выглядеть так:

$$D_2 = k_{d2}(err_2 - err_2^{(t-1)}) = k_{d2}(a(err_y + b/a err_z) - a(err_y^{(t-1)} + b/a err_z^{(t-1)})).$$

После раскрытия скобок:

$$D_2 = k_{d2}a(err_y - err_y^{(t-1)} + b/a(err_z - err_z^{(t-1)})) = k_{d2}'(err_y - err_y^{(t-1)} + a'(err_z - err_z^{(t-1)})),$$

где $k_{d2}' = k_{d2}a$; $a' = b/a$.

Для дифференциальной части третьего привода все аналогично и после повторения операций принимает вид

$$D_3 = k_{d3}'(err_y - err_y^{(t-1)} + c'(err_z - err_z^{(t-1)})),$$

где $k_{d3}' = k_{d3}c$; $c' = d/c$.

Таким образом, решение задачи управления сводится к подбору восьми коэффициентов: k_{p1} , p_{d1} – для первого привода, k_{p2}' , k_{d2}' и a' – для второго привода, k_{p3}' , k_{d3}' и c' – для третьего привода.

Применение ГА для оптимизации подбора параметров управления. Первоначально задача подбора параметров была решена путем эмпирического перебора. Удалось подобрать такие значения, при которых манипулятор стал стабильно достигать цели, но для этого ему требовалось большое число шагов (порядка 40). В качестве альтернативного способа подбора

оптимальных значений были рассмотрены эволюционные алгоритмы, основное место среди которых занимает ГА [6, 12].

Существует множество вариантов реализации ГА, и большинство из них оперируют бинарными генами. В рассматриваемой задаче параметры представляют собой числа с плавающей точкой, поэтому требовалось найти метод кодирования набора чисел с плавающей точкой в бинарные хромосомы. В качестве языка программирования использован Python3, стандартные методы битового кодирования в котором привязаны к количеству байт на число (1, 2, 4 или 8). Алгоритм позволяет производить поиск оптимальных значений в заданном интервале, прямое кодирование чисел в биты приведет как минимум к тому, что не все эти биты будут задействованы.

С другой стороны, можно представить интервал значений параметра как линейку, на которой будет столько делений, сколько мы сможем закодировать в выбранное число байт. Таким образом, было принято решение о преобразовании всех дробных чисел в целые, описывающие номер деления на этой виртуальной линейке:

$$i = \left\lfloor (v - \min) \frac{2^{8b}}{(\max - \min)} \right\rfloor,$$

где i – целое число, характеризующее значение v в заданном интервале $[\min, \max]$; b – число байт в i .

Обратное преобразование:

$$v = i \frac{(\max - \min)}{2^{8b}} + \min.$$

После данного преобразования получаем хромосому длиной, равной произведению числа параметров на b . Генерация случайной

особи представляет собой генерацию нужного количества случайных байт. Рекомендуется, чтобы мутация каждого гена производилась с вероятностью $1/8b$. Отбор сделали элитарным, чтобы не терять хорошие решения.

Сам алгоритм представляет собой последовательность перечисленных далее шагов.

1. Случайным образом генерируется нулевая популяция заданного размера.

2. Выполняется оценка каждой особи популяции.

3. Создается новая популяция, и в нее переносятся N лучших особей из предыдущей.

4. Случайным образом выбираются две разные особи из предыдущей популяции. Выбор осуществляется с помощью модуля случайного значения, полученного из нормального распределения, с математическим ожиданием в 0 и среднеквадратичным отклонением в 30 % от размера популяции.

5. С вероятностью P_1 выполняется двухточечный кроссинговер.

6. Выбранные ранее или полученные после скрещивания особи добавляются в новую популяцию.

7. Повторяются шаги 4–6 до тех пор, пока размер новой популяции не будет равен заданному размеру.

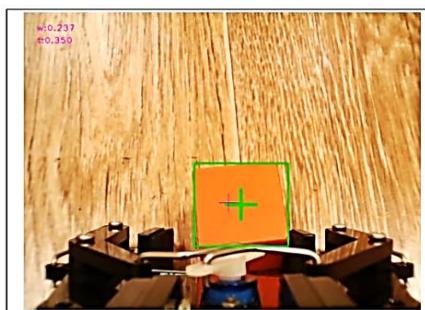
8. Для всех особей со 2-й с вероятностью P_2 применяется оператор мутации.

9. Выполняется оценка всех особей.

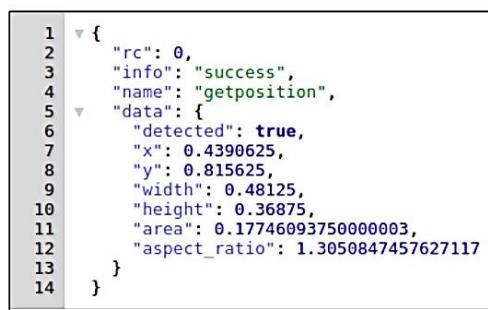
10. Повторяются шаги 3–9 столько раз, сколько поколений было задано при запуске.

11. Выбирается лучшая особь в качестве результата.

Подбор оптимальных параметров управления. В качестве цели были выставлены значения, полученные из машинного зрения (рис. 2).



а)



б)

Рис. 2. Результат работы машинного зрения: а) вид объекта из телекамеры; б) результат ответа на запрос к модулю машинного зрения

Fig. 2. Machine vision result: а) an object from a web-camera; б) the result of a query response to the machine vision module

Ошибки вычисляются следующим образом:

$$\text{err}_z = 0.5 - \text{width},$$

$$\text{err}_x = 0.44 - x,$$

$$\text{err}_y = 0.82 - y.$$

Для оценки качества коэффициентов в REST API манипулятора были добавлены новые Action и Service [9] с одинаковым именем checkcoeffs. Action принимает в качестве параметров восемь значений коэффициентов. Работа Action повторяет работу Action захвата объекта до момента захвата, то есть робот итеративно движется к объекту, но не зажимает его захватом, а возвращается в стартовое положение. При вычислении оценки была применена парадигма минимизации ошибки: чем меньше значение оценки, тем лучше. Для реализации этого условия во время работы программы проверки коэффициентов идет суммирование ошибки положения объекта по результатам машинного зрения. Оценка вычисляется по формуле $S = \sum_{i=2} d_i((\text{err}_{zi}) + (\text{err}_{yi}) + (\text{err}_{xi}))$, где d_i – число шагов с успешной детекцией объекта на i -м шаге процедуры. Первый шаг из оценки исключен, так как ошибка первого шага зависит от начального положения манипулятора и объекта и никак не зависит от проверяемых коэффициентов.

Если в какой-то момент объект выходит из поля зрения телекамеры, ошибка приравнивается к постоянному, заведомо большому значению. Объект может выйти из поля зрения телекамеры на первом же шаге, а может и через некоторое ненулевое количество шагов. Если придется выбирать из плохих решений, очевидно, что, чем больше шагов робот сделает до потери объекта из поля зрения, тем лучше. Поэтому финальное значение ошибки неудачного захвата объекта вычисляется как постоянное заведомо большое значение за вычетом числа шагов с успешно обнаруженным объектом. Также было принято решение ограничить число шагов до 100.

Проверка коэффициентов осуществляется полностью на стороне робота. Для получения значения оценки используется запрос информации из Service checkcoeffs. Ответ в поле data содержит число с плавающей точкой.

Для подбора параметров был запущен описанный выше ГА со следующими параметрами: размер популяции – 80, вероятность мутации особи – 0.1, вероятность скрещивания – 0.5, элитарная часть – 10 %, число поколений – 15.

Диапазоны параметров выбраны на основе значений, подобранных эмпирически на преды-

дущем этапе. В качестве размера интервала выступило удвоенное значение соответствующего параметра.

При каждом выполнении функции проверки (функции пригодности) производилась запись проверяемых коэффициентов и результата проверки для последующего анализа. Значения функции пригодности во времени представлены на рисунке 3 (важно отметить, что оценки непригодных решений (когда робот не дошел до объекта) на данном графике не отражены).

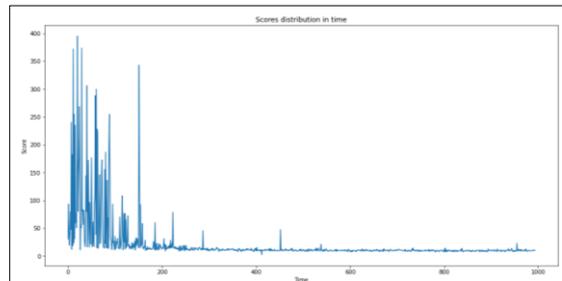


Рис. 3. Значения функции пригодности во времени

Fig. 3. Fitness function values in time

О том, что значения коэффициентов сходятся, можно судить по графикам изменения значений коэффициентов (<http://www.swsys.ru/uploaded/image/2024-3/1.jpg>).

Результаты экспериментов

Поскольку функция оценки качества выбора коэффициентов зависит от множества случайных факторов, было произведено по 30 запусков для коэффициентов, подобранных вручную, и коэффициентов, полученных с помощью ГА. Результаты распределения значений представлены на рисунках 4 и 5.

С новыми коэффициентами манипулятор стал быстрее достигать объекта. Для достижения цели при использовании эмпирически подобранных коэффициентов в среднем требуется около 40 шагов, при использовании коэффициентов ГА – около 10.

Анализ результатов

Проведенные эксперименты показали, что манипулятор достигает цели значительно быстрее, количество шагов, необходимых для захвата объекта, сократилось примерно в четыре раза. Значения отклонений от цели сходятся равномернее без резких скачков в сто-

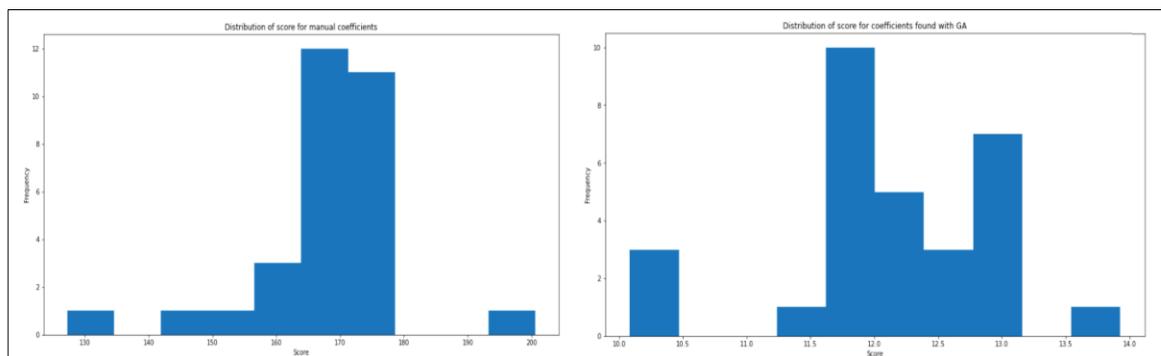


Рис. 4. Сравнение оценок коэффициентов, подобранных вручную (слева) и с помощью ГА (справа)

Fig. 4. Comparison of manually selected estimated coefficients (left) and using GA (right)

рону увеличения. Однако следует отметить, что работа манипулятора не является идеальной: при некоторых положениях объекта заметны довольно сильные колебания манипулятора относительно цели.

Выводы

На примере четырехосного робота-манипулятора была построена система управления объектом с множеством недоопределенных параметров, работающая по принципу глобальной отрицательной обратной связи. Параметры демонстрируемой системы подбирались с применением ГА на физическом объекте в режиме реального времени и без применения математической модели объекта управления. Таким образом, была наглядно продемонстрирована возможность реализации подобных систем и применения ГА в режиме реального времени. Представленные методы технологии проектирования интеллектуальных систем могут быть перенесены и на другие объекты управления.

В данном случае применение только ГА не позволило достичь полной робастности управления объектом, так как резкие и сильные из-

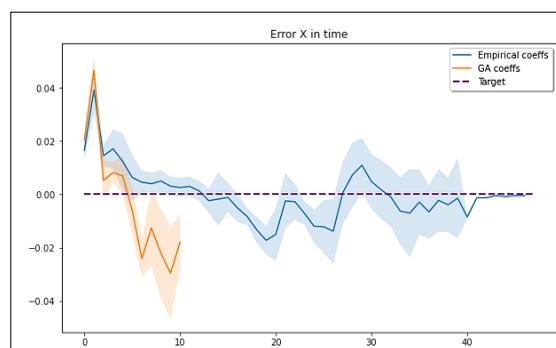


Рис. 5. Сравнение ошибки err_x для коэффициентов, подобранных вручную и с помощью ГА

Fig. 5. Comparison of err_x error for coefficients selected manually and using GA

менения окружающей среды могут привести к значительным колебаниям ошибок по мере приближения к цели. Создана платформа для дальнейших исследований и демонстрации возможностей алгоритмов. Следующим этапом может стать, например, применение нечетких и квантовых нейронных сетей для выбора коэффициентов управления.

Список литературы

1. Pascher M., Kronhardt K., Goldau F.F., Frese U., Gerken J. In time and space: Towards usable adaptive control for assistive robotic arms. ArXiv, 2023, art. 2307.02933. URL: <https://arxiv.org/pdf/2307.02933> (дата обращения: 21.03.2024).
2. Guist S., Schneider J., Ma H., Chen L. et al. A robust open-source tendon-driven robot arm for learning control of dynamic motions. ArXiv, 2023, art. 2307.02654. URL: <https://arxiv.org/pdf/2307.02654> (дата обращения: 21.03.2024).
3. Cuniato E., Geles I., Zhang W., Andersson O., Tognon M., Siegwart R. Learning to open doors with an aerial manipulator. ArXiv, 2023, art. 2307.15581. URL: <https://arxiv.org/abs/2307.15581> (дата обращения: 21.03.2024).
4. Параев Ю.И., Колесникова С.И., Цветницкая С.А. Управление роботом-манипулятором в условиях неопределенности // Вестн. ТГУ. Управление, вычислительная техника и информатика. 2021. № 57. С. 4–12. doi: 10.17223/19988605/57/1.

5. Лебедев С.К., Колганов А.Р. Модификация и настройка ПИД-регуляторов систем позиционирования промышленных манипуляторов // ЭиЭ: теория и практика: матер. III Всеросс. науч.-практич. конф. 2017. С. 1–13.
6. Рядчиков И.В., Гусев А.А., Сеченев С.И., Никульчев Е.В. Генетический алгоритм поиска параметров ПИД-регуляторов системы стабилизации шагающего робота // Тр. НГТУ им. Р.Е. Алексеева. 2019. № 1. С. 52–60.
7. Santos M.C., Molina L., Carvalho E.A.N., Freire E.O., Carvalho J.G.N., Santos P.C. FABRIK-R: An extension developed based on FABRIK for robotics manipulators. IEEE Access, 2021, vol. 9, pp. 53423–53435. doi: 10.1109/ACCESS.2021.3070693.
8. Кузнецов Е.А., Ульянов С.В. Разработка интеллектуальной системы управления многозвенным роботом-манипулятором // Системный анализ в науке и образовании. 2022. № 3. С. 161–179. URL: <http://sanse.ru/download/477> (дата обращения: 21.03.2024).
9. Катулин М.С., Кузнецов Е.А., Решетников А.Г., Рябов А.Р., Семашко С.В., Ульянов С.В. Механизмы обмена информацией и передачи знаний для задач взаимодействия в среде автономных интеллектуальных робототехнических систем в нештатных ситуациях // Системный анализ в науке и образовании. 2021. № 4. С. 44–62. URL: <http://sanse.ru/download/450> (дата обращения: 21.03.2024).
10. Бухтояров И.В., Кошкина И.Н. Создание ПД регулятора для квадрокоптера // StudNet. 2020. № 10. URL: <https://stud.net.ru/sozдание-pd-regulyatora-dlya-kvadrokoptera/> (дата обращения: 21.03.2024).
11. Кравченко П.П., Куликов Л.И., Щербинин В.В. Применение метода оптимизированных дельта-преобразований в задаче управления посадкой беспилотного летательного аппарата // Изв. РАН. Теория и системы управления. 2019. № 5. С. 130–144. doi: 10.1134/S0002338819050093.
12. Сайранов А.С., Касаткина Е.В., Нефедов Д.Г., Русяк И.Г. Применение генетических алгоритмов для управления организационными системами при возникновении нештатных ситуаций // Компьютерные исследования и моделирование. 2019. Т. 11. № 3. С. 533–556. doi: 10.20537/2076-7633-2019-11-3-533-556.

Software & Systems

doi: 10.15827/0236-235X.142.301-309

2024, 37(3), pp. 301–309

Physical model of a robot manipulator: Optimization of real-time intelligent control processes using a genetic algorithm

Mikhail S. Katulin ¹✉, Andrey G. Reshetnikov ¹, Andrey R. Ryabov ¹, Sergey V. Ulyanov ^{1,2}

¹ Joint Institute for Nuclear Research, M.G. Meshcheryakov Laboratory of Information Technologies, Dubna, 141980, Russian Federation

² Dubna University, Institute of System Analysis and Management, Dubna, 141980, Russian Federation

For citation

Katulin, M.S., Reshetnikov, A.G., Ryabov, A.R., Ulyanov, S.V. (2024) 'Physical model of a robot manipulator: Optimization of real-time intelligent control processes using a genetic algorithm', *Software & Systems*, 37(3), pp. 301–309 (in Russ.). doi: 10.15827/0236-235X.142.301-309

Article info

Received: 18.10.2023

After revision: 23.03.2024

Accepted: 02.04.2024

Abstract. When developing an intelligent control system, it becomes necessary to select the optimal parameters for the system to achieve a management goal. A selection of system parameters involves using a mathematical model or similar methods. This paper considers the problem of configuring an intelligent control system with undefined (not taken into account in the mathematical model) parameters. As an example of such system, the paper uses a control system for a four-link robot manipulator in a feedback loop based on a machine vision system. This example clearly demonstrates the operation of a control system characterized by external factors that affect system behavior and that are not always taken into account. In the example under consideration, the external factors are drive backlashes, errors/inaccuracies during the operation of a machine vision system, or a sudden change in a target position. The work describes in detail the robot manipulator design, a machine vision system and the controllers used. The first step of configuring the system involves empirical selection of the desired parameters, then the authors perform optimization using a genetic algorithm. The paper presents a method of applying a genetic algorithm to optimize the control system parameters on a physical object in real time. The paper considers algorithmic features of the applied genetic algorithm, including the features of implementing the fitness function. There are optimization results and a comparative analysis of the system operation with different parameters. The paper demonstrates the possibility of implementing such systems and using genetic algorithms in real time. The presented methods of intelligent systems design technology can be transferred to other control objects.

Keywords: genetic algorithm, manipulator, machine vision, global negative feedback

References

1. Pascher, M., Kronhardt, K., Goldau, F.F., Frese, U., Gerken, J. (2023) 'In time and space: Towards usable adaptive control for assistive robotic arms', *ArXiv*, art. 2307.02933, available at: <https://arxiv.org/pdf/2307.02933> (accessed March 21, 2024).
2. Guist, S., Schneider, J., Ma, H., Chen, L. et al. (2023) 'A robust open-source tendon-driven robot arm for learning control of dynamic motions', *ArXiv*, art. 2307.02654, available at: <https://arxiv.org/pdf/2307.02654> (accessed March 21, 2024).
3. Cuniato, E., Geles, I., Zhang, W., Andersson, O., Tognon, M., Siegwart, R. (2023) 'Learning to open doors with an aerial manipulator', *ArXiv*, art. 2307.15581, available at: <https://arxiv.org/abs/2307.15581> (accessed March 21, 2024).
4. Paraev, Yu.I., Kolesnikova, S.I., Tsvetnitskaya, S.A. (2021) 'Control of a robot manipulator in conditions of uncertainty', *Tomsk State University J. of Control and Comput. Sci.*, (57), pp. 4–12 (in Russ.). doi: 10.17223/19988605/57/1.
5. Lebedev, S.K., Kolganov, A.R. (2017) 'Modification and adjustment of PID controllers for positioning systems of industrial manipulators', *Proc. III All-Russ. Sci. and Pract. Conf Energy and Energy Conservation: Theory and Pract.*, pp. 1–13 (in Russ.).
6. Ryadchikov, I.V., Gusev, A.A., Sechenev, S.I., Nikulchev, E.V. (2019) 'A genetic algorithm for searching for parameters of PID regulators of a walking robot stabilization system', *Proc. NNSTU n.a. R.E. Alekseev*, (1), pp. 52–60 (in Russ.).
7. Santos, M.C., Molina, L., Carvalho, E.A.N., Freire, E.O., Carvalho, J.G.N., Santos, P.C. (2021) 'FABRIK-R: An extension developed based on FABRIK for robotics manipulators', *IEEE Access*, 9, pp. 53423–53435. doi: 10.1109/ACCESS.2021.3070693.
8. Kuznetsov, E.A., Ulyanov, S.V. (2022) 'Developing of intelligent control system for robotic manipulator with several freedom degrees', *System Analysis in Science and Education*, (3), pp. 161–179, available at: <http://sanse.ru/download/477> (accessed March 21, 2024) (in Russ.).
9. Katulin, M.S., Kuznetsov, E.A., Reshetnikov, A.G., Ryabov, A.R., Semashko, S.V., Ulyanov, S.V. (2021) 'Mechanisms of information exchange and knowledge transfer for interaction tasks in the hazard environment of autonomous intelligent robotic systems', *System Analysis in Science and Education*, (4), pp 44–62, available at: <http://sanse.ru/download/450> (accessed March 21, 2024) (in Russ.).
10. Bukhtoyarov, I.V., Koshkina, I.N. (2020) 'The creation of a PD controller for the quadcopter', *StudNet*, (10), available at: <http://stud.net.ru/sozdanie-pd-regulyatora-dlya-kvadrokoptera/> (accessed March 21, 2024) (in Russ.).
11. Kravchenko, P.P., Kulikov, L.L., Scherbinin, V.V. (2019) 'Application of the method of optimized delta-transformations in the control landing problem for an unmanned aerial vehicle', *J. of Comput. and Systems Sci. Int.*, 58(5), pp. 786–800. doi: 10.1134/S0002338819050093.
12. Sairanov, A.S., Kasatkina, E.V., Nefedov, D.G., Rusyak, I.G. (2019) 'The application of genetic algorithms for organizational systems management in case of emergency', *Computer Research and Modeling*, 11(3), pp. 533–556. doi: 10.20537/2076-7633-2019-11-3-533-556.

Авторы

Катулин Михаил Сергеевич¹, инженер-программист, mikhail@katulin.ru
Решетников Андрей Геннадьевич¹, к.т.н., старший научный сотрудник, agreshetnikov@jinr.ru
Рябов Андрей Русланович¹, старший лаборант, rar@jinr.ru
Ульянов Сергей Викторович^{1,2}, д.ф.-м.н., главный научный сотрудник, ulyanovsv46_46@mail.ru

¹ Объединенный институт ядерных исследований, Лаборатория информационных технологий им. М.Г. Мещерякова, г. Дубна, 141980, Россия
² Университет «Дубна», Институт системного анализа и управления, г. Дубна, 141980, Россия

Authors

Mikhail S. Katulin¹, Engineer-Programmer, mikhail@katulin.ru
Andrey G. Reshetnikov¹, Cand. of Sci. (Engineering), Senior Researcher, agreshetnikov@jinr.ru
Andrey R. Ryabov¹, Senior Assistant, rar@jinr.ru
Sergey V. Ulyanov^{1,2}, Dr.Sci. (Physics and Mathematics), Chief Researcher, ulyanovsv46_46@mail.ru

¹ Joint Institute for Nuclear Research, M.G. Meshcheryakov Laboratory of Information Technologies, Dubna, 141980, Russian Federation
² Dubna University, Institute of System Analysis and Management, Dubna, 141980, Russian Federation

Использование метода валидации выделением особенностей для сопоставления наборов экспериментальных или смоделированных данных

Н.Ю. Шайманов¹, В.П. Авраамов¹, А.А. Иванов¹✉, С.П. Куксенко¹

¹ Томский государственный университет систем управления и радиоэлектроники (ТУСУР) – Научно-исследовательская лаборатория «Безопасность и электромагнитная совместимость радиоэлектронных средств», г. Томск, 634050, Россия

Ссылка для цитирования

Шайманов Н.Ю., Авраамов В.П., Иванов А.А., Куксенко С.П. Использование метода валидации выделением особенностей для сопоставления наборов экспериментальных или смоделированных данных // Программные продукты и системы. 2024. Т. 37. № 3. С. 310–317. doi: 10.15827/0236-235X.142.310-317

Информация о статье

Группа специальностей ВАК: 1.2.2

Поступила в редакцию: 31.01.2024

После доработки: 23.04.2024

Принята к публикации: 27.04.2024

Аннотация. Данная работа посвящена методу валидации выделением особенностей (FSV), широко используемому зарубежными специалистами в области радиотехники и электроники при количественной оценке совпадения двух наборов данных (например, при сопоставлении результатов компьютерного моделирования и эксперимента). Для этого создано новое свободно распространяемое программное средство на основе метода FSV, представлены теоретические сведения о методе, а также примеры его использования. Так, приведено описание процедуры сопоставления данных методом FSV. Отмечены особенности разбиения данных на поддиапазоны, основы получения мер разности между наборами и интерпретации рассчитанных мер при построении оценочных гистограмм. Разработана блок-схема, наглядно описывающая алгоритм работы метода FSV и специфику его программной реализации. Используя эту блок-схему, на языке C++ разработано программное средство на основе метода FSV. Приведены сведения об инструментах разработки программного средства, его программной архитектуре, графическом интерфейсе пользователя, а также о функциональных возможностях. Даны примеры использования метода FSV в виде гистограмм глобальной меры разности, полученных при сопоставлении четырех наборов данных, описывающих сложные и нелинейные частотные зависимости: напряжения в многосвязной линии передачи, эффективности экранирования металлического корпуса, модуля коэффициента передачи проводной структуры, напряженности электрического поля снаружи резонатора со щелью. В результате показано, что разработанное программное средство и метод FSV позволяют получить достоверную количественную оценку совпадения двух наборов данных даже тогда, когда их качественное визуальное сопоставление затруднительно.

Ключевые слова: метод валидации выделением особенностей, FSV, сопоставление данных, программное средство

Благодарности. Исследование выполнено за счет гранта РФФИ № 23-79-10165, <https://rscf.ru/project/23-79-10165/>

Введение. При решении исследовательских и инженерных задач в области радиотехники и электроники часто возникает необходимость сравнения двух наборов данных (например, результатов компьютерного моделирования и натурального эксперимента). Для этого, как правило, выполняется расчет погрешностей или отклонений, которые, однако, не всегда дают адекватную оценку схожести результатов, особенно, если сопоставляемые данные описывают сложные нелинейные зависимости.

Для сопоставления двух наборов данных предложен метод валидации выделением особенностей [1–3], известный в зарубежной литературе как feature selective validation method (далее – метод FSV). Данный метод позволяет получить количественную оценку качества совпадения двух наборов данных и избавиться

от элемента субъективности в процессе этой оценки. После своего появления метод FSV быстро завоевал популярность и стал широко использоваться для верификации математических моделей в области вычислительной электродинамики [4–6] и в других сферах [7, 8]. Метод также был включен в знаковый зарубежный стандарт IEEE STD 1597, регламентирующий процедуры тестирования математического и программного обеспечения, предназначенного для электромагнитного моделирования. Несмотря на широкую распространенность и удобство метода FSV, отечественные исследователи и разработчики его практически не применяют. Цель данной работы – восполнить этот пробел за счет демонстрации примеров использования метода FSV, а также разработки простого открытого программного средства (ПС), реализующего этот метод.

Алгоритм сопоставления данных методом FSV

Алгоритм сопоставления данных с помощью метода FSV включает четыре основных этапа: предобработка данных, их разделение на диапазоны, расчет мер разности между наборами данных, интерпретация рассчитанных мер для построения оценочной гистограммы (рис. 1).

На этапе предобработки осуществляются импорт сопоставляемых наборов данных, их интерполяция (для получения одинакового числа отсчетов) и последующее применение к ним быстрого преобразования Фурье. После быстрого преобразования Фурье число отсчетов в наборах остается таким же, как и в исходных данных до преобразования.

На втором этапе производится разделение данных, полученных после быстрого преобразования Фурье, на три диапазона: *DC* (постоянный ток), *Lo* (низкие частоты) и *Hi* (высокие частоты). К диапазону *DC* обычно относят все отсчеты с номерами *i* от 1 до 5, причем данные, соответствующие этим отсчетам, при дальней-

шей оценке не используются, то есть удаляются из наборов. Для определения границы между диапазонами *Lo* и *Hi* сначала в каждом из наборов данных определяются так называемые точки 40 % ($i_{40\%}$). Каждая такая точка соответствует номеру отсчета, на котором сумма всех значений данных на предшествующих ему отсчетах равна 40 % от общей суммы *S* всех значений в наборе данных. Затем из двух найденных точек $i_{40\%}$ выбирается наибольшая, которая назначается в качестве точки i_{bp} , разделяющей диапазоны *Lo* и *Hi*. Если исходные наборы содержат *K* отсчетов, в результате получим $Lo = X \{6, \dots, i_{bp}\}$ и $Hi = X \{i_{bp}+1, \dots, K\}$. В завершение этапа разделения данных к полученным диапазонам *Lo* и *Hi* применяется обратное преобразование Фурье.

Следует отметить, что описанный метод разделения данных на диапазоны не является единственно возможным и тем более оптимальным. Так, помимо жесткого разбиения наборов данных, широко распространены методы разделения на основе фильтрации, в результате которой в диапазоны *Lo* и *Hi* включаются пять или девять общих точек (отсчетов) [2].

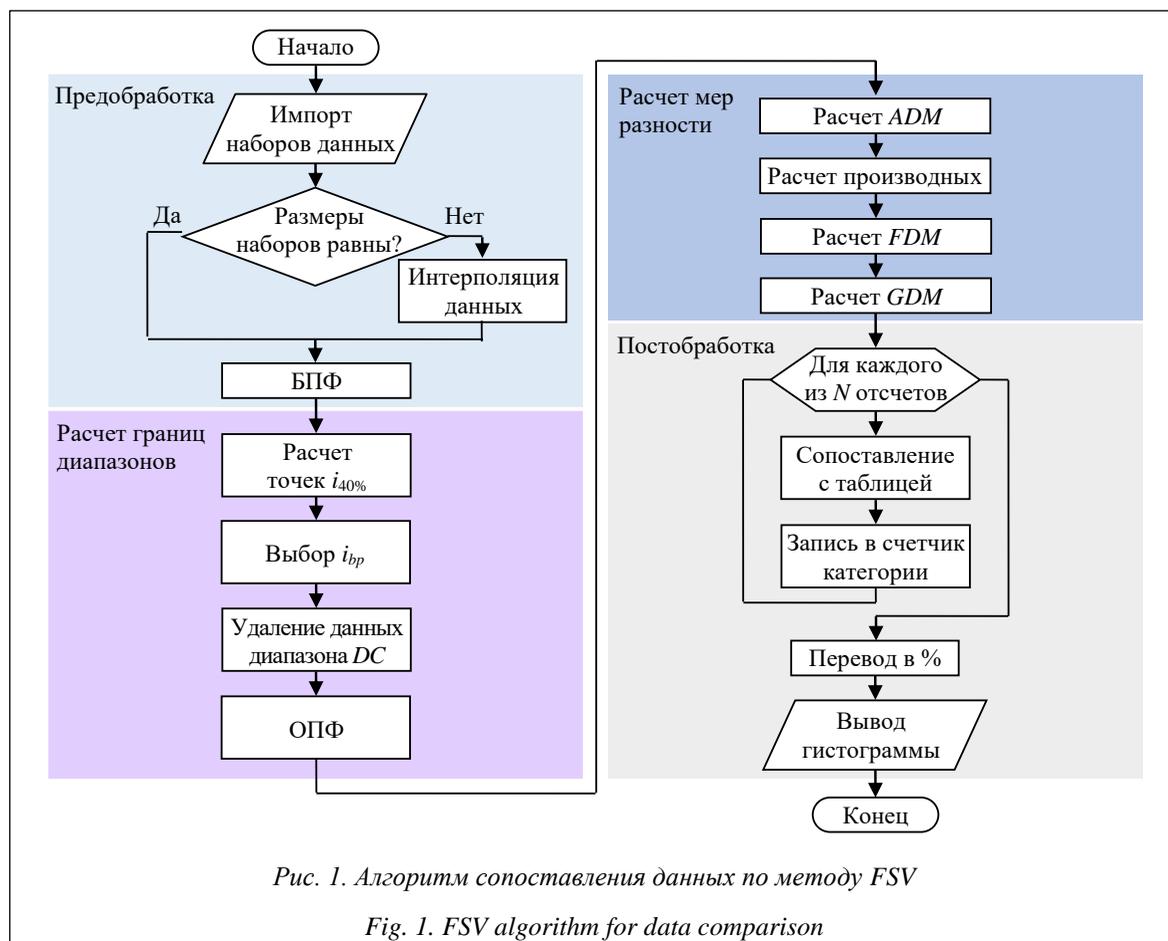


Рис. 1. Алгоритм сопоставления данных по методу FSV

Fig. 1. FSV algorithm for data comparison

На следующем этапе отсчеты, попавшие в диапазоны Lo и Hi , используются для расчета мер разности между исходными наборами данных (процедура расчета подробно изложена в работе [9]). Так, для сопоставления наборов по амплитудному критерию (тренду) применяется так называемая мера разницы амплитуд (amplitude difference measure, ADM). Различие между скоростями изменения наборов данных оценивается по значениям меры разности характеристик (feature difference measure, FDM). Для общей оценки сходимости часто используется глобальная мера разности (global difference measure, GDM), определяемая для n -го отсчета как

$$GDM(n) = \sqrt{ADM(n)^2 + FDM(n)^2}.$$

В рамках заключительного этапа с помощью вычисленных значений мер ADM , FDM или GDM строится доверительная гистограмма, отображающая оценку согласованности двух наборов данных. Для построения гистограммы применяется таблица интерпретации результатов метода FSV, устанавливающая соотношения между значениями рассчитанной меры и качественными категориями от «очень плохо» до «отлично». В алгоритме из рисунка 1 для получения оценочной гистограммы на этапе постобработки создаются шесть счетчиков (согласно числу качественных категорий), которые фиксируют резуль-

таты проверки каждого из N вычисленных значений GDM , FDM или ADM по таблице. На основе информации из счетчиков формируется, а затем выводится оценочная гистограмма с процентной шкалой.

Интерпретация результатов метода FSV

Interpretation of FSV results

Значение GDM	Результат
$GDM < 0,1$	Отлично
$0,1 \leq GDM < 0,2$	Очень хорошо
$0,2 \leq GDM < 0,4$	Хорошо
$0,4 \leq GDM < 0,8$	Удовлетворительно
$0,8 \leq GDM < 1,6$	Плохо
$1,6 \leq GDM$	Очень плохо

Разработка ПС

На базе алгоритма (рис. 1) разработано ПС для сопоставления двух наборов данных методом FSV (<https://talgat.org/>). Разработка выполнялась на базе MS Windows 10 с использованием возможностей фреймворка Qt Creator на языке C++. Также был применен ряд вспомогательных библиотек, таких как Armadillo (для реализации процедур быстрого преобразования Фурье и обратного преобразования Фурье), QtXlsx (для работы с файлами excel) и QCustomPlot (для построения и визуализации исход-

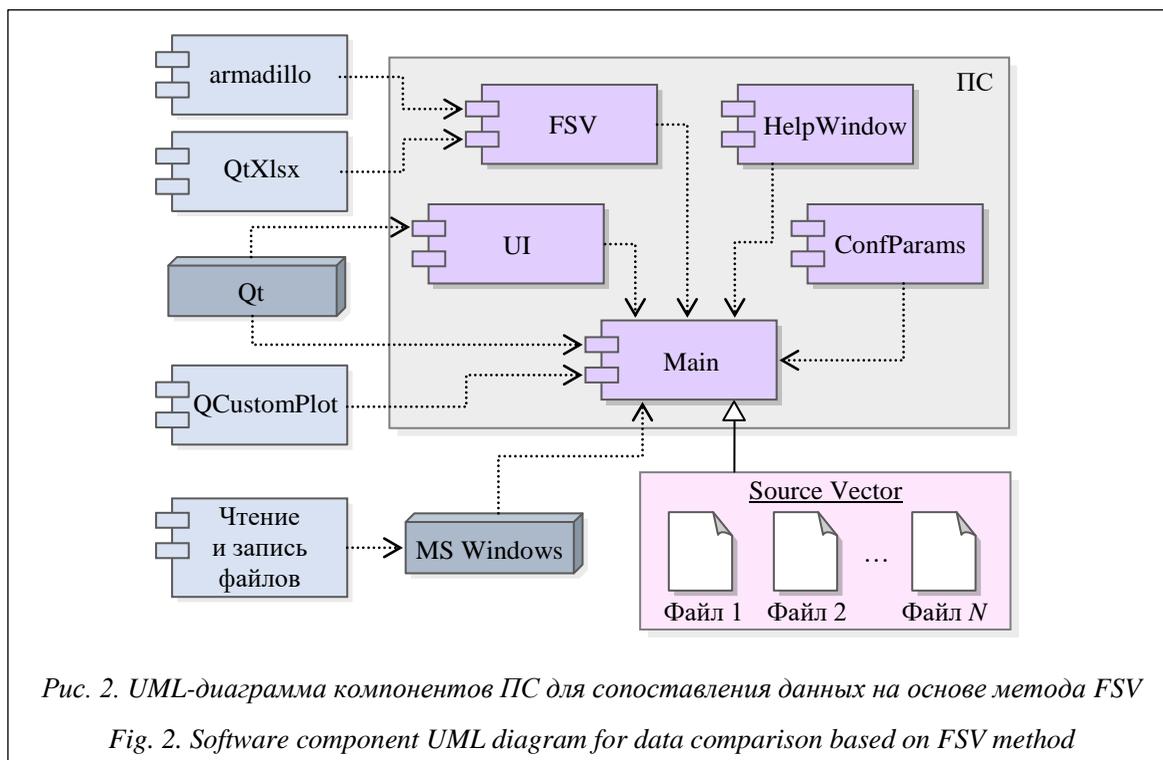


Рис. 2. UML-диаграмма компонентов ПС для сопоставления данных на основе метода FSV

Fig. 2. Software component UML diagram for data comparison based on FSV method

ных данных и оценочных гистограмм). На рисунке 2 приведена UML-диаграмма компонентов созданного ПС. Для объединения работы всех компонентов используется файл Main, управляющий вызовом функций и обработкой данных, а также реализующий процедуры логирования. Основой модуля является компонент FSV, выполняющий функции импорта и экспорта данных (из директории Source Vector), вычислений по методу FSV, а также построения оценочных гистограмм и рассчитанных мер для их визуализации в графическом интерфейсе (UI). Чтение и запись файлов реализуются базовыми средствами операционной системы.

Разработанное ПС также содержит ряд вспомогательных классов. Так, класс HelpWindow отвечает за отображение справочной информации по работе с ПС, а ConfParams – за конфигурирование (изменение стиля) графического интерфейса.

Для разработанных ПС установлены следующие минимальные системные требования: операционная система Windows 7,8 или 10; 78 Мб свободной памяти на диске; объем ОЗУ не менее 2 Гб; двухъядерный процессор с тактовой частотой не менее 2 ГГц; наличие устройств ввода (клавиатура, мышь) и вывода (монитор).

На рисунке 3 показан графический интерфейс разработанного ПС. В верхней части окна располагаются переключатели между вкладками, содержащими исходные данные, рассчитанные значения *ADM*, *FDM* и *GDM* и их гистограммы. Вкладка «Главная» включает в себя элементы управления, необходимые для импорта наборов данных в ПС. Данные могут быть загружены в популярных текстовых и табличных форматах: txt, csv, xls или xlsx.

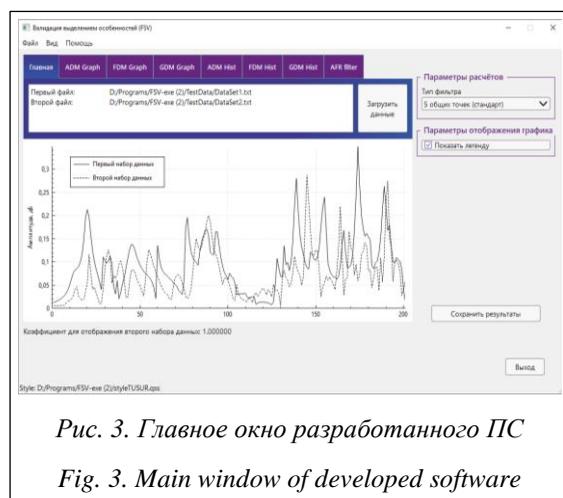


Рис. 3. Главное окно разработанного ПС

Fig. 3. Main window of developed software

Кроме того, доступна загрузка файлов в формате SnP (Touchstone), широко применяющемся в программах для электромагнитного и схемотехнического моделирования, а также в векторных и скалярных анализаторах цепей. Для загрузки в ПС данные должны быть разделены по двум отдельным файлам (допускается загрузка двух файлов разных форматов) либо объединены в один, содержащий несколько столбцов. Для удобной работы в ПС добавлено окно для предварительного просмотра содержимого импортируемых файлов (<http://www.swsys.ru/uploaded/image/2024-3/7.jpg>), а также реализован алгоритм кусочно-линейной интерполяции, позволяющий сравнивать наборы данных с разным числом отсчетов.

В правой части вкладки «Главная» доступен выбор типа фильтра, использующегося при разделении загруженных наборов данных на диапазоны. Всего в ПС реализованы четыре варианта разделения данных, в том числе предусматривающие наличие общих точек (отсчетов) в диапазонах *Lo* и *Hi*. Для наглядности выбранная пользователем процедура разделения данных демонстрируется в схематичном виде на графике, расположенном во вкладке AFR filter.

Результаты обработки данных методом FSV отображаются во вкладках ADM Graph–GDM Hist, каждая из которых имеет встроенную графическую сцену для вывода значений мер или их гистограмм (<http://www.swsys.ru/uploaded/image/2024-3/8.jpg>). Для настройки этих сцен в ПС реализованы функции масштабирования, визуализации линий сетки, отображения результатов в процентной шкале и пр. С целью последующего использования результатов вычислений их можно сохранить в табличных форматах csv, xls, xlsx или экспортировать из ПС в виде растровых изображений в формате jpg.

Примеры использования

Для демонстрации возможностей метода FSV и разработанного ПС выполнено сравнение четырех разных наборов данных. Во всех случаях сходимость оценивалась по гистограмме *GDM*, а для разделения наборов на диапазоны применялась фильтрация с пятью общими точками. Первые сопоставленные наборы показаны на рисунке 4а. Данные взяты из [10] и представляют собой частотные зависимости напряжения *U* на одном из проводников многопроводной линии передачи, полученные с помощью численного метода конечных элементов

и гибридной модели, сочетающей квазистатистический и схмотехнический виды анализа. Из рисунка 4а видно, что до отсчета 600 данные отлично согласуются. С увеличением номера отсчета появляются различия по амплитуде и положениям локальных минимумов, что усложняет оценку сходимости. При этом в целом различия выглядят адекватными и допустимыми для двух принципиально разных подходов к моделированию. Эта качественная оценка подтверждается гистограммой *GDM*, показанной на рисунке 4б. Так, видно, что преимущественно (для 81 % отсчетов) данные сходятся с оценками «отлично», «очень хорошо» и «хорошо». Причем в категорию «очень хорошо» попадает до 40 % данных из наборов.

Следующие наборы данных, показанные на рисунке 5а, взяты из [11]. Они описывают частотные зависимости эффективности экранирования прямоугольного металлического кор-

пуса со щелью, полученные методом конечных элементов и полуаналитической моделью. Видно, что до отсчета 500 данные хорошо согласуются. Затем в диапазоне от 500 до 1 000 совпадают только точки минимумов, а по амплитудам имеются значительные различия. При этом во всех последующих отсчетах (от 1 000 до 1 200) наборы существенно различаются. В результате получение однозначной качественной оценки сходимости результатов не представляется возможным. Однако оценку можно получить с использованием метода FSV. Так, из гистограммы *GDM* (рис. 5б), видно, что почти для 30 % отсчетов совпадение оценивается как «удовлетворительное». При этом к категориям «отлично» и «очень хорошо» относятся всего лишь 24 % от общего числа данных.

Для третьего примера использованы данные из [12], описывающие частотные зависимости

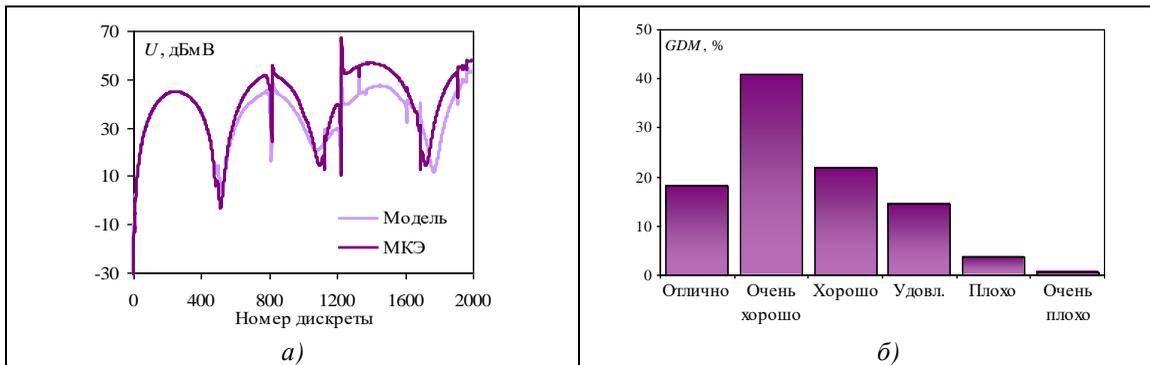


Рис. 4. Пример тестового использования метода FSV: а) данные, рассчитанные гибридной моделью и методом конечных элементов; б) их гистограмма *GDM*

Fig. 4. Example of using the FSV method: а) data calculated using the hybrid model and FEM; б) *GDM* histogram based on these data

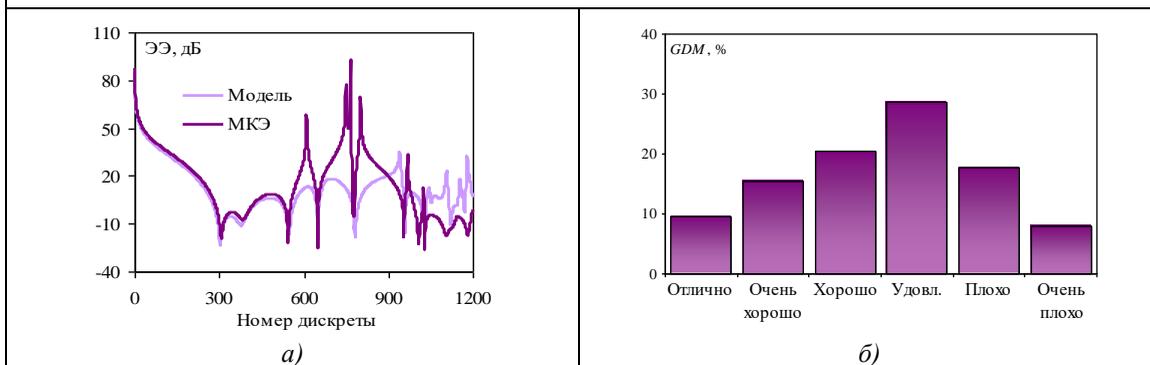


Рис. 5. Пример тестового использования метода FSV: а) данные, рассчитанные полуаналитической моделью и методом конечных элементов; б) их гистограмма *GDM*

Fig. 5. Example of using the FSV method: а) data calculated using the semi-analytical model and FEM; б) *GDM* histogram based on these data

нормированного коэффициента передачи $|S_{21}|$ между концами кабеля, находящегося внутри объемного резонатора, полученные моделированием и экспериментально (посредством измерений). Выбранные наборы данных показаны на рисунке 6а. Видно, что результаты моделирования и эксперимента очень близки, однако на измеренной частотной зависимости $|S_{21}|$ имеется множество осцилляций, которые не позволяют дать однозначную оценку сходимости данных без использования метода FSV. Гистограмма GDM, полученная для исследуемых наборов, представлена на рисунке 6б. Видно, что почти для 90 % отсчетов данные согласуются с оценками «отлично», «очень хорошо» и «хорошо». При этом оставшиеся 10 % отнесены к категории «удовлетворительно», а данные, оцениваемые «плохо» и «очень плохо», в наборах отсутствуют.

Последние использованные наборы данных показаны на рисунке 7а. Они представляют со-

бой частотные зависимости модуля напряженности электрического поля $|E|$, полученные при моделировании и измерениях снаружи экранирующего корпуса с излучающей антенной внутри. Видно, что положения максимумов $|E|$ на зависимостях, полученных разными подходами, сходятся, но имеются расхождения по их амплитудам. При этом качественная оценка сходимости данных осложняется из-за их большого динамического диапазона.

Гистограмма GDM, соответствующая результатам моделирования и измерений $|E|$, показана на рисунке 7б. Видно, что для 27 % отсчетов совпадение оценивается как «очень хорошее». При этом почти 35 % отчетов отнесено к категориям «удовлетворительно», «плохо» и «очень плохо».

Заключение

Приведены теоретические основы и алгоритм работы метода FSV, предназначенного

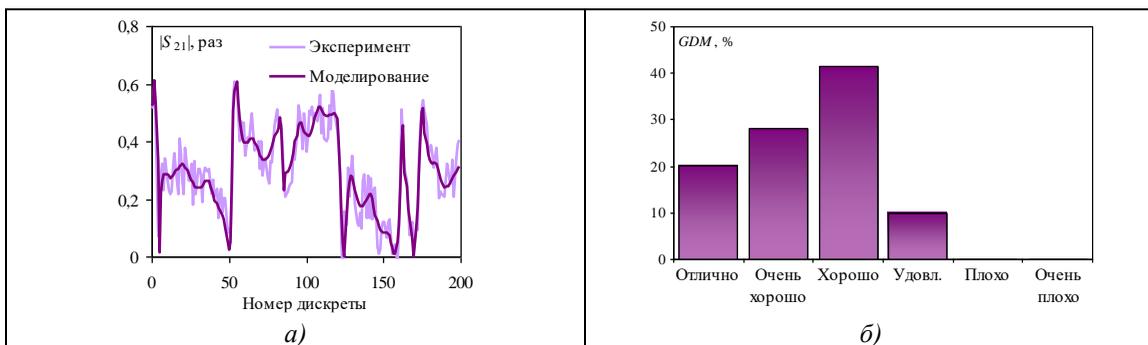


Рис. 6. Пример тестового использования метода FSV: а) данные моделирования и измерений $|S_{21}|$; б) их гистограмма GDM

Fig. 6. Example of using the FSV method: а) $|S_{21}|$ modeling and measurement data; б) corresponding GDM histogram

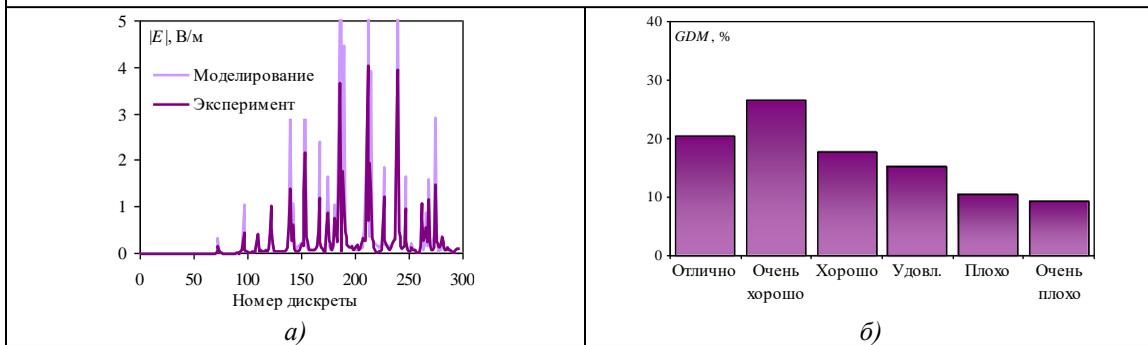


Рис. 7. Пример тестового использования метода FSV: а) данные моделирования и измерений $|E|$; б) их гистограмма GDM

Fig. 7. Example of using the FSV method: а) $|E|$ modeling and measurement data; б) corresponding GDM histogram

для количественной оценки совпадения двух наборов данных. Представлены результаты разработки ПС на основе метода FSV. Для демонстрации возможностей этого метода и ПС сопоставлены четыре набора данных. В результате показано, что разработанное ПС и метод FSV позволяют получить достоверные количественные оценки сходимости двух наборов данных даже тогда, когда их качественное ви-

зуальное сопоставление затруднительно. Поэтому метод FSV может быть рекомендован к использованию, как минимум, специалистам, разрабатывающим математическое и программное обеспечение для моделирования сложных технических систем. Отметим, что в настоящий момент представленное ПС доступно для свободного скачивания на сайте лаборатории разработчика.

Список литературы

1. Duffy A., Orlandi A. The influence of data density on the consistency of performance of the feature selective validation (FSV) technique. *ACES*, 2006, vol. 21, no. 2, pp. 164–172.
2. Duffy A., Zhang G. FSV: State of the art and current research fronts. *IEEE Electromagnetic Compatibility Magazine*, 2020, vol. 9, no. 3, pp. 55–62. doi: 10.1109/MEMC.2020.9241552.
3. Bongiorno J., Mariscotti A. Uncertainty and sensitivity of the feature selective validation (FSV) method. *Electronics*, 2022, vol. 11, no. 16, art. 2532. doi: 10.3390/electronics11162532.
4. Kadem K., Benyoubi F., Bensetti M. et al. An efficient method for dimensioning magnetic shielding for an induction electric vehicle charging system. *PIER*, 2021, vol. 170, pp. 153–167. doi: 10.2528/PIER21031903.
5. Zhang G., Duffy A. Applying FSV to the comparison of return path integrity in high speed circuit designs. *IEEE L-EMCPA*, 2021, vol. 3, no. 2, pp. 78–81. doi: 10.1109/LEMCPA.2021.3064420.
6. Rabat A., Bonnet P., Drissi K.E.K., Girard S. An analytical evaluation of the shielding effectiveness of enclosures containing complex apertures. *IEEE Access*, 2021, vol. 9, pp. 147191–147200. doi: 10.1109/ACCESS.2021.3123441.
7. Stöhr T., Rasek G.A., Murthy N.S. Agreement quantification of a numerical EMC computer model and test infrastructure for the HV power train emissions for an electric vehicle. *Proc. EMC Europe*, 2023, pp. 1–6. doi: 10.1109/EMCEurope57790.2023.10274221.
8. Zhang G., Zappala D., Crabtree C., Donaghy-Spargo C., Hogg S., Duffy A. Validation of a non-contact technique for torque measurements in wind turbines using an enhanced transient FSV approach. *Measurement*, 2020, vol. 151, art. 107261. doi: 10.1016/j.measurement.2019.107261.
9. Шайманов Н.Ю., Осинцев А.В., Иванов А.А. Об использовании метода FSV в задачах электромагнитной совместимости // *Электронные средства и системы управления: мат. конф.* 2023. С. 50–52.
10. Алхадж Хасан А. Виртуальные испытания структуры с модальным резервированием в системе TALGAT // *СИБРЕСУРС: мат. конф.* 2022. С. 102–107.
11. Ivanov A.A., Demakov A.V., Komnatov M.E., Gazizov T.R. Semi-analytical approach for calculating shielding effectiveness of an enclosure with a filled aperture. *Electrica*, 2022, vol. 22, no. 2, pp. 220–225. doi: 10.54614/electrica.2022.21152.
12. Williams A.J.M., Benson T.M., Duffy A.P. Determining the accuracy of TLM in simulating the behaviour of resonant cavities with arbitrary dielectric loading. *IJE*, 1997, vol. 83, no. 5, pp. 645–660. doi: 10.1080/002072197135201.

Applying the feature selective validation method to compare experimental or simulated datasets

Nikita Yu. Shaymanov¹, Vladimir P. Avraamov¹,
Anton A. Ivanov¹✉, Sergey P. Kuksenko¹

¹ Tomsk State University of Control Systems and Radioelectronics (TUSUR) –
Research Laboratory “Safety and Electromagnetic Compatibility of Radioelectronic Equipment”,
Tomsk, 634050, Russian Federation

For citation

Shaymanov, N.Yu., Avraamov, V.P., Ivanov, A.A., Kuksenko, S.P. (2024) ‘Applying the feature selective validation method to compare experimental or simulated datasets’, *Software & Systems*, 37(3), pp. 310–317 (in Russ.). doi: 10.15827/0236-235X.142.310-317

Article info

Received: 31.01.2024

After revision: 24.04.2024

Accepted: 27.04.2024

Abstract. This paper is intended to introduce domestic engineers and researchers to the feature selective validation (FSV) method, which is widely used by foreign specialists in radio-engineering and electronics for quantitative evaluation of the

agreement between two datasets (for example, when comparing the results of simulation and experiment). For this purpose, the paper presents a new open source software tool based on the FSV method. It also gives theoretical information about the method and shows examples of its use, describes a FSV data comparison procedure in detail. The paper notes the peculiarities of data partitioning into sub-ranges, the basics of calculating measures indicating the difference between sets, and the interpretation of calculated measures when constructing evaluation histograms. The developed block diagram visually describes the algorithm of the FSV method and the specifics of its software implementation. A software tool based on the FSV method was developed in C++ using this block diagram. The paper provides information about the development tools, software architecture, graphical user interface, and functionalities. The final part of the paper presents examples of applying the FSV method in the form of global difference measure histograms obtained by comparing four data sets describing complex and non-linear frequency dependencies: voltage in a multi-conductor transmission line, shielding effectiveness of a metal enclosure, modulus of the transmission coefficient for a wire structure, electric field strength outside an apertured cavity. The results showed that the developed software tool and the FSV method allow obtaining a reliable quantitative assessment of the agreement between two data sets, even when their qualitative visual comparison is difficult.

Keywords: feature selective validation, FSV, datasets analysis, validation, software

Acknowledgements. The research was carried out at the expense of Russian Science Foundation grant 23-79-10165, <https://rscf.ru/project/23-79-10165/>

References

1. Duffy, A., Orlandi, A. (2006) 'The influence of data density on the consistency of performance of the feature selective validation (FSV) technique', *ACES*, 21(2), pp. 164–172.
2. Duffy, A., Zhang, G. (2020) 'FSV: State of the art and current research fronts', *IEEE Electromagnetic Compatibility Magazine*, 9(3), pp. 55–62. doi: 10.1109/MEMC.2020.9241552.
3. Bongiorno, J., Mariscotti, A. (2022) 'Uncertainty and sensitivity of the feature selective validation (FSV) method', *Electronics*, 11(16), art. 2532. doi: 10.3390/electronics11162532.
4. Kadem, K., Benyoubi, F., Bensetti, M. et al. (2021) 'An efficient method for dimensioning magnetic shielding for an induction electric vehicle charging system', *PIER*, 170, pp. 153–167. doi: 10.2528/PIER21031903.
5. Zhang, G., Duffy, A. (2021) 'Applying FSV to the comparison of return path integrity in high speed circuit designs', *IEEE L-EMCPA*, 3(2), pp. 78–81. doi: 10.1109/LEMCPA.2021.3064420.
6. Rabat, A., Bonnet, P., Drissi, K.E.K., Girard, S. (2021) 'An analytical evaluation of the shielding effectiveness of enclosures containing complex apertures', *IEEE Access*, 9, pp. 147191–147200. doi: 10.1109/ACCESS.2021.3123441.
7. Stöhr, T., Rasek, G.A., Murthy, N.S. (2023) 'Agreement quantification of a numerical EMC computer model and test infrastructure for the HV power train emissions for an electric vehicle', *Proc. EMC Europe*, pp. 1–6. doi: 10.1109/EMCEurope57790.2023.10274221.
8. Zhang, G., Zappala, D., Crabtree, C., Donaghy-Spargo, C., Hogg, S., Duffy, A. (2020) 'Validation of a non-contact technique for torque measurements in wind turbines using an enhanced transient FSV approach', *Measurement*, 151, art. 107261. doi: 10.1016/j.measurement.2019.107261.
9. Shaymanov, N.Yu., Osintsev, A.V., Ivanov, A.A. (2023) 'On the use of FSV method in electromagnetic compatibility problems', *Proc. Conf. Electronic Aids and Control Systems*, pp. 50–52 (in Russ.).
10. Alhaj Hasan, A. (2022) 'Virtual experiment of a structure with modal reservations in the TALGAT system', *Proc. Conf. SIBRESOURCE*, pp. 102–107 (in Russ.).
11. Ivanov, A.A., Demakov, A.V., Komnatnov, M.E., Gazizov, T.R. (2022) 'Semi-analytical approach for calculating shielding effectiveness of an enclosure with a filled aperture', *Electrica*, 22(2), pp. 220–225. doi: 10.54614/electrica.2022.21152.
12. Williams, A.J.M., Benson, T.M., Duffy, A.P. (1997) 'Determining the accuracy of TLM in simulating the behaviour of resonant cavities with arbitrary dielectric loading', *IJE*, 83(5), pp. 645–660. doi: 10.1080/002072197135201.

Авторы

Шайманов Никита Юрьевич¹,

инженер, vishado1@mail.ru

Авраамов Владимир Петрович¹,

инженер, avp@2i.tusur.ru

Иванов Антон Андреевич¹, к.т.н.,

старший научный сотрудник, anton.ivvv@gmail.com

Куксенко Сергей Петрович¹,

д.т.н., доцент, зав. лабораторией,

ksergp@tu.tusur.ru

Authors

Nikita Yu. Shaymanov¹,

Engineer, vishado1@mail.ru

Vladimir P. Avraamov¹,

Engineer, avp@2i.tusur.ru

Anton A. Ivanov¹, Cand. of Sci. (Engineering),

Senior Researcher, anton.ivvv@gmail.com

Sergey P. Kuksenko¹, Dr.Sci. (Engineering),

Associate Professor, Head of Laboratory,

ksergp@tu.tusur.ru

¹ Томский государственный университет систем управления и радиоэлектроники (ТУСУР) – Научно-исследовательская лаборатория «Безопасность и электромагнитная совместимость радиоэлектронных средств», г. Томск, 634050, Россия

¹ Tomsk State University of Control Systems and Radioelectronics (TUSUR) – Research Laboratory “Safety and Electromagnetic Compatibility of Radioelectronic Equipment”, Tomsk, 634050, Russian Federation

Применение реализованных в MATLAB методов Big Data в задаче выделения геомагнитных пульсаций типа Pc3 из данных INTERMAGNET

А.Г. Коробейников ¹✉

¹ СПбФ ИЗМИРАН, г. Санкт-Петербург, 199034, Россия

Ссылка для цитирования

Коробейников А.Г. Применение реализованных в MATLAB методов Big Data в задаче выделения геомагнитных пульсаций типа Pc3 из данных INTERMAGNET // Программные продукты и системы. 2024. Т. 37. № 3. С. 318–323. doi: 10.15827/0236-235X.142.318-323

Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 14.02.2024

После доработки: 18.03.2024

Принята к публикации: 25.03.2024

Аннотация. При решении фундаментальных и прикладных задач в различных предметных областях широко применяется система MATLAB. С использованием реализованных в MATLAB методов технологии Big Data решена и рассмотренная в данной статье задача выделения геомагнитных пульсаций типа Pc3 из данных, поставляемых магнитными обсерваториями, входящими в международную сеть INTERMAGNET. Задача решается в несколько этапов. На первом этапе выполняется предобработка исходных данных. В данном случае это удаление выбросов. На втором этапе производится проектирование полосового цифрового фильтра оптимального порядка с бесконечно-импульсной характеристикой (БИХ-фильтр) методом Золотарева–Кауэра. Параметры фильтра задаются исходя от диапазона периода геомагнитных пульсаций Pc3 – 10–45 секунд. На следующем этапе осуществляется прогон предобработанного набора данных через спроектированный БИХ-фильтр с последующим устранением нелинейного искажения фазы фильтра. Выбор метода проектирования обусловлен следующими характеристиками разработанного БИХ-фильтра: наличие уравновешенного поведения пульсаций АЧХ как в полосе пропускания, так и в полосе задержки; ослабление требований к форме амплитудно-частотной характеристики в полосе задержки, позволяющее добиться быстрого перехода между полосой пропускания и полосой задержки; минимальные порядок и сложность БИХ-фильтра, спроектированного по методу цифровых фильтров Золотарева. Одним из требований при решении задачи является условие применения реализованных в MATLAB технологий Big Data. Это требование обусловлено большим объемом исходных данных: непрерывная съемка данных о состоянии геомагнитного поля по четырем каналам в течение целого года с частотой съема, равной одной секунде. Полученные результаты можно использовать для решения как прикладных задач, например, анализ состояния космической погоды, так и фундаментальных, например, дальнейшее развитие теории геомагнитных пульсаций. Проектирование полосового БИХ-фильтра и расчеты проводились в MATLAB R2023b.

Ключевые слова: Big Data, INTERMAGNET, MATLAB, Pc3, БИХ-фильтр, геомагнитные пульсации, полосовой фильтр

Благодарности. Работа выполнена при поддержке гранта РФФИ № 23-27-00011

Введение. В качестве программного инструментария система MATLAB используется в процессе решения задач в самых разных предметных областях [1, 2]. В данной работе рассмотрена задача, результаты решения которой относятся к области геомагнетизма – выделение *геомагнитных пульсаций* (ГП) из данных, полученных магнитной обсерваторией Lycksele (Швеция, Геологическая служба Швеции, международный IAGA-код LYC), входящей в международную сеть ИНТЕРМАГНЕТ [3–5]. Выбор этой обсерватории обусловлен передачей от нее в ИНТЕРМАГНЕТ достаточно качественных данных измерений состояний геомагнитного поля.

Физически ГП представляют собой изменения напряженности геомагнитного поля в разном диапазоне частот в зависимости от времени и различных факторов. Формально ГП

обычно представляют в виде магнитогидродинамических волн в околоземной плазме [6]. Возбуждение таких волн происходит в силу разнообразных физических процессов, которые возникают, например, в ионосфере, магнитосфере или в солнечном ветре. Отсюда следует, что ГП имеют высокую информативность, так как их анализ позволяет получать много информации, например, о параметрах среды в областях генерации ГП, об особенностях развития геомагнитных бурь и суббурь, о потоках высыпавшихся частиц и путях их распространения к земной поверхности. Фиксация ГП на поверхности Земли происходит путем регистрации ультранизкочастотных электромагнитных колебаний [6]. Таким образом, можно сделать вывод, что анализ ГП – актуальная задача, связанная с решением как прикладных проблем, например, радиосвязь, диагно-

стика текущего состояния околоземной среды и прогноз космической погоды, так и фундаментальных, например, дальнейшее развитие теории ГП [7].

В статье сделан акцент на выделение ГП типа Pс3, то есть ГП с периодом колебаний, попадающим в диапазон 10–45 секунд, и амплитудой магнитной индукции 0,4–20 нТл. Такие ГП относятся к одним из самых распространенных, регистрируемых на земной поверхности. В большинстве случаев ГП Pс3 – это результат резонансных колебаний силовых линий геомагнитного поля в магнитосфере Земли. В работе предложен подход, позволяющий производить выделение ГП Pс3 из данных, полученных в геомагнитных обсерваториях.

Представленные в работе результаты относятся также к области проектирования *цифровых фильтров* (ЦФ), параметры которых задаются исходя из свойств ГП Pс3. Аналогичные исследования как в России, так и за рубежом начали проводиться достаточно давно. Например, в [8] описаны основы ЦФ, которые использовали тогда в процессе обработки данных измерений с сейсмологических, гравиметрических и магнитометрических датчиков. В настоящее время методы проектирования ЦФ базируются на современных информационных технологиях [9–11]. В последнее время при обработке данных делается акцент на использовании методов технологии Big Data [12]. Необходимо отметить, что публикаций о применении технологий Big Data, реализованных в MATLAB с представлением программного кода, для решения задачи выделения конкретного типа ГП Pс3 в доступных источниках не обнаружено. Вычислительные эксперименты были проведены в системе MATLAB R2023b под управлением Windows 11.

Постановка задачи

Проектирование оптимального рекурсивного полосового ЦФ с *бесконечно-импульсной характеристикой* (БИХ) необходимо провести при помощи метода Золотарева–Кауэра [13, 14]. Под оптимальным БИХ-фильтром понимается ЦФ минимального порядка при заданных требованиях к *амплитудно-частотной характеристике* (АЧХ), к полосе пропускания и полосе задержки [13]. Выбор данного ЦФ обусловлен его основными особенностями.

Во-первых, поведение изменений АЧХ как в полосе пропускания, так и в полосе задержки уравновешенное. Амплитуды изменений в любой из этих полос не зависят друг от друга.

Так как условие наличия жесткого требования к форме АЧХ в полосе задержки отсутствует, существует возможность быстрого перехода между полосой пропускания и полосой задержки, то есть создание очень быстрого уменьшения АЧХ, позволяющее более эффективно осуществлять дифференциацию частот. Проектирование ЦФ с подобными характеристиками с помощью других линейных фильтров практически невозможно.

Во-вторых, порядок ЦФ Золотарева и его сложность минимальные.

Такие ЦФ также относят к эллиптическим. Это связано с тем, что при их описании используют математический аппарат на базе эллиптических функций [13, 15].

Можно отметить, что эти ЦФ называют так из-за дробей Золотарева, применяемых при описании передаточных функций ЦФ. Причем нули передаточных функций расположены так, что минимумы АЧХ находятся в полосе задержки.

Исходя из свойств ГП Pс3 зададим требования к полосе пропускания и к АЧХ в полосе пропускания и полосе задержки проектируемого ЦФ:

$f_{\chi} = 1/45$ (с) ≈ 20 мГц – левая граничная частота полосы пропускания;

$f_{\chi} = 1/10$ (с) ≈ 100 мГц – правая граничная частота полосы пропускания;

$a_{\max} = 0.001$ (дБ) – максимальное допустимое затухание АЧХ в полосе пропускания;

$a_{1\min} = a_{2\min} = 150$ (дБ) – максимальное допустимое затухание АЧХ в левой ($a_{1\min}$) и правой ($a_{2\min}$) полосах задержки.

Заключительным требованием при постановке задачи является условие применения реализованных в MATLAB технологий Big Data. Это требование вытекает из большого объема исходных данных: непрерывная съемка данных о состоянии геомагнитного поля по четырем каналам в течение целого года с частотой дискретизации $f_d = 1$ (с). В силу этого необходимо применение специальных средств, таких как Datastore.

Проектирование ЦФ

Зададим параметры для полосового ЦФ, соответствующие принятым условиям. Код на MATLAB можно написать, например, так:
 $Fs_{45_s} = 1/45$; $Fs_{10_s} = 1/10$; % Диапазон частот Pс3
 $Fs = 1$; % Частота дискретизации 1 секунда
 $F_{pass1} = Fs_{45_s}$; % Левая граница частоты полосы пропускания

```
Fpass2 = Fs_10_s; % Правая граница частоты
полосы пропускания
Fstop1 = 0.99*Fpass1; % Левая граничная ча-
стота полосы задержки
Fstop2 = 1.01*Fpass2; % Правая граничная ча-
стота полосы задержки
Astop1 = 150; % Величина затухания в левой
полосе задержки (дБ)
Astop2 = 150; % Величина затухания в правой
полосе задержки (дБ)
Apass = 0.001; % Величина затухания в полосе
пропускания (дБ)
```

Создадим полосовой ЦФ с БИХ-фильтром при помощи следующей команды:

```
digital_filter_Pc4= designfilt('bandpassiir', ...;
% Полосовой БИХ-фильтр
'StopbandFrequency1', Fstop1, ...; % Левая гра-
ничная частота полосы задержки
'PassbandFrequency1', Fpass1, ...; % Левая гра-
ница частоты полосы пропускания
'PassbandFrequency2', Fpass2, ...; % Правая гра-
ница частоты полосы пропускания
'StopbandFrequency2', Fstop2, ...; % Правая гра-
ничная частота полосы задержки
'StopbandAttenuation1', Astop1, ...; % Величина за-
тухания в левой полосе задержки (дБ)
'PassbandRipple', Apass, ...; % Величина затуха-
ния в полосе пропускания (дБ)
'StopbandAttenuation2', Astop2, ...; % Величина за-
тухания в правой полосе задержки (дБ)
'DesignMethod', 'ellip', ...; % Метод Золотарева
(эллиптический)
'MatchExactly', 'passband', ...; % Полосовой ЦФ
'SampleRate', Fs); % Частота дискретизации
```

В результате выполнения этой команды будет спроектирован оптимальный устойчивый БИХ-фильтр 50-го порядка, имеющий заданные характеристики. На рисунке 1 представлена АЧХ этого ЦФ.

Понятно, что ЦФ можно спроектировать иначе [16]. В этом случае появится дополнительная информация, например, карта нулей и полюсов передаточных функций у спроектированного ЦФ.

Подготовка данных для проверки спроектированного цифрового БИХ-фильтра

Для проверки полученного ЦФ были использованы данные измерений состояний геомагнитного поля в реальном масштабе времени с частотой дискретизации $f_d = 1$ (с), полученные

геомагнитной обсерваторией Lycksele за период с 1 января по 31 декабря 2023 г. Файлы с исходными данными содержат четыре столбца – LYCX, LYCY, LYCZ и LYCF, где LYCX, LYCY, LYCZ – компоненты вектора геомагнитного поля, а LYCF – модуль этого вектора. Все значения приводятся в нанотеслах. Код на MATLAB для доступа к этим данным и их визуализации приведен в [17]. При небольшой модификации, связанной с конкретными вычислительными ресурсами, после выполнения кода появится график исходных данных (рис. 2). Применяя методы визуализации, входящие в Big Data, можно решать множество различных задач, например, максимальное погружение в данные и (или) обнаружение отклонений и аномалий. На рисунке 2 видно, что в исходных данных имеются выбросы, появившиеся, вероятнее всего, в результате прекращения работы магнитометрических сенсоров.

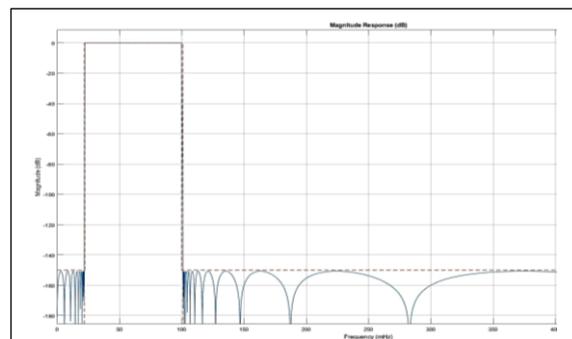


Рис. 1. АЧХ спроектированного полосового цифрового БИХ-фильтра

Fig. 1. Amplitude-frequency characteristics of the designed bandpass digital BIR filter

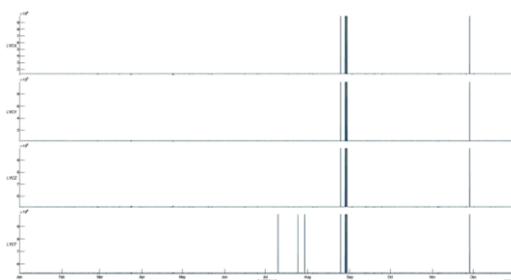


Рис. 2. Исходные данные с LYC за период 01.01.2023–31.12.2023

Fig. 2. LYC initial data for the period 01.01.2023–31.12.2023

В таких случаях общепринятой практикой считается занесение числа 99 999 в соответствующую БД. Для дальнейшей работы эти данные необходимо удалить следующим образом:

```

toDelete_XYZF = (LYC.LYCX == 99999) |
(LYC.LYCY == 99999) ...
| (LYC.LYCZ == 99999) | (LYC.LYCF =
= 99999);
LYC(toDelete_XYZF,:) = [];
clearvars toDelete_XYZF;
stackedplot(LYC).

```

Обработанные при помощи такого кода очищенные данные представлены на рисунке 3.

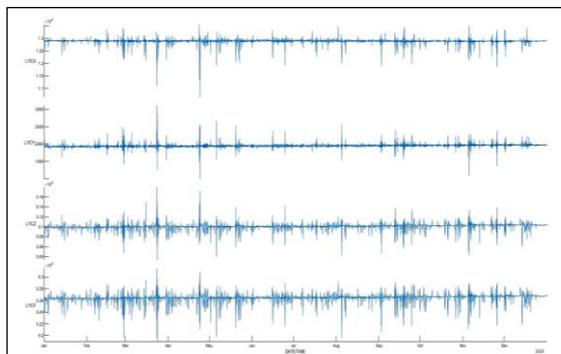


Рис. 3. Очищенные исходные данные

Fig. 3. Cleaned initial data

Выделения ГП типа Pc3

На первом шаге необходимо удалить тренд в очищенных данных, например, так: `LYC_dettrend = detrend(LYC,3)`; здесь цифра 3 задает тренд в виде кубического сплайна.

Далее пропускаем покомпонентно очищенные данные без тренда через спроектированный ЦФ, учитывая при этом фазовую задержку при помощи функции `filtfilt`:

```

LYC.LYCX = filtfilt(digital_filter_Pc3, LYC_dettrend.LYCX);
LYC.LYCY = filtfilt(digital_filter_Pc3, LYC_dettrend.LYCY);
LYC.LYCZ = filtfilt(digital_filter_Pc3, LYC_dettrend.LYCZ);
LYC.LYCF = filtfilt(digital_filter_Pc3, LYC_dettrend.LYCF).

```

В результате выполнения представленного кода в `timetable LYC` будет содержаться покомпонентно информация о ГП Pc3.

Исходя из решаемой задачи можно осуществлять еще какую-либо обработку данных, например, нормировку.

Анализ полученных результатов

При помощи методов визуализации можно определить даты, когда происходили магнитные бури. Например, анализируя x -компоненту ГП (`LYCX`), можно сказать, что 23 и 24 апреля 2023 г. происходила сильная магнитная буря (<http://www.swsys.ru/uploaded/image/2024-3/9.jpg>).

Можно видеть несимметричный по времени характер развития магнитной бури. Это дает различную информацию, например, о вариации динамического давления солнечного ветра и межпланетного магнитного поля, которые изменяют токовые системы в магнитосфере Земли и вызывают ГП.

Выводы

Предлагаемый подход дает возможность достаточно эффективно решать задачи предварительной обработки большого объема исходных данных и выделения из них ГП Pc3. Он базируется на проектировании ЦФ и использовании инструментария визуального анализа, позволяющего представлять в графической форме сразу весь набор исследуемых данных. Следует отметить, что в настоящее время при фильтрации геофизических данных также применяются методы спектрального анализа [18, 19].

Полученные результаты могут быть основанием для дальнейшей обработки и интерпретации данных, используемых, например, при прогнозировании космической погоды или анализе свойств возмущений геомагнитного поля в диапазоне длиннопериодных пульсаций.

Список литературы

1. Коробейников А.Г. Применение искусственных нейронных сетей в системах автоматического управления магнитной левитацией // Программные продукты и системы. 2022. Т. 35. № 3. С. 452–457. doi: 10.15827/0236-235X.139.452-457.
2. Егоров Е.Е. Моделирование работы манипуляционного робота в программном пакете Matlab Robotics Toolbox // Политехнический молодежный журнал. 2020. № 01. С. 1–12.
3. Коробейников А.Г. Обработка и анализ данных с российского сегмента мировой сети магнитных обсерваторий ИНТЕРМАГНЕТ // Международный журнал гуманитарных и естественных наук. 2018. № 8. С. 91–98.
4. Вишняков Д.Д., Арутюнян Д.А., Шклярук А.Д. Оценка планетарного индекса геомагнитной активности по данным стационарных магнитовариационных обсерваторий сети INTERMAGNET с применением собственного программного обеспечения // Гелиогеофизические исследования. 2023. № 38. С. 40–45.
5. Воробьев А.В., Воробьева Г.П. Подход к оценке относительной информационной эффективности магнитных обсерваторий сети INTERMAGNET // Геомагнетизм и аэронавигация. 2018. Т. 58. № 5. С. 648–652. doi: 10.1134/s0016794018050164.

6. Гульельми А.В., Потапов А. С. Частотно-модулированные ультранизкочастотные волны в околоземном космическом пространстве // УФН. 2021. Т. 191. № 5. С. 475–491. doi: 10.3367/UFN.2020.06.038777.
7. Mishin V.V., Tsegmed B., Klibanova Yu.Yu., Kurikalova M.A. Burst geomagnetic pulsations as indicators of sub-storm expansion onsets during storms. *JGR Space Phys.*, 2020, vol. 125, art. e2020JA028521. doi: 10.1029/2020JA028521.
8. Кулханек О. Введение в цифровую фильтрацию в геофизике; [пер. с англ.]. М.: Недра, 1981. 198 с.
9. Potthast R., Walter A., Rhodin A. A localized adaptive particle filter within an operational NWP framework. *Monthly Weather Review*, 2019, vol. 147, pp. 345–362. doi: 10.1175/MWR-D-18-0028.1.
10. van Leeuwen P.J., Kunsch H.R., Nerger L., Potthast R., Reich S. Particle filters for high-dimensional geoscience applications: A review. *Quarterly J. of the Royal Meteorological Society*, 2019, vol. 145, no. 723, pp. 2335–2365. doi: 10.1002/qj.3551.
11. Климова Е.Г. Локальный ансамблевый алгоритм усвоения данных для нелинейных геофизических моделей // СибЖВМ. 2023. Т. 26. № 1. С. 27–42. doi: 10.15372/SJNM20230103.
12. Мошак Н.М., Рудинская С.Р., Груздев А.А. Третья платформа информации и Big Data // Научные технологии в космических исследованиях Земли. 2023. Т. 15. № 4. С. 47–59. doi: 10.36724/2409-5419-2023-15-4-47-59.
13. Солонина А.И. Цифровая обработка сигналов в зеркале MATLAB. СПб: БХВ-Петербург, 2018. 560 с.
14. Будунова К.А., Кравченко В.Ф. Математические методы синтеза частотно-избирательных фильтров // Физические основы приборостроения. 2022. Т. 11. № 1. С. 2–21. doi: 10.25210/jfop-2201-002021.
15. Юлина А.О., Синкевич Г.И. История развития теории эллиптических функций в работах Абеля, Якоби, Вейерштрасса, Сомова // ТВИМ. 2021. № 3. С. 79–92.
16. Коробейников А.Г. Применение MATLAB при проектировании цифровых фильтров для выделения геомагнитных пульсаций типа Pc5 // Программные продукты и системы. 2023. Т. 36. № 2. С. 257–262. doi: 10.15827/0236-235X.142.257-262.
17. Коробейников А.Г. Применение методов Big Data для сравнения данных геомагнитных обсерваторий сети INTERMAGNET // Изв. вузов. Приборостроение. 2023. Т. 66. № 12. С. 993–1001. doi: 10.17586/0021-3454-2023-66-12-993-1001.
18. Белашев Б.З. Спектральный анализ данных геофизического мониторинга // Тр. КарНЦ РАН. 2023. № 4. С. 5–15. doi: 10.17076/mat1768.
19. Лыгин И.В., Брагина А.А., Вишняков Д.Д. Подходы прогноза геомагнитных вариаций // Гелиогеофизические исследования. 2023. № 41. С. 42–53.

Software & Systems

doi: 10.15827/0236-235X.142.318-323

2024, 37(3), pp. 318–323

Identifying Pc3 type geomagnetic pulsations from INTERMAGNET data: MATLAB Big Data methods

Anatoly G. Korobeynikov ¹ SPbF IZMIRAN, St. Petersburg, 199034, Russian Federation**For citation**Korobeynikov, A.G. (2024) 'Identifying Pc3 type geomagnetic pulsations from INTERMAGNET data: MATLAB Big Data', *Software & Systems*, 37(3), pp. 318–323 (in Russ.). doi: 10.15827/0236-235X.142.318-323**Article info**

Received: 14.02.2024

After revision: 18.03.2024

Accepted: 25.03.2024

Abstract. MATLAB system is widely used in solving fundamental and applied problems in various subject areas. This work describes using Big Data technology methods implemented in MATLAB to solve the problem of identifying Pc3 type geomagnetic pulsations from data supplied by magnetic observatories included in the international INTERMAGNET network. The solution involves passing the preprocessed data set through a designed infinite pulse bandpass digital filter (infinite impulse response filter – IIR filter) and then removing the nonlinear phase distortion of the filter. The IIR filter is designed using the Zolotarev–Kauer method. The filter parameters are based on the period range of Pc3 geomagnetic pulsations 10÷45 seconds. The choice of a design method is determined by the developed IIR filter characteristics: 1) a balanced behavior of frequency response ripples both in the passband and delay band; 2) relaxing the requirements for the frequency response shape in the delay band that allows a quick transition between the passband and the delay band; 3) the order of the IIR filter designed using the Zolotarev digital filter method and its complexity are minimal. One of the requirements when solving the indicated problem is the condition of using Big Data technologies implemented in MATLAB. This requirement arises from a large amount of initial data: continuous recording of data on the geomagnetic field state through four channels for a whole year with a sampling frequency equal to one second. The obtained results can be used to solve both applied problems, for example, analyzing the space weather state, and fundamental ones, for example, further development of the theory of geomagnetic pulsations. The IIR bandpass filter design and calculations were carried out in MATLAB R2023b.

Keywords: Big Data, INTERMAGNET, MATLAB, Pc3, IIR-filter, geomagnetic ripple, bandpass filter**Acknowledgements.** The work was supported by RNF grant no. 23-27-00011

References

1. Korobeynikov, A.G. (2022) 'Applying artificial neural networks in automatic control systems for magnetic levitation', *Software & Systems*, 35(3), pp. 452–457 (in Russ.). doi: 10.15827/0236-235X.139.452-457.
2. Egorov, E. (2020) 'Simulating the manipulation robot operation in the Matlab Robotics Toolbox software', *Politechnical Student J.*, (01), pp. 1–12 (in Russ.).
3. Korobeynikov, A.G. (2018) 'Processing and analysis of data from the Russian segment of the world network of magnetic observatory INTERMAGNET', *Int. J. of Humanities and Natural Sciences*, (8), pp. 91–98 (in Russ.).
4. Vishnyakov, D.D., Arutyunyan, D.A., Shklyaruk, A.D. (2023) 'Assessment of the planetary index in geomagnetic activity from data of stationary magnetic observatories of the INTERMAGNET network using proprietary software', *Heliogeophysical Research*, (38), pp. 40–45 (in Russ.).
5. Vorobev, A.V., Vorobeva, G.R. (2018) 'Approach to assessment of the relative informational efficiency of INTERMAGNET magnetic observatories', *Geomagnetism and Aeronomy*, 58(5), pp. 625–628 (in Russ.). doi: 10.1134/S0016794018050164.
6. Guglielmi, A.V., Potapov, A.S. (2021) 'Frequency-modulated ULF waves in near-Earth space', *UFN*, 191(5), pp. 475–491 (in Russ.). doi: 10.3367/UFNr.2020.06.038777.
7. Mishin, V.V., Tsegmed, B., Klibanova, Yu.Yu., Kurikalova, M.A. (2020) 'Burst geomagnetic pulsations as indicators of substorm expansion onsets during storms', *JGR Space Phys.*, 125, art. e2020JA028521. doi: 10.1029/2020JA028521.
8. Kulhánek, O. (1976) *Introduction to Digital Filtering in Geophysics*. Elsevier Publ., 168 p. (Russ. ed.: (1981) Moscow, 198 p.).
9. Potthast, R., Walter, A., Rhodin, A. (2019) 'A localized adaptive particle filter within an operational NWP framework', *Monthly Weather Review*, 147, pp. 345–362. doi: 10.1175/MWR-D-18-0028.1.
10. van Leeuwen, P.J., Kunsch, H.R., Nerger, L., Potthast, R., Reich, S. (2019) 'Particle filters for high-dimensional geoscience applications: A review', *Quarterly J. of the Royal Meteorological Society*, 145(723), pp. 2335–2365. doi: 10.1002/qj.3551.
11. Klimova, E.G. (2023) 'A local ensemble data assimilation algorithm for nonlinear geophysical models', *SJNM*, 26(1), pp. 27–42 (in Russ.). doi: 10.15372/SJNM20230103.
12. Moshak, N.N., Rudinskaya, S.R., Gruzdev, A.A. (2023) 'Third platform for informatization and big data', *H&ES Research*, 15(4), pp. 47–59 (in Russ.). doi: 10.36724/2409-5419-2023-15-4-47-59.
13. Solonina, A.I. (2018) *Digital Signal Processing in the MATLAB Mirror*. St. Petersburg, 560 p. (in Russ.).
14. Budunova, K.A., Kravchenko, V.F. (2022) 'Mathematical methods of frequency selective filters synthesis', *Phys. Bases of Instrumentation*, 11(1), pp. 2–21 (in Russ.). doi: 10.25210/jfop-2201-002021.
15. Yulina, A.O., Sinkevich, G.I. (2021) 'The history of the development of the theory of elliptic functions in the works of Abel, Jacobi, Weierstrass, Somov', *TJCSTM*, (3), pp. 79–92 (in Russ.).
16. Korobeynikov, A.G. (2023) 'Applying MATLAB in the design of digital filters for selecting Pc5 geomagnetic pulsations', *Software & Systems*, 36(2), pp. 257–262 (in Russ.). doi: 10.15827/0236-235X.142.257-262.
17. Korobeynikov, A.G. (2023) 'Application of big data methods for comparing data of geomagnetic observatories in the INTERMAGNET network', *J. of Instrument Engineering*, 66(12), pp. 993–1001 (in Russ.). doi: 10.17586/0021-3454-2023-66-12-993-1001.
18. Belashev, B.Z. (2023) 'Spectral analysis of geophysical monitoring data', *Proc. of the Karelian Research Centre of the RAS*, (4), pp. 5–15 (in Russ.). doi: 10.17076/mat1768.
19. Lygin, I.V., Bragina, A.A., Vishnyakov, D.D. (2023) 'Methods for predicting geomagnetic variations', *Heliogeophysical Research*, (41), pp. 42–53 (in Russ.).

Авторы

Коробейников Анатолий Григорьевич¹, д.т.н.,
профессор, зам. директора по науке,
Korobeynikov_A_G@mail.ru

Authors

Anatoly G. Korobeynikov¹, Dr.Sci. (Engineering),
Professor, Deputy Director on Scientific Work,
Korobeynikov_A_G@mail.ru

¹ СПбФ ИЗМИРАН, г. Санкт-Петербург,
199034, Россия

¹ SPbF IZMIRAN, St. Petersburg,
199034, Russian Federation

Программная реализация алгоритмов для создания прототипов баз знаний на основе визуального моделирования и трансформаций

Н.О. Дородных¹, А.Ю. Юрин¹✉

¹ Институт динамики систем и теории управления имени В.М. Матросова СО РАН, г. Иркутск, 664033, Россия

Ссылка для цитирования

Дородных Н.О., Юрин А.Ю. Программная реализация алгоритмов для создания прототипов баз знаний на основе визуального моделирования и трансформаций // Программные продукты и системы. 2024. Т. 37. № 3. С. 324–333. doi: 10.15827/0236-235X.142.324-333

Информация о статье

Группа специальностей ВАК: 2.3.5

Поступила в редакцию: 16.02.2024

После доработки: 07.05.2024

Принята к публикации: 14.05.2024

Аннотация. В работе описывается специализированная веб-ориентированная программная система Knowledge Modeling System (KMS), рассчитанная на экспертов и системных аналитиков. Система обеспечивает возможность построения визуальных концептуальных моделей в форме диаграмм переходов состояний, деревьев событий и отказов, а также автоматического получения кода баз знаний на основе их преобразования. Полученные таким способом базы знаний могут быть применены в качестве прототипов при дальнейшей разработке интеллектуальных систем. Созданная система основана на принципах визуального моделирования и модельных трансформаций. Последний принцип обеспечивает возможность описания соответствия между элементами различных моделей (нотаций) и их преобразования. Для формирования операторов преобразования (трансформации) концептуальных моделей описаны их модели (метамоделли), включающие основные элементы и отношения. Трансформации наглядно представлены в виде набора соответствий между элементами моделей и реализованы на языке общего назначения PHP. В качестве целевых языков определены CLIPS и OWL2 DL. Приводится описание разработанной системы, включая метод создания баз знаний на основе преобразования концептуальных моделей, основные функции, архитектуру. Реализованный метод включает этапы построения концептуальной модели предметной области, представления концептуальных моделей в формате XML, анализа XML-структуры модели, формирования модели онтологии или продукций, модификации полученных знаний, генерации кода базы знаний на целевом языке. Представлены примеры использования системы KMS в контексте создания прототипов баз знаний в области техногенной безопасности в части решения задач диагностирования и прогнозирования технического состояния объектов и систем: описания динамики развития аварии при истечении нефти из резервуара и планирования алгоритма анализа отказа.

Ключевые слова: программная система, визуальное моделирование, база знаний, прототип, продукция, онтология, дерево событий, диаграмма переходов состояний, трансформация, генерация кода

Благодарности. Работа выполнена в рамках госзадания Минобрнауки России по проекту № 121030500071-2

Введение. Интеллектуальные системы поддержки принятия решений представляют собой сложные программные комплексы, ориентированные на решение слабоформализованных прикладных задач. Они широко применяются в таких предметных областях, как медицина [1], финансы, производство, транспорт [2], энергетика, а повышение эффективности их разработки является одной из важных научно-технических проблем [3, 4].

В последнее время для представления знаний в интеллектуальных системах особую популярность получили формализмы, основанные на семантических технологиях, такие как онтологии [5] и графы знаний [6]. Остаются достаточно распространенными и востребованными языки представления знаний (ЯПЗ), обеспечивающие поддержку логических правил (продукций) [7]. При этом разработка как онтологических, так и продукционных баз зна-

ний продолжает оставаться одним из самых сложных и трудоемких этапов при создании интеллектуальных систем, поскольку требует значительных ресурсов и привлечения специалистов различной квалификации. Одним из перспективных средств решения данной проблемы являются методы автоматического извлечения знаний из различных информационных источников (например, БД, электронных документов, таблиц). Как правило, для обработки данных такие методы используют алгоритмы машинного обучения [8]. Однако в ряде случаев обрабатываемые данные не сопровождаются явной семантикой, необходимой для машинной интерпретации своего содержания, а накапливаемая в этих форматах информация часто является либо неструктурированной, либо неунифицированной (форма ее представления не соответствует какому-либо общепринятому стандарту). Это затрудняет активное

применение данных методов на практике. Также важно отметить, что большинство решений, предлагаемых в области разработки баз знаний, направлены на программистов, а вовлечение конечных пользователей (экспертов предметной области, системных аналитиков) в этот процесс весьма ограничено. Поэтому создание новых методов и программных средств, повышающих эффективность построения баз знаний интеллектуальных систем на основе автоматизированного преобразования данных из различных источников информации и направленных на непрограммирующих пользователей, является актуальной и перспективной областью научных исследований.

В данной работе описывается веб-ориентированная программная система Knowledge Modeling System (KMS) (<http://kms.knowledge-core.ru/>), реализующая алгоритмы, которые позволяют в интерактивном режиме создавать концептуальные модели, отражающие знания предметной области с помощью специализированных нотаций. Построенные таким образом визуальные модели (диаграммы) выступают в качестве основного информационного источника при автоматическом формировании кодов баз знаний на определенном ЯПЗ. При этом в качестве целевых платформ, ориентированных на описание логических правил, используются ЯПЗ CLIPS и таблицы решений в формате CSV, в случае применения онтологий – OWL в формате RDF/XML. Разработанная система была использована для прототипирования баз знаний в области техногенной безопасности при решении задач диагностирования и прогнозирования технического состояния объектов и систем [9].

Подходы к разработке баз знаний

Базы знаний могут быть созданы вручную или автоматически на основе информации из различных источников. Как правило, при ручном построении баз знаний используют различные специализированное ПО, автоматизирующее этапы получения, структурирования и представления знаний. Подходы, лежащие в основе данных программных средств, можно разделить на три основные группы.

Текстовые. Обеспечивают прямое взаимодействие пользователя с конструкциями определенного ЯПЗ. Примерами реализации этого подхода для поддержки формализма правил являются средства VISUAL JESS, ClipsWin и DroolsExpert ([\[ase/6.0.0.CR1/drools-expert-docs/html/\]\(https://docs.jboss.org/drools/rel-ase/6.0.0.CR1/drools-expert-docs/html/\)\). Для онтологических баз знаний существуют такие системы, как Protégé и FluentEditor.](https://docs.jboss.org/drools/rel-</p></div><div data-bbox=)

Табличные. Обеспечивают построение пользователем таблиц решений [10, 11] и их возможную автоматическую трансляцию в коды баз знаний.

Графические. Обеспечивают создание визуальных примитивов (например, блоков, стрелок и т.д.), соответствующих элементам баз знаний, с последующей кодогенерацией на определенном ЯПЗ. Этот подход является наиболее перспективным, так как позволяет значительно расширить круг разработчиков баз знаний за счет непрограммирующих пользователей, владеющих навыками визуального моделирования.

Как правило, в графическом подходе используют либо универсальные семантические графовые модели (например, VisiRule (http://www.lpa.co.uk/ind_hom.htm), VIPR, SCg [12]), либо специализированные графические нотации, моделирующие логические и причинно-следственные зависимости (например, URML, RVML [13]). Данное направление обладает достаточно большим количеством различных решений, однако имеет и ряд значимых ограничений. В частности, большинство подходов содержат нестандартные условные обозначения (визуальные артефакты) или вводят ограничения на именование понятий и отношений, что не всегда интуитивно понятно пользователю. Более того, некоторые решения не учитывают специфику создания баз знаний на общепринятых ЯПЗ (например, CLIPS, Jess, Drools, SWRL, OWL) и ориентируются исключительно на свой собственный формат представления знаний.

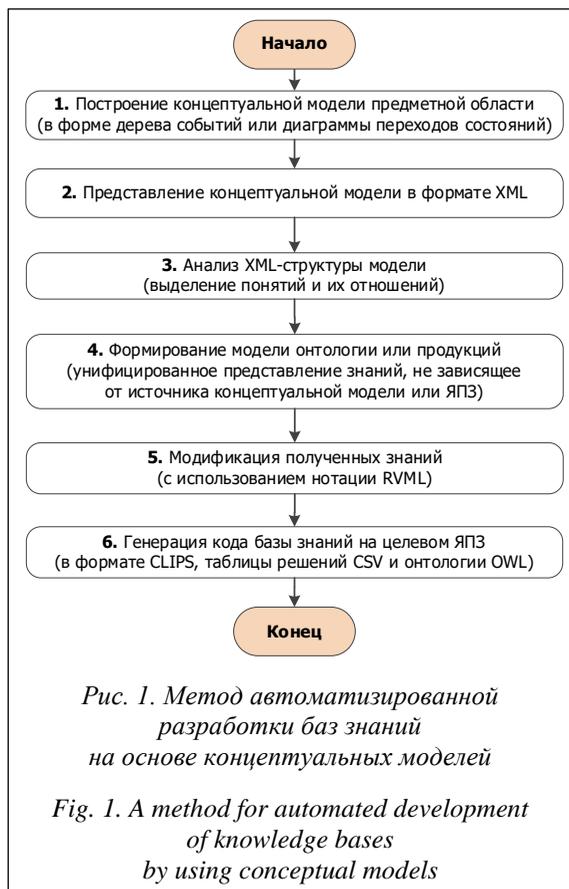
Следует отметить, что очень часто на стадиях извлечения и структурирования знаний применяют различные когнитивные или концептуальные модели: концепт-карты, деревья событий и отказов, диаграммы переходов состояний, диаграммы Исикавы и др. Данные визуальные модели являются достаточно удобным и интуитивно понятным средством представления знаний для специалистов-предметников, при этом существует ПО в виде графических редакторов, реализующих их поддержку. Однако эти редакторы не предназначены для целенаправленной разработки баз знаний и не дают какой-либо возможности для интеграции построенных визуальных моделей в базы знаний на определенном ЯПЗ. Эта особенность также затрудняет практическое использование

построенных моделей для создания баз знаний при разработке интеллектуальных систем.

Таким образом, с целью преодоления названных ограничений в данной работе предлагается специализированное программное средство Knowledge Modeling System (KMS), поддерживающее полный цикл автоматизированной разработки баз знаний, начиная от создания визуальных моделей предметной области и заканчивая кодогенерацией с возможностью модификации и проверки полученного кода.

Реализуемый метод, модели и трансформации

Разработанная KMS основана на принципах технологии PEsOT [14], которая рассматривает процесс создания интеллектуальных систем и баз знаний в контексте последовательного преобразования моделей [15, 16]. В частности, KMS реализует метод автоматизированной разработки баз знаний на основе трансформации концептуальных моделей (рис. 1).



На этапе 1 пользователь (эксперт) строит исходные концептуальные модели, описывающие некоторую предметную область. В ка-

честве исходных концептуальных моделей в KMS используются деревья событий и диаграммы переходов состояний.

Дерево событий – это графическое представление некоторой последовательности событий, которые могут произойти в результате определенного исходного события (начальной точки диаграммы) или условий (ГОСТ Р 54142-2010). Деревья событий обычно используются для анализа и оценки рисков в различных областях, таких как природная и техногенная безопасность, управление проектами, инженерия и т.д.

Модель, описывающая основные элементы данной диаграммы и их отношения, приведена на рисунке 2.

Диаграммы переходов состояний используются для моделирования поведения системы или объекта [17], позволяя визуализировать последовательность состояний и переходов между ними в виде ориентированного графа. Диаграммы переходов состояний активно используются в проектировании и анализе сложных технических и социальных систем.

Основные элементы и отношения данной диаграммы отображены на рисунке 3.

На этапе 2 построенные концептуальные модели (диаграммы) представляются (сериализуются) в формате XML с использованием специально разработанной спецификации. Формат XML удобен для представления данных диаграмм, так как позволяет описывать структуру диаграмм с помощью набора тегов и атрибутов, а также осуществлять их импорт и экспорт.

На этапах 3 и 4 выполняется анализ XML-структуры концептуальных моделей, в результате которого выделяются элементы и отношения между ними. На их основе автоматически формируется модель продукций (правил) или онтологии, при этом пользователь только выбирает необходимую целевую модель. Приведенные модели являются частью внутреннего представления знаний в системе KMS, которая не зависит от реализации конкретного ЯПЗ (например, CLIPS, Jess, Drools, SWRL, OWL и т.д.). Это дает возможность на этапе 5 визуализировать, модифицировать и проверять полученные знания в виде как правил, так и онтологии.

На этапе 6 осуществляется генерация кода баз знаний для целевой платформы путем преобразования сформированной модели продукций или онтологии.

В целом преобразование исходных деревьев событий и диаграмм переходов состояний можно описать с помощью специальных языков трансформации моделей, например TMRL [18].

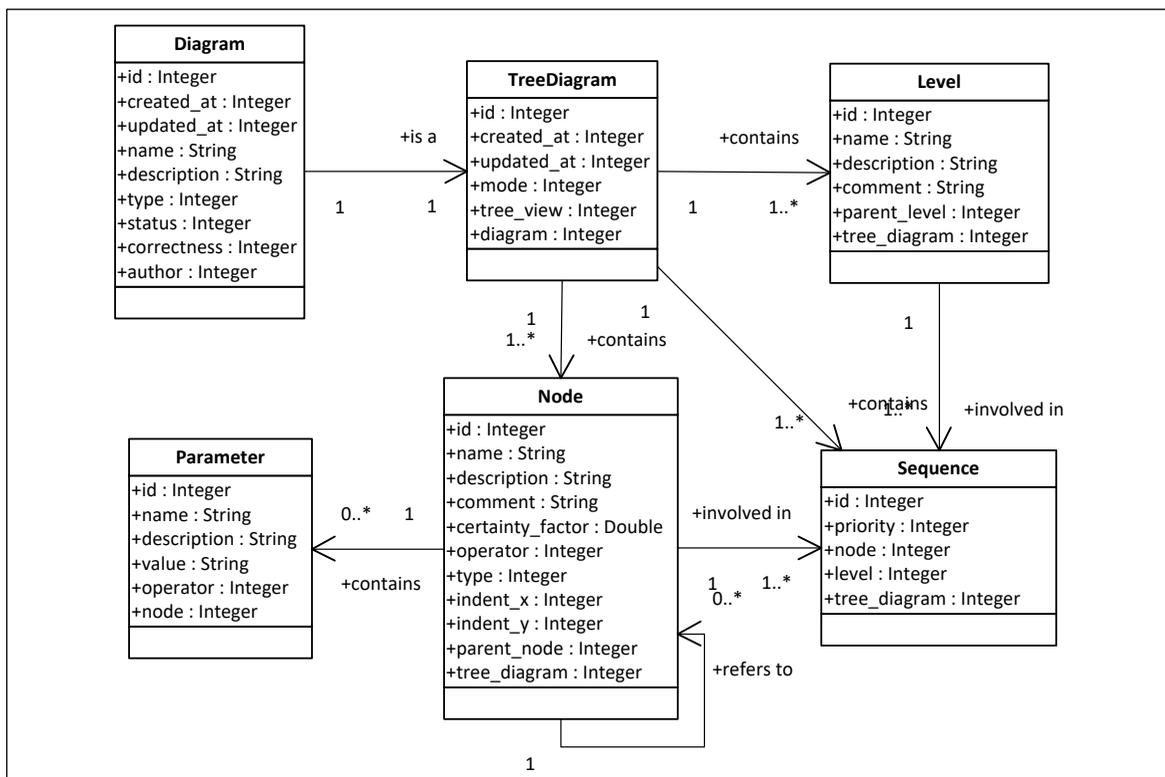


Рис. 2. Модель диаграммы деревьев событий

Fig. 2. Event tree diagram model

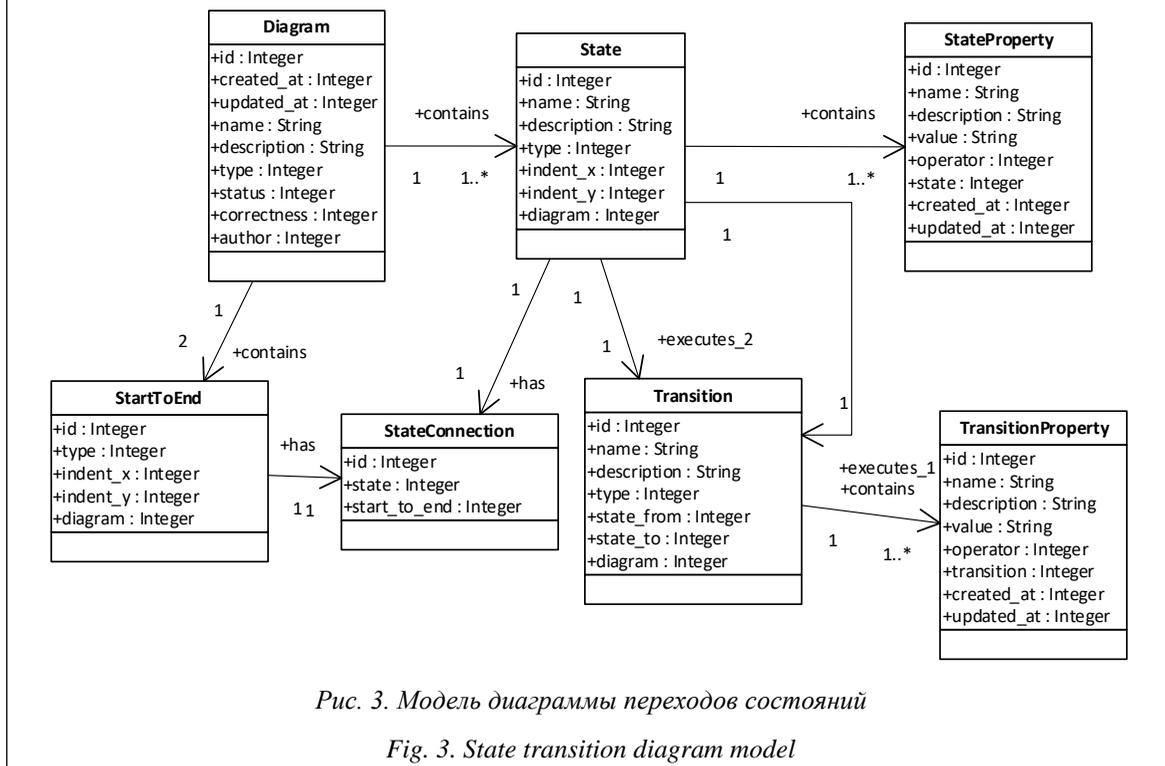


Рис. 3. Модель диаграммы переходов состояний

Fig. 3. State transition diagram model

В данной работе для реализации трансформаций использован язык программирования общего назначения РНР. При этом все трансфор-

мации можно наглядно представить в виде набора соответствий, представленных в таблицах 1, 2.

Таблица 1

Основные соответствия между элементами деревьев событий, модели продукций и продукционной базы знаний на CLIPS

Table 1

Basic correspondences between elements of event trees, a product model and a product knowledge base in the CLIPS KRL

Элемент формата представления		Элемент модели продукций	Оператор CLIPS
деревьев событий	диаграммы переходов состояний		
TreeDiagram	Diagram	Model	-
Node (name)	State (name)	FactTemplate / Fact	Deftemplate
Parameter (name)	StateProperty (name)	Slot	(slot "<name>")
Sequence	Transition (name)	RuleTemplate / Rule	defrule
Node (parent_node)	Transition (state_from)	FactTemplate / Condition	deftemplate / defrule (consequent)
Node (id)	Transition (state_to)	FactTemplate / Action	deftemplate / defrule (antecedent)

Таблица 2

Основные соответствия между элементами деревьев событий, модели онтологии и онтологической базы знаний на OWL2 DL

Table 2

Basic correspondences between elements of event trees, an ontology model and an ontology knowledge base in the OWL2 DL KRL

Формат представления		Элемент модели онтологии	OWL2 DL
деревьев событий	диаграммы переходов состояний		
TreeDiagram	Diagram	Ontology	owl:Ontology
Node	State	Class	owl:Class
Node (name)	State (name)	Class (name)	rdf:ID
Node (parent_node-id)	Transition (name)	Relationship	owl:ObjectProperty
Node (parent_node)	Transition (state_from)	Rhs	owl:ObjectProperty (rdfs:domain)
Node (id)	Transition (state_to)	Lhs	owl:ObjectProperty (rdfs:range)
Parameter	StateProperty	Property	owl:DatatypeProperty
Parameter (name)	StateProperty (name)	Property (name)	owl:DatatypeProperty (rdfs:domain)
Parameter (value)	StateProperty (value)	Property (value)	owl:DatatypeProperty (rdfs:range)
Parameter (description)	StateProperty (description)	Property (description)	rdfs:comment

Полученный таким образом код баз знаний может быть использован в сторонних програмных средствах, например, PKBD (<http://www.knowledge-core.ru/index.php?p=pkbd&lan=ru>) или Protégé, с целью уточнения, модификации и дальнейшего использования при разработке интеллектуальных систем.

Назначение и основные функции

Разработанная KMS предназначена для моделирования знаний некой предметной обла-

сти в форме концептуальных моделей и для создания на их основе прототипов баз знаний (<http://www.swsys.ru/uploaded/image/2024-3/10.jpg>).

Основная цель использования KMS – автоматизировать процессы концептуализации, формализации предметных знаний, а также кодогенерации прототипов баз знаний.

Представим основные функции KMS:

– создание, редактирование, просмотр и удаление пользователей;

- аутентификация и авторизация действий пользователей;
- поддержка интернационализации (английский и русский языки);
- создание, редактирование, просмотр и удаление проектов, в рамках которых ведется работа с различными диаграммами;
- создание, редактирование, просмотр и удаление диаграмм деревьев событий и диаграмм переходов состояний в рамках определенного проекта;
- проверка корректности построенных диаграмм;
- импорт и экспорт диаграмм деревьев событий и диаграмм переходов состояний в виде сериализованных файлов в формате XML;
- импорт онтологий в формате OWL (RDF/XML) в KMS, и их преобразование в диаграммы деревьев событий;
- предоставление REST API для взаимодействия с моделями продукций и онтологий с целью модификации данных моделей средствами других редакторов (RVML, KBDS);
- экспорт (генерация) построенных диаграмм в файлы форматов ЯПЗ CLIPS и OWL (RDF/XML), а также в электронные таблицы формата CSV;
- дополнительная возможность генерации спецификации виртуального ассистента на основе построенных диаграмм.

Архитектура и особенности реализации

Разработанное веб-ориентированное программное средство KMS обладает клиент-сер-

верной архитектурой (рис. 4), в которую входят следующие основные модули:

- подсистема построения концептуальных моделей, включающая два визуальных редактора: редактор классических и расширенных диаграмм деревьев событий и редактор диаграмм переходов состояний;
- модуль администрирования (авторизация/аутентификация);
- модуль управления универсальным уровнем представления знаний (предоставляет доступ к моделям продукций и онтологии);
- генераторы программных кодов и спецификаций баз знаний;
- модули импорта и экспорта диаграмм;
- модуль взаимодействия с внешними системами через REST API.

Программная реализация осуществлена с применением языков PHP 8 и JavaScript с использованием стилей Bootstrap 5. При этом визуальные редакторы реализованы с использованием библиотеки JsPlumb Toolkit. Для хранения данных применяется СУБД PostgreSQL 15. Исходный код системы открыт и доступен на GitHub (<https://github.com/Lab42-Team/kms>).

KMS поддерживает работу трех типов пользователей: *администратор* – зарегистрированный пользователь, который может производить любые действия в рамках KMS; *обычный пользователь* – зарегистрированный пользователь, который может создавать и изменять различные диаграммы в рамках определенного проекта; *гость* – незарегистрированный пользователь, который может только просматривать от-

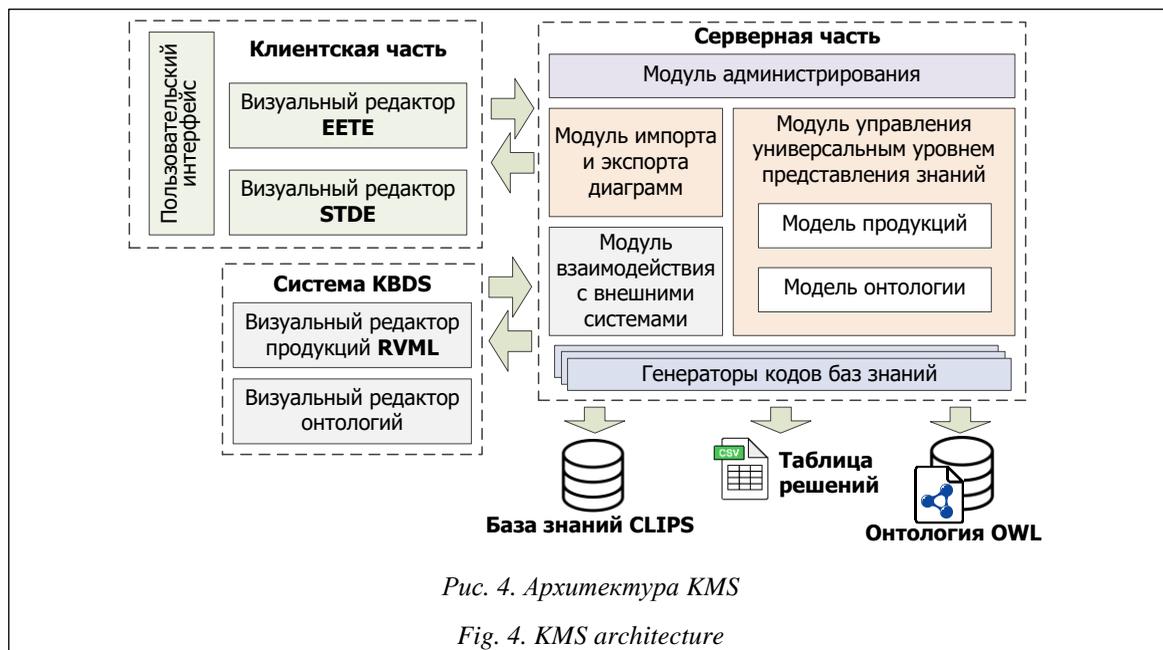


Рис. 4. Архитектура KMS

Fig. 4. KMS architecture

крытые (публичные) диаграммы и общую информацию по KMS.

Пример применения

Одним из примеров применения KMS является создание прототипов баз знаний в области техногенной безопасности при решении задач диагностирования и прогнозирования технического состояния объектов и систем [9].

Проблема оценки и повышения безопасности промышленных объектов с течением времени сохраняет свою актуальность, что обусловлено высокими темпами старения (деградации) оборудования во многих отраслях промышленности, превышающими темпы его замены и модернизации, как по субъективным, так и по объективным причинам. Значительное повышение надежности и безопасности нефтехимического оборудования можно обеспечить путем создания и активного использования методов и средств искусственного интеллекта, в частности, интеллектуальных систем поддержки принятия решений.

В рамках разработки прототипов баз знаний для ИАС «Экспертиза ПБ» были построены деревья событий, описывающие динамику развития аварий, в частности, истечение нефти (рис. 5).

В результате преобразования были получены наборы шаблонов, фактов и правил (<http://www.swsys.ru/uploaded/image/2024-3/11.jpg>).

После уточнения построенных диаграмм и тестовых прогонов (отладки) в PKBD сгенерирован программный код базы знаний на CLIPS, который затем был перенесен в ИАС «Экспертиза ПБ». Приведем фрагмент полученного кода:

```

;***** Templates *****
(deftemplate Event-0
  (slot name (default "РАЗРУШЕНИЕ СОСЕДНЕГО ОБОРУДОВАНИЯ"))
  (slot veroyatnost)
)
;***** Rules *****
(defrule 0-Event-0->Event-0+Event-0
(declare (salience 1))
(Event-0 ;Event-0
  (name "ФАКЕЛЬНОЕ ГОРЕНИЕ СТРУИ")
  (veroyatnost "0.04")
)
=>
(assert
  (Event-0 ;Event-0
    (name "РАЗРУШЕНИЕ СОСЕДНЕГО ОБОРУДОВАНИЯ")
    (veroyatnost "0.02")
  )
  ...

```

В рамках решения задачи информационного моделирования области анализа отказов технических систем была разработана специальная онтологическая база знаний. В частности, на рисунке 6 приведена визуальная модель в форме диаграммы переходов состояний (для ее построения использована разработанная авторами система), описывающая алгоритм анализа отказа.

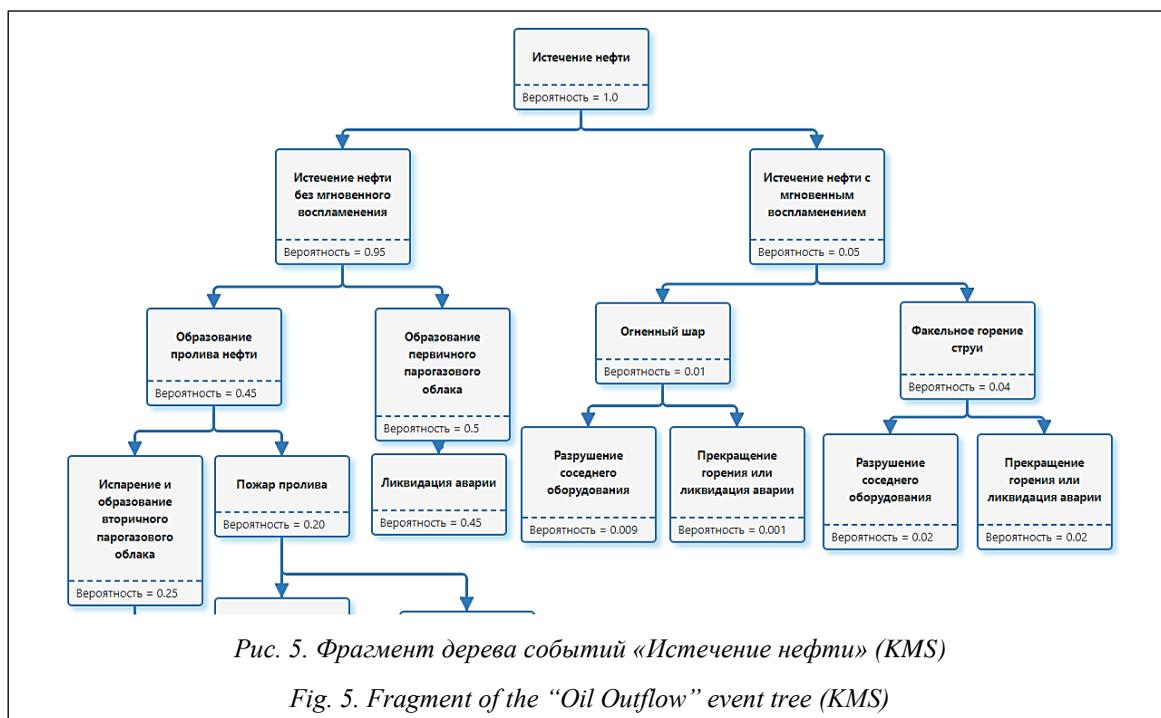


Рис. 5. Фрагмент дерева событий «Истечение нефти» (KMS)

Fig. 5. Fragment of the “Oil Outflow” event tree (KMS)

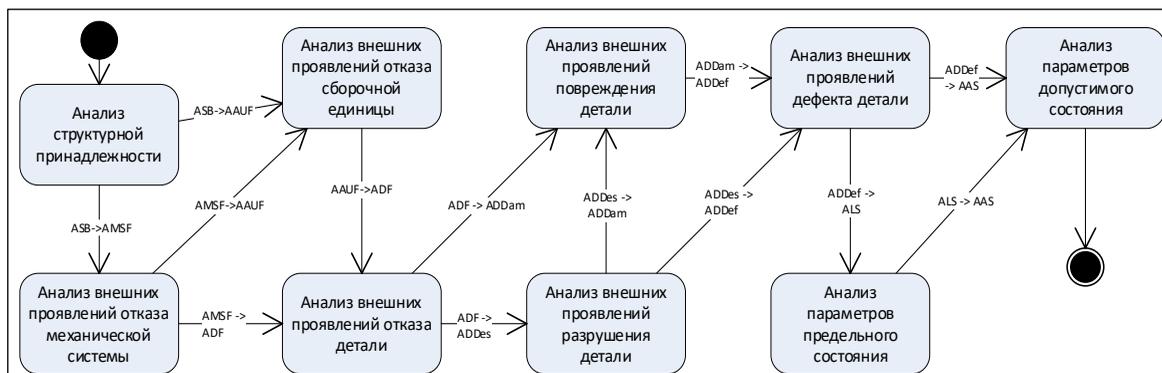


Рис. 6. Фрагмент диаграммы переходов состояний, описывающий алгоритм анализа отказов (KMS)

Fig. 6. Fragment of the "failure analysis algorithm" state transition diagram (KMS)

В результате выполнения преобразования данной диаграммы синтезирован код онтологии в формате OWL2 DL:

```

<owl:Class rdf:ID="Состояние"/>
<owl:Class rdf:ID="Переход"/>

<owl:Class
rdf:ID="АнализВнешнихПроявленийОтказаДе
тали">
  <rdfs:subClassOf rdf:re-
source="Состояние"/>
</owl:Class>

<owl:Class
rdf:ID="АнализВнешнихПроявленийПоврежде
нияДетали">
  <rdfs:subClassOf rdf:re-
source="Состояние"/>
</owl:Class>
<owl:DatatypeProperty
rdf:ID="свойстваВведены">
  <rdfs:domain rdf:re-
source="Переход"/>
  <rdfs:range rdf:re-
source="http://www.w3.org/2001/XMLSchema
#string"/>
</owl:DatatypeProperty>
...
    
```

При этом все состояния интерпретировались как классы (owl:Class), а переходы как объекты-свойства (owl:ObjectProperty). Свойства (характеристики) переходов выражались через свойства-значения (owl:DatatypeProperty), которые задаются для специального служебного класса перехода.

Заключение

Актуальность проблемы автоматизации разработки интеллектуальных систем обуслови-

вает необходимость создания нового методологического и программного инструментария, направленного на повышение эффективности процессов извлечения, структурирования и формализации знаний.

В данном контексте автоматизация создания баз знаний интеллектуальных систем остается перспективной областью научных исследований. Особенно, когда речь идет о вовлечении в данный процесс непрограммирующих пользователей, а также об использовании уже накопленной ранее информации, представленной, в частности, в форме визуальных концептуальных моделей.

В данной работе предлагается веб-ориентированное программное средство KMS, реализующее алгоритмы визуального моделирования знаний предметной области в форме диаграмм переходов состояний и деревьев событий, которые в дальнейшем являются основным информационным источником для автоматического формирования кодов баз знаний как продукционного, так и онтологического типа. Данные особенности можно выделить в качестве преимуществ системы, так как они способствуют более полному вовлечению в процесс разработки непрограммирующих пользователей и минимизируют ошибки программирования.

В качестве примера использования KMS рассмотрено создание прототипов баз знаний при решении задач диагностирования и прогнозирования технического состояния объектов и систем в нефтехимии.

Список литературы

1. Saibene A., Assale M., Giltri M. Expert systems: Definitions, advantages and issues in medical field applications. Expert Systems with Applications, 2021, vol. 177, art. 114900. doi: 10.1016/j.eswa.2021.114900.

2. Wagner W.P. Trends in expert system development: A longitudinal content analysis of over thirty years of expert system case studies. *Expert Systems with Applications*, 2017, vol. 76, pp. 85–96. doi: 10.1016/j.eswa.2017.01.028.
3. Сухих Н.Н., Рукавишников В.Л. Экспертные системы – средство информационной поддержки принятия решений экипажем самолета // *Изв. вузов*. 2022. № 2. С. 19–25.
4. Гаврилова Т.А., Кудрявцев Д.В., Муромцев Д.И. Инженерия знаний. Модели и методы. СПб: Лань, 2023. 324 с.
5. Karabulut E., Pileggi S.F., Groth P., Degeler V. Ontologies in digital twins: A systematic literature review. *Future Generation Computer Systems*, 2024, vol. 153, pp. 442–456. doi: 10.1016/j.future.2023.12.013.
6. Ji S., Pan S., Cambria E., Martinen P., Yu P.S. A survey on knowledge graphs: Representation, acquisition and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, vol. 33, no. 2, pp. 494–514. doi: 10.1109/TNNLS.2021.3070843.
7. Nowak-Brzezińska A., Wakulicz-Deja A. Exploration of rule-based knowledge bases: A knowledge engineer's support. *Information Sciences*, 2019, vol. 485, pp. 301–318. doi: 10.1016/j.ins.2019.02.019.
8. Martinez-Rodriguez J.L., Hogan A., Lopez-Arevalo I. Information extraction meets the semantic web: A survey. *Semantic Web*, 2020, vol. 11, pp. 255–335. doi: 10.3233/SW-180333.
9. Берман А.Ф., Николайчук О.А., Павлов А.И., Юрин А.Ю., Кузнецов К.А. Информационно-аналитическая поддержка экспертизы промышленной безопасности объектов химии, нефтехимии и нефтепереработки // *ХНГМ*. 2018. № 8. С. 30–36.
10. Димитров В.П., Борисова Л.В., Хубиян К.Л. Моделирование знаний в задаче поиска причин неисправностей // *Инженерные технологии и системы*. 2021. Т. 31. № 3. С. 364–379. doi: 10.15507/2658-4123.031.202103.364-379.
11. Nalepa G.J., Kluza K. UML representation for rule-based application models with XTT2-based business rules. *Int. J. of Software Eng. and Knowledge Eng.*, vol. 22, no. 4, pp. 485–524. doi: 10.1142/S021819401250012X.
12. Голенков В.В., Шункевич Д.В., Давыденко И.Т., Гракова Н.В. Принципы организации и автоматизации процесса разработки семантических компьютерных систем // *OSTIS*. 2019. № 3. С. 53–90 (на англ.).
13. Дородных Н.О., Юрин А.Ю., Коршунов С.А. Средства поддержки моделирования логических правил в нотации RVML // *Программные продукты и системы*. 2018. Т. 31. № 4. С. 667–672. doi: 10.15827/0236-235X.124.667-672.
14. Yurin A.Yu. Technology for prototyping expert systems based on transformations (PESoT): A method. *CEUR Workshop Proc. Proc. III Sci.-Pract. Workshop Inform. Tech.: Algorithms, Models, Systems*, 2020, vol. 2677, pp. 36–50.
15. Brambilla M., Cabot J., Wimmer M. *Model-driven software engineering in practice*, Springer Publ., 2017, 280 p. doi: 10.2200/s00441ed1v01y201208swe001.
16. Ogunyomi B., Rose L.M., Kolovos D.S. Incremental execution of model-to-text transformations using property access traces. *Software & System Modeling*, 2019, vol. 18, pp. 367–383. doi: 10.1007/s10270-018-0666-5.
17. Хопкрофт Дж., Мотвани Р., Ульман Дж. Введение в теорию автоматов, языков и вычислений; [пер. с англ.]. М.: Вильямс, 2008. 528 с.
18. Дородных Н.О., Юрин А.Ю. Язык для описания моделей трансформаций. *CEUR Workshop Proc. Proc. I Sci.-Pract. Workshop Inform. Tech.: Algorithms, Models, Systems*, 2018, vol. 2221, pp. 70–75 (на русс.).

Software & Systems

doi: 10.15827/0236-235X.142.324-333

2024, 37(3), pp. 324–333

Software implementation of algorithms for knowledge base prototyping with visual modelling and transformations

Nikita O. Dorodnykh¹, Aleksandr Yu. Yurin¹✉¹Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences, Irkutsk, 664033, Russian Federation

For citation

Dorodnykh, N.O., Yurin A.Yu. (2024) 'Software implementation of algorithms for knowledge base prototyping with visual modelling and transformations', *Software & Systems*, 37(3), pp. 324–333 (in Russ.). doi: 10.15827/0236-235X.142.324-333

Article info

Received: 16.02.2024

After revision: 07.05.2024

Accepted: 14.05.2024

Abstract. The paper describes a special purpose web-oriented software system called Knowledge Modeling System (KMS) designed for experts and system analysts. The system enables building visual conceptual models in the form of state transition diagrams, event and failure trees, as well as automatic generation of knowledge base code based on their transformation. The obtained knowledge bases can become prototypes in further development of intelligent systems. The created system bases on the principles of visual modelling and model transformations. The latter principle enables describing the correspondence between the elements of different models (notations) and their transformation. Forming transformation operators for conceptual models involves describing their models (metamodels) that include basic elements and relations. The paper shows transformations as a set of correspondences between model elements, they are implemented in the general-purpose language PHP. CLIPS and OWL2 DL are the target languages. The paper describes the developed system, including the method of creating knowledge bases using the transformation of conceptual models, main functions and architecture. The implemented method includes the stages of building a subject area conceptual model, XML representation of

conceptual models, analysis of XML-structure of the model, formation of ontology model or products, modification of the obtained knowledge, generation of knowledge base code in the target language. The authors of the paper present examples of using KMS in terms of creating knowledge bases in the field of technological safety. The examples show how to solve the problems of diagnosing and forecasting the technical state of objects and systems for describing accident dynamics when oil flows out of the reservoir, as well as planning the failure analysis algorithm.

Keywords: software, visual modeling, knowledge base, prototype, rules, ontology, event tree, state transition diagram, model transformation, code generation

Acknowledgements. The work was carried out within the framework of the state assignment of the Ministry of Education and Science of the Russian Federation under the project No. 121030500071-2

References

1. Saibene, A., Assale, M., Giltri, M. (2021) 'Expert systems: Definitions, advantages and issues in medical field applications', *Expert Systems with Applications*, 177, art. 114900. doi: 10.1016/j.eswa.2021.114900.
2. Wagner, W.P. (2017) 'Trends in expert system development: A longitudinal content analysis of over thirty years of expert system case studies', *Expert Systems with Applications*, 76, pp. 85–96. doi: 10.1016/j.eswa.2017.01.028.
3. Sukhikh, N.N., Rukavishnikov, V.L. (2022) 'Expert systems: Means of information support for aircraft crew decision making', *Russ. Aeronautics*, 65, pp. 243–250.
4. Gavrilova, T.A., Kudryavtsev, D.V., Muromtsev, D.I. (2023) *Knowledge Engineering. Models and Methods*. St. Petersburg, 324 p. (in Russ.).
5. Karabulut, E., Pileggi, S.F., Groth, P., Degeler, V. (2024) 'Ontologies in digital twins: A systematic literature review', *Future Generation Computer Systems*, 153, pp. 442–456. doi: 10.1016/j.future.2023.12.013.
6. Ji, S., Pan, S., Cambria, E., Marttinen, P., Yu, P.S. (2022) 'A survey on knowledge graphs: Representation, acquisition and applications', *IEEE Transactions on Neural Networks and Learning Systems*, 33(2), pp. 494–514. doi: 10.1109/TNNLS.2021.3070843.
7. Nowak-Brzezińska, A., Wakulicz-Deja, A. (2019) 'Exploration of rule-based knowledge bases: A knowledge engineer's support', *Information Sciences*, 485, pp. 301–318. doi: 10.1016/j.ins.2019.02.019.
8. Martinez-Rodriguez, J.L., Hogan, A., Lopez-Arevalo, I. (2020) 'Information extraction meets the semantic web: A survey', *Semantic Web*, 11, pp. 255–335. doi: 10.3233/SW-180333.
9. Berman, A.F., Nikolaychuk, O.A., Pavlov, A.I., Yurin, A.Yu., Kuznetsov, K.A. (2018) 'Informational and analytical support of industrial safety expert review of chemical and petrochemical objects', *Chemical and Petroleum Engineering*, (8), pp. 30–36 (in Russ.).
10. Dimitrov, V.P., Borisova, L.V., Khubiyan, K.L. (2021) 'Knowledge modeling in troubleshooting', *Engineering Technologies and Systems*, 31(3), pp. 364–379 (in Russ.). doi: 10.15507/2658-4123.031.202103.364-379.
11. Nalepa, G.J., Kluza, K. (2012) 'UML representation for rule-based application models with XTT2-based business rules', *Int. J. of Software Eng. and Knowledge Eng.*, 22(4), pp. 485–524. doi: 10.1142/S021819401250012X.
12. Golenkov, V., Shunkevich, D., Davydenko, I., Grakova, N. (2019) 'Principles of organization and automation of the semantic computer systems development', *OSTIS*, (3), pp. 53–90.
13. Dorodnykh, N.O., Yurin, A.Yu., Korshunov, S.A. (2018) 'Support tools for modeling logical rules in the RVML notation', *Software & Systems*, 31(4), pp. 667–672 (in Russ.). doi: 10.15827/0236-235X.124.667-672.
14. Yurin, A.Yu. (2020) 'Technology for prototyping expert systems based on transformations (PESoT): A method', *CEUR Workshop Proc. Proc. III Sci.-Pract. Workshop Inform. Tech.: Algorithms, Models, Systems*, 2677, pp. 36–50.
15. Brambilla, M., Cabot, J., Wimmer, M. (2017) *Model-driven Software Engineering in Practice*. Springer Publ., 280 p. doi: 10.2200/s00441ed1v01y201208swe001.
16. Ogunyomi, B., Rose, L.M., Kolovos, D.S. (2019) 'Incremental execution of model-to-text transformations using property access traces', *Software & System Modeling*, 18, pp. 367–383. doi: 10.1007/s10270-018-0666-5.
17. Hopcroft, J.E., Motwani, R., Ullman, J.D. (2001) *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley Publ., 521 p. (Russ. ed.: (2018) Moscow, 528 p.).
18. Dorodnykh, N.O., Yurin, A.Yu. (2018) 'A domain-specific language for transformation models', *CEUR Workshop Proc. Proc. I Sci.-Pract. Workshop Inform. Tech.: Algorithms, Models, Systems*, 2221, pp. 70–75 (in Russ.).

Авторы

Дородных Никита Олегович¹, к.т.н.,
старший научный сотрудник, nikidorny@icc.ru
Юрин Александр Юрьевич¹,
д.т.н., доцент, ведущий научный сотрудник,
зав. лабораторией, iskander@icc.ru

¹ Институт динамики систем
и теории управления
имени В.М. Матросова СО РАН,
г. Иркутск, 664033, Россия

Authors

Nikita O. Dorodnykh¹, Cand. of Sci. (Engineering),
Senior Researcher, nikidorny@icc.ru
Aleksandr Yu. Yurin¹, Dr.Sci. (Engineering),
Associate Professor, Leading Researcher,
Head of Laboratory, iskander@icc.ru

¹ Matrosov Institute for System Dynamics
and Control Theory of Siberian Branch
of Russian Academy of Sciences,
Irkutsk, 664033, Russian Federation

Организация хранения данных функционирования объектов киберфизических систем

Д.А. Коростелев^{1,2}, А.Г. Подвесовский^{1,2}✉, А.А. Захарова¹

¹ Институт проблем управления им. В.А. Трапезникова РАН,
г. Москва, 117997, Россия

² Брянский государственный технический университет,
г. Брянск, 241035, Россия

Ссылка для цитирования

Коростелев Д.А., Подвесовский А.Г., Захарова А.А. Организация хранения данных функционирования объектов киберфизических систем // Программные продукты и системы. 2024. Т. 37. № 3. С. 334–343. doi: 10.15827/0236-235X.142.334-343

Информация о статье

Группа специальностей ВАК: 2.3.8

Поступила в редакцию: 16.04.2024

После доработки: 05.05.2024

Принята к публикации: 14.05.2024

Аннотация. В статье представлены результаты исследований в области организации и программной поддержки хранения данных функционирования объектов киберфизических систем, получаемых в ходе экспериментов. Определены основные особенности их хранения: использование специализированных форматов, хранение больших объемов данных разной степени структурированности, поддержка универсальных структурированных форматов данных, хранение сведений о структуре самих экспериментов и сценариях их проведения. Проанализированы технологии построения систем хранения и специализированные форматы данных, учитывающие эти особенности. Предложен подход к построению программной системы хранения и предобработки экспериментальной информации о функционировании объектов киберфизической системы. Архитектура и особенности реализации системы рассмотрены на примере хранения данных экспериментов с группами совместно действующих беспилотных мобильных средств при решении задач перемещения в пространстве и транспортировки грузов. Архитектура предполагает наличие серверного приложения и веб-клиента, взаимодействующего с ним с помощью API-интерфейса, построенного на основе принципов REST. В состав серверного приложения также входят подсистема предобработки результатов экспериментов и конвертер форматов источников данных. Для хранения экспериментальной информации предложено использовать двухуровневую структуру модели данных. Верхний уровень предназначен для хранения информации о структуре и параметрах экспериментов и реализован в виде БД в системе PostgreSQL. На нижнем уровне осуществляется хранение потоковой информации, получаемой от первичных источников данных. Данный уровень реализован в виде облачного хранилища, построенного на основе Яндекс Облака с использованием технологии S3. Результаты апробации разработанной системы при проведении реальных экспериментов подтвердили обоснованность выбранных архитектурных решений. Практическая значимость выполненных исследований и разработок заключается в развитии подходов к построению масштабируемых программных систем хранения данных функционирования объектов киберфизических систем, что способствует систематизации задач и сценариев управления этими объектами при разработке новых технологий и инструментов управления.

Ключевые слова: киберфизическая система, объекты киберфизических систем, экспериментальные данные, форматы данных, модель данных, СУБД, REST API, S3-хранилище

Благодарности. Исследование выполнено при поддержке РФФИ, проект № 23-19-00342, <https://rscf.ru/project/23-19-00342/>

Введение. Одно из ключевых направлений цифровизации любой сферы деятельности связано с развитием и повсеместным внедрением *киберфизических систем* (КФС). В общем случае под КФС понимаются комплексные системы, объединяющие информационно-вычислительные и физические компоненты, находящиеся в тесном взаимодействии между собой и функционирующие посредством интеграции информационно-коммуникационных, вычислительных и физических технологий. Примерами КФС могут служить различные технические устройства и комплексы: робототехнические системы, беспилотные транспортные системы и средства, системы умного города, сенсорные

системы и др. Характерной чертой современных КФС является взаимодействие большого числа объектов различных типов, как полностью автономных, так и частично автоматизированных и управляемых операторами. За счет автономности объектов КФС, а также интеллектуальности и распределенного характера управления ими становятся возможными решение принципиально новых задач автоматизации производственных и технологических процессов, а также значительное упрощение, ускорение и удешевление подходов к решению.

С другой стороны, развитие индустрии КФС, повышение их технического уровня и уровня интеллектуальности обуславливают актуаль-

ность задач управления такими системами, включая управление как отдельными объектами КФС, так и группой взаимодействующих объектов, в том числе в условиях неполноты и разномасштабности информации, а также распределенности и разнородности источников ее получения.

Исследования в области создания новых и совершенствования существующих подходов, методов и средств управления объектами КФС неизбежно связаны с необходимостью проведения большого количества и вычислительных, и натуральных экспериментов. При этом основу управления объектами КФС всегда составляют процессы сбора, обработки и анализа разнородной информации, получаемой непосредственно от самих объектов, а также от внешних источников данных, осуществляющих мониторинг их работы. В подобной ситуации становятся важными задачи накопления и упорядочения информации, получаемой от объектов КФС и в ходе постановки экспериментов, и при выполнении реальных заданий.

Таким образом, необходим инструмент, позволяющий осуществлять сбор, предварительную обработку и хранение информации, получаемой от объектов КФС в ходе выполнения экспериментов с ними. Важным требованием к указанному инструменту является поддержка набора форматов данных, учитывающих особенности представления и хранения такой информации и обеспечивающих возможность ее дальнейшего использования для обработки и анализа в рамках создания методов и построения систем управления КФС.

Анализ особенностей хранения экспериментальной информации, получаемой от объектов КФС

С точки зрения процессов сбора и хранения информации КФС можно рассматривать как совокупность мобильных и стационарных источников разнородных данных, относящихся к различным областям. Для хранения таких данных, как правило, существуют специализированные форматы, традиционно применяемые в соответствующей области. Так, одним из наиболее распространенных форматов хранения данных в КФС, основанных на беспилотных объектах, является ROS Bag. Данный формат используется в системе ROS и обеспечивает хранение данных изображений, видеопотоков, лидарных данных и других сообщений от различных сенсоров и устройств [1, 2]. Помимо

этого, для хранения изображений и видеопотоков широко используются специально разработанные для этого графические и видеоформаты. Применительно к КФС можно отдельно отметить формат GeoTIFF [3, 4], предназначенный для хранения изображений с геопространственной привязкой – этот формат может применяться для хранения данных беспилотных транспортных систем, работающих в таких приложениях, как картография и навигация.

Таким образом, можно выделить *первую особенность* хранения экспериментальной информации, получаемой от объектов КФС, – возможность использования специализированных форматов, применяемых в соответствующих областях.

Информация, получаемая от КФС, характеризуется большими объемами данных различной степени структурированности с преобладанием полуструктурированных и неструктурированных данных. Для хранения подобных данных существуют специально разработанные подходы и форматы хранения [5]. Следует отметить формат HDF5. Он предназначен для хранения больших объемов данных с поддержкой иерархической структуры и системы метаданных. Известен опыт успешного применения данного формата для хранения информации, которая получается с датчиков беспилотных аппаратов [6, 7]. Кроме того, в качестве средств хранения больших объемов информации можно указать облачные хранилища, построенные на базе технологий AWS S3 (<https://aws.amazon.com/s3/>) и характеризующиеся высокой доступностью и масштабируемостью. Также популярным и востребованным подходом хранения данных сложной структуры и больших объемов является использование NoSQL СУБД [8, 9].

С учетом этого *вторая особенность* хранения экспериментальной информации, получаемой от объектов КФС, состоит в необходимости хранения больших объемов данных разной степени структурированности.

Наряду со специализированными форматами представления данных, получаемых от сенсоров и устройств, для хранения данных КФС также требуются универсальные структурированные форматы данных, такие как CSV, XML, JSON [10]. Они активно используются при анализе данных, а также удобны для передачи и обмена данными. Это обуславливает *третью особенность* хранения экспериментальной информации, получаемой от объектов КФС, – необходимость поддержки универ-

сальных структурированных форматов данных.

В контексте хранения информации, получаемой от объектов КФС при проведении экспериментов с ними, приходится сталкиваться с *четвертой особенностью* – необходимость хранения сведений о структуре самого эксперимента (сценариях его проведения, участвующих в нем объектах, используемых источниках данных и проч.).

Учитывая высокую степень разнородности информации, используемой в управлении объектами КФС, многообразие механизмов ее получения, а также форматов представления и хранения, а также принимая во внимание недостаток универсальных программных инструментов, поддерживающих хранение и преобразование такой информации при проведении экспериментов с объектами КФС, можно сделать вывод об актуальности исследований, направленных на создание методов и средств хранения и преобразования экспериментальной информации, получаемой от объектов КФС и используемой для моделирования сценариев управления этими объектами.

Фактически речь идет о проектировании и разработке системы поддержки хранения экспериментальной информации, получаемой от объектов КФС, с учетом выделенных особенностей хранения такой информации, способов ее получения и механизмов взаимодействия с данными.

Отдельно следует отметить требование, связанное с обеспечением удобства использования данных при их автоматизированной обработке, суть которого заключается в обеспечении возможности оперировать универсальными форматами данных при их анализе. Иными словами, требование состоит в предоставлении дополнительных инструментов преобразования информации путем ее конвертации в соответствующие форматы.

Предметная область и терминология

Описание предлагаемого подхода к хранению экспериментальной информации, получаемой от объектов КФС, и программных решений, реализующих данный подход, проведем на примере КФС, представляющих собой группы совместно действующих беспилотных мобильных средств различного уровня автономности, создаваемые для выполнения некой задачи или комплекса задач. В качестве задач рассмотрим перемещение мобильных средств

в пространстве и транспортировку грузов с их помощью.

Введем ряд понятий, которые далее будут использованы для выделения сущностей, необходимых при построении модели данных.

Объекты КФС, задействованные в решении целевой задачи, будем называть *актерами*. Примерами акторов в контексте рассматриваемой предметной области могут являться беспилотные мобильные средства различных типов (летательные аппараты, платформы и др.), объекты инфраструктуры (погрузчики, зарядные станции) и т.д. Предполагается, что каждый актер имеет некоторый уровень автономности, который может зависеть от конкретной выполняемой задачи и от характера взаимодействия с другими актерами.

Объединение задач, решаемых группой акторов, по некоторому целевому признаку будем называть *миссией*. Примером миссии может быть транспортировка разнородных грузов в маршрутной сети, содержащей источники и пункты назначения, с использованием комплекса беспилотных летательных аппаратов и наземных транспортных средств [11, 12].

Каждая миссия характеризуется своей *структурой*, определяемой типом и характером взаимосвязи решаемых задач, и *набором параметров*, которые можно задавать при планировании миссии, а также при необходимости варьировать в процессе ее исполнения.

Планом миссии будем называть ее структурное представление, в котором определены все параметры. В общем случае план миссии строится с учетом ее цели, начальных условий, ограничений и множества акторов. При этом построение плана осуществляется на двух уровнях: планирование заданий отдельным актерам (уровень задания, исполнительный) и планирование взаимодействия между актерами (уровень миссии, тактический). Также при построении плана могут быть заданы механизмы оценки эффективности выполнения миссии в целом или ее отдельных составляющих.

Под *сценарием управления* понимаем набор взаимосвязанных управляющих воздействий, направленных на реализацию плана миссии, предназначенный для обеспечения эффективного решения комплекса задач, с которыми связана миссия.

Наконец, *эксперимент* – специально организованная миссия, в рамках которой при реализации определенного сценария управления осуществляется сбор данных, характеризующих состояние и функционирование акторов,

а также объектов внешней среды в процессе исполнения миссии. Таким образом, для эксперимента план миссии и сценарии управления дополняются требованиями к составу, источникам и форматам собираемых в ходе эксперимента данных.

Проектирование и реализация программной системы

Для хранения и предобработки данных функционирования объектов КФС разработана программная система на основе клиент-серверной архитектуры (рис. 1). В качестве серверной подсистемы выступает .NET-приложение, предоставляющее для взаимодействия с ней API-интерфейс, построенный с использованием принципов REST (<https://restapitutorial.ru/>).

Применение REST API позволяет в полной мере задействовать возможности протокола HTTP за счет применения различных поддерживаемых им методов для разных типов действий над сущностями, используемыми в приложении. Основные методы HTTP:

- GET для получения как сведений о конкретной сущности, так и списка соответствующих сущностей с возможностью фильтрации и постраничной навигации; результат запроса

представляет собой описание одной или нескольких сущностей в JSON-формате;

- POST для создания новых сущностей; описание сущности представляется в JSON-формате в теле запроса;

- PUT для обновления уже имеющихся сущностей; обновленное описание сущности представляется в JSON-формате в теле запроса;

- DELETE для удаления сущности; идентификатор удаляемой сущности передается как часть URL в запросе.

Для каждой сущности используется свой базовый адрес URL, что позволяет структурировать REST API, обеспечивая тем самым более простое и понятное взаимодействие с ним со стороны веб-клиента. Полное описание API доступно по ссылке <https://cps-control.ru/swagger/index.html>.

Веб-клиент представляет собой одностраничное веб-приложение, выполненное на базе платформы Angular в виде набора HTML-страниц, JS-скриптов и таблиц стилей CSS (<http://www.swsys.ru/uploaded/image/2024-3/12.jpg>). Веб-клиент изначально загружается с сервера и выполняется в браузере исследователя. Взаимодействие веб-клиента и REST API осуществляется с помощью указанных выше видов асинхронных HTTP-запросов.

В состав серверной компоненты также входят два дополнительных модуля:

- подсистема предобработки результатов экспериментов, предназначенная для первичной обработки загружаемых файлов с экспериментальными данными, включая сжатие, объединение, разбиение и очистку данных;

- конвертер форматов источников данных, используемый для преобразования информации из исходных форматов, поддерживаемых датчиками объектов КФС, в табличный формат, пригодный для дальнейшей обработки и анализа.

С учетом особенностей хранения информации, получаемой в ходе экспериментов с объектами КФС, для ее обработки и хранения предлагается использовать двухуровневую структуру модели данных. На верхнем уровне данная модель определяет структуру данных базовых сущностей и включает следующую информацию:

- набор взаимосвязанных параметров, определяющих структуру и план миссии;
- начальные условия, задающие состояние КФС на момент начала исполнения миссии;
- состав акторов, участвующих в миссии;
- ограничения, которые необходимо учитывать при построении плана миссии;

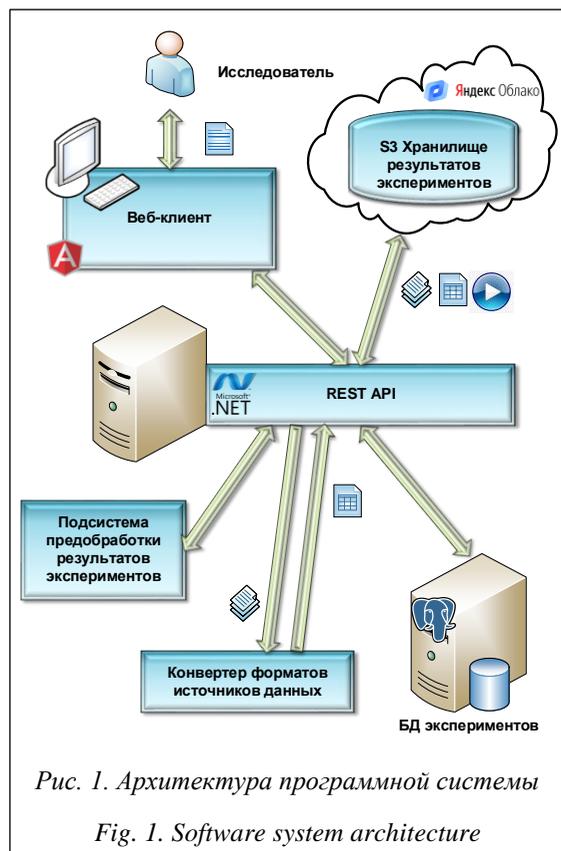


Рис. 1. Архитектура программной системы

Fig. 1. Software system architecture

– сведения о первичных источниках данных, используемых при реализации сценария управления и проведении эксперимента.

Этот уровень может быть представлен в формализованном виде. Для выполнения подобной формализации необходимо задавать структуру сценария управления с помощью схем данных, характерных для реляционных или объектно-реляционных СУБД, поскольку это позволяет однозначно выделять элементы структуры модели данных и устанавливать взаимосвязь между ними. Подобное требование также обусловлено необходимостью в дальнейшем осуществлять выборки сценариев управления по различным наборам условий с целью дальнейшего анализа и оптимизации данных сценариев.

Таким образом, при описании схемы данных на верхнем уровне должны быть предусмотрены следующие сущности: справочники типов акторов, акторов, грузов, форматов первичных источников данных, сценарии управле-

ния, сведения об использовании грузов в плане миссии, об использовании акторов в плане миссии; о заданиях, об этапах заданий, о первичных источниках данных.

Учитывая эти требования, для хранения данных на верхнем уровне была разработана БД в СУБД PostgreSQL. Физическая модель БД представлена на рисунке 2. В ее структуре предусмотрены таблицы-справочники с информацией о транспортируемых грузах (габариты, тип крепления, масса и т.п.) – Cargoes, об акторах (объектах КФС) – CPSObjects, о типах акторов – CPSObjectTypes, о форматах, используемых в источниках данных, – DataSourceFormats.

Также в структуре БД предусмотрены таблицы для хранения сведений об экспериментах и связанных с ними сценариях управления.

На нижнем уровне модель данных оперирует гетерогенной информацией, получаемой от первичных источников данных. Этими источниками являются датчики акторов и внеш-

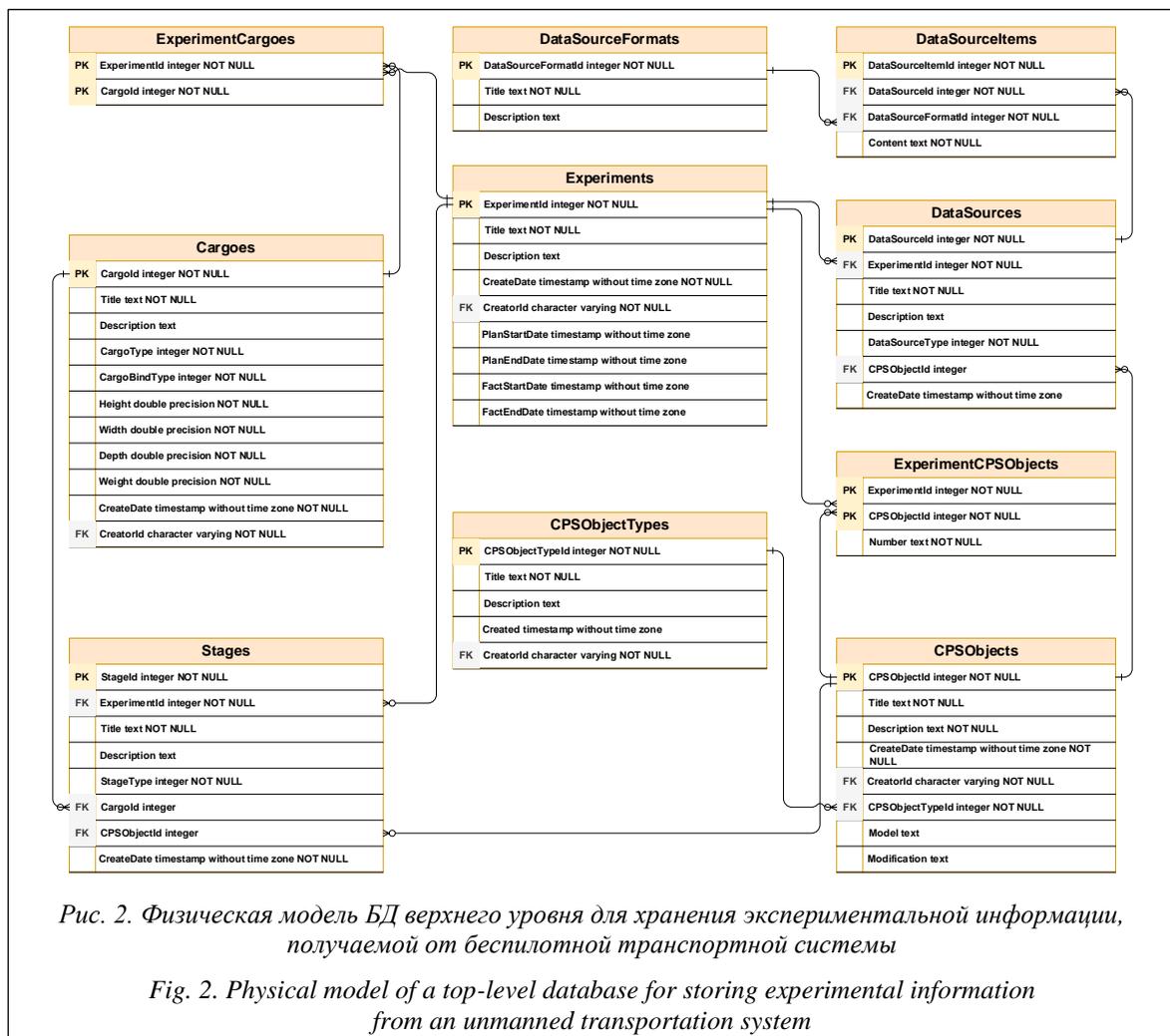


Рис. 2. Физическая модель БД верхнего уровня для хранения экспериментальной информации, получаемой от беспилотной транспортной системы

Fig. 2. Physical model of a top-level database for storing experimental information from an unmanned transportation system

них систем наблюдения. Получаемая от них информация может представляться в форме видеопотока, аудиопотока, потока числовых векторов и т.п. При этом обязательным требованием ко всем потокам данных, используемых в рамках сценария, является наличие привязки каждого фрейма данных ко времени. Объем подобных данных достаточно велик и не всегда структурирован, поэтому для дальнейшей работы с ним должны быть предусмотрены средства хранения его в первоначальном виде, а также методы очистки и предобработки данных.

С учетом того, что сами потоки данных зачастую могут быть представлены в виде файлов соответствующего формата, для их хранения целесообразно использовать облачные хранилища на основе технологии S3. Соответственно, для работы с файлами, содержащими показания датчиков, использовано S3-хранилище, построенное на базе Яндекс Облака. Взаимодействие с S3-хранилищем реализовано с помощью библиотеки AWS SDK для .NET и осуществляется путем преобразования запросов от веб-клиента к REST API в запросы к S3 API Яндекс Облака. Поддерживаются запросы на загрузку файла в бакет, удаление файла из бакета и обновление файла в бакете.

С целью обеспечения взаимосвязи между моделями верхнего и нижнего уровней на верхнем уровне должна быть предусмотрена дополнительная сущность – сведения об объекте S3-хранилища, используемого для хранения первичных данных определенного источника в указанном формате. Для этого в таблице DataSourceItems размещено поле Content строкового типа, в котором хранится в сериализованном виде (на базе формата JSON) описание связанного объекта бакета в S3-хранилище.

Описание и обсуждение экспериментов

В рамках тестирования и апробации разработанной системы был проведен ряд экспери-

ментов с участием следующих беспилотных мобильных средств: планер (с установленным GPS-датчиком), мультикоптер (квадрокоптер с установленными датчиками следующих видов: барометр, RGB-камера, инерциальная система наклона, GPS-датчик) и мобильный ровер (с установленными датчиками следующих видов: GPS-модуль, лазерный дальномер, модуль технического зрения).

Информация об одной из проведенных серий экспериментов, результаты которых были помещены в разработанное хранилище, представлена в таблице. В качестве примера можно привести эксперимент по сбору данных от сенсоров (координат объекта в пространстве и наблюдаемых с него видеопотоков) при перемещении мобильного ровера с установленной стереопарой. При продолжительности эксперимента 5 мин. были получены log-файл состояния размером 1,5 МБ и файл ROS-bag со стереопары в формате RGB и RGBD размером файла 5,57 Гб.

Полученные файлы с экспериментальной информацией были помещены в хранилище системы: метаданные сохранялись в таблицах БД PostgreSQL, а сами файлы помещались в S3-хранилище Яндекс Облака. Связь между хранилищами обеспечивалась за счет структурированного в формате JSON объекта в таблице базы PostgreSQL.

Отметим, что для некоторых форматов (например, log-файлов) имеется возможность конвертировать их в CSV-формат, более удобный для последующей автоматизированной обработки и анализа данных.

Как можно заметить, даже небольшой по продолжительности эксперимент (длительностью порядка 5 мин.) может порождать значительный объем информации (более 5 Гб), что подтверждает адекватность решения об использовании специализированного S3-хранилища исходной экспериментальной информации. Вместе с тем хранение файлов без метаданных о структуре и параметрах проведенных

Информация о результатах одной серии экспериментов с объектами КФС

Information on the results of one series of experiments with CPS objects

Тип объекта КФС	Количество экспериментов	Средняя продолжительность эксперимента, мин.	Количество источников информации	Суммарный объем полученных экспериментальных данных
Планер	3	19,3	1	154,7 Мб
Мультикоптер	2	10	1	17,78 Мб
Ровер	1	5	2	5,57 Гб

экспериментов и используемых форматах файлов существенно затруднило бы дальнейшее использование накопленной информации в автоматизированных системах анализа данных. Однако в рамках разработанной программной системы указанная проблема решается за счет структурированного хранения сведений об экспериментах и объектах КФС путем установления однозначного соответствия между записями в БД PostgreSQL и объектами (файлами) S3-хранилища. Таким образом, можно утверждать, что предложенная архитектура системы пригодна для решения задач хранения экспериментальной информации при исследовании моделей управления объектами КФС.

В целом следует отметить, что для хранения экспериментальной информации в разных предметных областях большинство исследователей предпочитают использовать либо локальные файловые хранилища, либо облачные (<https://www.stemcell.com/efficient-research/storing-data>). В этом случае структурирование экспериментов осуществляется за счет иерархического упорядочения файлов и каталогов. Данный подход имеет ряд существенных недостатков: низкая надежность хранения информации и сложность организации удаленного доступа при использовании собственных файловых хранилищ; ограниченный объем хранилищ (как локальных, так и облачных); сложность поиска информации по метаданным при удаленном доступе к хранилищу. С другой стороны, наблюдается рост интереса к использованию S3-хранилищ для хранения экспериментальных данных (<https://aws.amazon.com/ru/blogs/robotics/record-store-robot-data-rosbag/>). Однако хранение только самих данных эксперимента без метаданных затрудняет широкое применение подобных подходов, поскольку это существенно ограничивает возможности поиска и выборки экспериментов по заданным критериям. В то же время существуют и специализированные онлайн-хранилища, учитывающие метаданные. Например, подобные системы используются при проведении биологических [13] или обобщенных вычислительных экспериментов [5]. Однако хранилища подобного типа ориентированы на иную структуру метаданных и не подходят для применения в задачах хранения данных функционирования объектов КФС.

Предложенные в статье подходы и описанная система лишены указанных выше ограничений за счет

– использования структурированного хранения метаданных о проведенных эксперимен-

тах, что обеспечивает возможности поиска и выборки экспериментов по заданным критериям для их дальнейшего анализа;

– применения надежных и масштабируемых S3-хранилищ, гарантирующих сохранность результатов экспериментов, а также обеспечивающих возможность удаленного доступа и отсутствие ограничений по объему хранимой информации.

Заключение

Создание технологий и средств хранения информации, получаемой от объектов КФС в процессе их функционирования, является важным и перспективным направлением в рамках создания подходов и методов управления объектами КФС на основе анализа и интерпретации разнородных данных разного уровня структуризации. Предложенная система позволяет решать задачи хранения данных, получаемых в ходе экспериментов с объектами КФС, как в исходном, сыром, виде, так и в форматах, пригодных для автоматизированной обработки. Одним из перспективных путей использования этих данных является создание наборов данных для тестирования различных сценариев управления. Дополнительно за счет хранения в структурированном виде данных об экспериментах система позволяет накапливать соответствующий опыт, который в дальнейшем может быть использован при постановке и планировании новых экспериментов.

Принципы построения системы хранения данных были описаны на примере задач перемещения беспилотных мобильных средств в пространстве при решении задач транспортировки грузов. Вместе с тем с точки зрения архитектуры, применяемых подходов к хранению данных и форматов для их представления система в достаточной мере не зависит как от конфигурации КФС, так и от области ее применения. Благодаря этому имеется возможность масштабирования системы за счет как распределенного накопления и хранения разнородных данных, получаемых от объектов КФС и иных источников информации, так и расширения состава поддерживаемых типов объектов, форматов источников данных и методов их предварительной обработки, что дает возможность создать универсальную систему, позволяющую решать более широкий класс задач в области хранения и предварительной обработки разнородных данных. В свою очередь, данное обстоятельство будет способствовать

систематизации различных классов задач и сценариев управления объектами КФС, что может быть полезным как при разработке новых технологий и инструментов управления, так и при адаптации существующего математического аппарата и алгоритмического обеспечения при создании таких инструментов.

Еще одним актуальным направлением развития предложенного подхода является его расширение до уровня хранения не только самих данных функционирования объектов КФС,

но и моделей сценариев управления объектами. Таким образом, от системы хранения данных можно перейти к созданию на ее основе библиотек и репозиториев моделей по разным классам ситуационных задач управления объектами КФС. Наличие подобных репозиториев является необходимым условием развития ситуационного подхода к управлению объектами КФС, поскольку именно таким образом возможно накопление необходимой базы прецедентов.

Список литературы

1. Sánchez M., Morales J., Martínez J.L., Fernández-Lozano J.J., García-Cerezo A. Automatically annotated dataset of a ground mobile robot in natural environments via gazebo simulations. *Sensors*, 2022, vol. 22, no. 15, art. 5599. doi: 10.3390/s22155599.
2. Bosello M., Aguiari D., Keuter Y., Pallotta E. et al. Race against the machine: A fully-annotated, open-design dataset of autonomous and piloted high-speed flight. *IEEE Robotics and Automation Letters*, 2024, vol. 9, no. 4, pp. 3799–3806. doi: 10.1109/LRA.2024.3371288.
3. Bennett M.K., Younes N., Joyce K. Automating drone image processing to map coral reef substrates using google earth engine. *Drones*, 2020, vol. 4, no. 3, art. 50. doi: 10.3390/drones4030050.
4. Zhuo X., Koch T., Kurz F., Fraundorfer F., Reinartz P. Automatic UAV image geo-registration by matching UAV images to georeferenced image data. *Remote Sensing*, 2017, vol. 9, no. 4, art. 376. doi: 10.3390/rs9040376.
5. Подвесовский А.Г., Коростелёв Д.А., Лупачёв Е.А., Беляков Н.В. Построение хранилища обобщенных вычислительных экспериментов на основе онтологического подхода // Онтология проектирования. 2022. Т. 12. № 1. С. 41–56. doi: 10.18287/2223-9537-2022-12-1-41-56.
6. Wiemann T., Igelbrink F., Pütz S., Hertzberg J. A file structure and reference data set for high resolution hyperspectral 3D point clouds. *IFAC-PapersOnLine*, 2019, vol. 52, no. 8, pp. 403–408. doi: 10.1016/j.ifacol.2019.08.101.
7. Morast E. A deep learning based approach to object recognition from LiDAR data along Swedish railroads. TRITA, TRITA-ABE-MBT-22448. URL: <https://www.diva-portal.org/smash/get/diva2:1678019/FULLTEXT01.pdf> (дата обращения: 15.04.2024).
8. Wang S., Li G., Yao X., Zeng Y., Pang L., Zhang L. A distributed storage and access approach for massive remote sensing data in MongoDB. *ISPRS Int. J. Geo-Inf.*, 2019, vol. 8, no. 12, art. 533. doi: 10.3390/ijgi8120533.
9. Béjar-Martos J.A., Rueda-Ruiz A.J., Ogayar-Anguita C.J., Segura-Sánchez R.J., López-Ruiz A. Strategies for the storage of large LiDAR datasets – a performance comparison. *Remote Sensing*, 2022, vol. 14, no. 11, art. 2623. doi: 10.3390/rs14112623.
10. Allahham MHD S., Al-Sa'd M.F., Al-Ali A. et al. DroneRF dataset: A dataset of drones for RF-based detection, classification and identification. *Data in Brief*, 2019, vol. 26, art. 104313. doi: 10.1016/j.dib.2019.104313.
11. Захарова А.А., Кутахов В.П., Мещеряков Р.В., Подвесовский А.Г., Смолин А.Л. Моделирование задач транспортировки грузов в беспилотной авиационной транспортной системе // Авиакосмическое приборостроение. 2023. № 3. С. 3–15.
12. Zakharova A., Podvesovskii A. Model for optimization of heterogeneous cargo transportation using UAVs, taking into account the priority of delivery tasks. In: *SIST. Proc. ADOP*, 2023, vol. 362, pp. 257–268. doi: 10.1007/978-981-99-4165-0_24.
13. Morrell W.C., Birkel G.W., Forrer M. et al. The experiment data depot: A web-based software tool for biological experimental data storage, sharing, and visualization. *ACS Synth. Biol.*, 2017, vol. 6, no. 12, pp. 2248–2259. doi: 10.1021/acssynbio.7b00204.

Cyber-physical system objects: Organizing operation data storage

Dmitry A. Korostelyov^{1,2}, Aleksandr G. Podvesovskii^{1,2} ✉, Alena A. Zakharova¹

¹ V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, 117997, Russian Federation

² Bryansk State Technical University, Bryansk, 241035, Russian Federation

For citation

Korostelyov, D.A., Podvesovskii, A.G., Zakharova, A.A. (2024) 'Cyber-physical system objects: Organizing operation data storage', *Software & Systems*, 37(3), pp. 334–343 (in Russ.). doi: 10.15827/0236-235X.142.334-343

Article info

Received: 16.04.2024

After revision: 05.05.2024

Accepted: 14.05.2024

Abstract. The paper presents the results of research in the field of organization and software support when storing operation data from cyber-physical system objects obtained during experiments. The identified features of their storage are the following: special data formats, large amounts of data with different structuring degrees, supporting universal structured data formats, storing information about the structure and scenarios of experiments. The authors of the paper have analyzed technologies for building storage systems and special-purpose data formats that consider these features. They also proposed an approach to building a software system for storing and preprocessing experimental information about the functioning of cyber-physical system objects. The architecture and peculiarities of the system implementation are considered on the example of storing data of experiments with groups of jointly operating unmanned mobile vehicles when solving tasks of moving in space and transporting cargoes. The architecture assumes a server application and a web-client interacting with it by means of an API-interface based on REST principles. The server application also includes a subsystem for preprocessing experimental results and a data source format converter. The paper proposes to use a two-level structure of the data model for storing the experimental information. The upper level is intended for storing information about the structure and parameters of experiments and is implemented as a database in the PostgreSQL system. The lower level stores streaming information received from primary data sources. This level is implemented as a cloud storage built based on Yandex Cloud using S3 technology. The results of testing the developed system in real experiments confirmed the validity of the selected architectural solutions. Practical significance of the performed researches consists in developing approaches to the construction of scalable software systems for storing operating data of cyber-physical system objects. This contributes to organizing tasks and control scenarios of these objects when developing new control technologies and tools.

Keywords: cyberphysical system, cyberphysical system objects, experimental data, data formats, data model, DMS, REST API, S3 storage

Acknowledgements. This study was funded by the Russian Science Foundation, project no. 23-19-00342, <https://rscf.ru/en/project/23-19-00342/>

References

1. Sánchez, M., Morales, J., Martínez, J.L., Fernández-Lozano, J.J., García-Cerezo, A. (2022) 'Automatically annotated dataset of a ground mobile robot in natural environments via gazebo simulations', *Sensors*, 22(15), art. 5599. doi: 10.3390/s22155599.
2. Bosello, M., Aguiari, D., Keuter, Y., Pallotta, E. et al. (2024) 'Race against the machine: A fully-annotated, open-design dataset of autonomous and piloted high-speed flight', *IEEE Robotics and Automation Letters*, 9(4), pp. 3799–3806. doi: 10.1109/LRA.2024.3371288.
3. Bennett, M.K., Younes, N., Joyce, K. (2020) 'Automating drone image processing to map coral reef substrates using google earth engine', *Drones*, 4(3), art. 50. doi: 10.3390/drones4030050.
4. Zhuo, X., Koch, T., Kurz, F., Fraundorfer, F., Reinartz, P. (2017) 'Automatic UAV image geo-registration by matching UAV images to georeferenced image data', *Remote Sensing*, 9(4), art. 376. doi: 10.3390/rs9040376.
5. Podvesovskii, A.G., Korostelyov, D.A., Lupachev, E.A., Belyakov, N.V. (2022) 'Building a repository of generalized computational experiments based on the ontological approach', *Ontology of Designing*, 12(1), pp. 41–56 (in Russ.). doi: 10.18287/2223-9537-2022-12-1-41-56.
6. Wiemann, T., Igelbrink, F., Pütz, S., Hertzberg, J. (2019) 'A file structure and reference data set for high resolution hyperspectral 3D point clouds', *IFAC-PapersOnLine*, 52(8), pp. 403–408. doi: 10.1016/j.ifacol.2019.08.101.
7. Morast, E. 'A deep learning based approach to object recognition from LiDAR data along Swedish railroads', *TRITA, TRITA-ABE-MBT-22448*, available at: <https://www.diva-portal.org/smash/get/diva2:1678019/FULLTEXT01.pdf> (accessed April 15, 2024).
8. Wang, S., Li, G., Yao, X., Zeng, Y., Pang, L., Zhang, L. (2019) 'A distributed storage and access approach for massive remote sensing data in MongoDB', *ISPRS Int. J. Geo-Inf.*, 8(12), art. 533. doi: 10.3390/ijgi8120533.
9. Béjar-Martos, J.A., Rueda-Ruiz, A.J., Ogayar-Angueta, C.J., Segura-Sánchez, R.J., López-Ruiz, A. (2022) 'Strategies for the storage of large LiDAR datasets – a performance comparison', *Remote Sensing*, 14(11), art. 2623. doi: 10.3390/rs14112623.
10. Allahham, MHD S., Al-Sa'd, M.F., Al-Ali, A. et al. (2019) 'DroneRF dataset: A dataset of drones for RF-based detection, classification and identification', *Data in Brief*, 26, art. 104313. doi: 10.1016/j.dib.2019.104313.
11. Zakharova, A.A., Kutakhov, V.P., Meshcheryakov, R.V., Podvesovskii, A.G., Smolin A.L. (2023) 'Modeling cargo transportation tasks in an unmanned air transportation system', *Aerospace Instrument-Making*, (3), pp. 3–15 (in Russ.).
12. Zakharova, A., Podvesovskii, A. (2023) 'Model for optimization of heterogeneous cargo transportation using UAVs, taking into account the priority of delivery tasks', in *SIST. Proc. ADOP*, 362, pp. 257–268. doi: 10.1007/978-981-99-4165-0_24.
13. Morrell, W.C., Birkel, G.W., Forrer, M. et al. (2017) 'The experiment data depot: A web-based software tool for biological experimental data storage, sharing, and visualization', *ACS Synth. Biol.*, 6(12), pp. 2248–2259. doi: 10.1021/acssynbio.7b00204.

Авторы

Коростелев Дмитрий Александрович ^{1, 2},

к.т.н., доцент,
nigm85@mail.ru

Подвесовский Александр Георгиевич ^{1, 2},

к.т.н., доцент, профессор,
apodv@tu-bryansk.ru

Захарова Алена Александровна ¹,

д.т.н., доцент, главный научный сотрудник,
zaawmail@gmail.com

¹ Институт проблем управления
им. В.А. Трапезникова РАН,
г. Москва, 117997, Россия

² Брянский государственный технический
университет, г. Брянск, 241035, Россия

Authors

Dmitry A. Korostelyov ^{1, 2},

Cand. of Sci. (Engineering),
Associate Professor, nigm85@mail.ru

Aleksandr G. Podvesovskii ^{1, 2},

Cand. of Sci. (Engineering), Associate Professor,
Professor, apodv@tu-bryansk.ru

Alena A. Zakharova ¹,

Dr.Sci. (Engineering), Associate Professor,
Chief Researcher, zaawmail@gmail.com

¹ V.A. Trapeznikov Institute of Control Sciences
of Russian Academy of Sciences,
Moscow, 117997, Russian Federation

² Bryansk State Technical University,
Bryansk, 241035, Russian Federation

Применение MATLAB/GNU Octave для расчета погонных параметров многопроводных линий передачи методом моментов

А.Е. Максимов ¹✉, П.П. Снетков ¹, А.А. Иванов ¹, С.П. Куксенко ¹

¹ Томский государственный университет систем управления и радиоэлектроники (ТУСУР) – Научно-исследовательская лаборатория «Безопасность и электромагнитная совместимость радиоэлектронных средств», г. Томск, 634050, Россия

Ссылка для цитирования

Максимов А.Е., Снетков П.П., Иванов А.А., Куксенко С.П. Применение MATLAB/GNU Octave для расчета погонных параметров многопроводных линий передачи методом моментов // Программные продукты и системы. 2024. Т. 37. № 3. С. 344–353. doi: 10.15827/0236-235X.142.344-353

Информация о статье

Группа специальностей ВАК: 1.2.2

Поступила в редакцию: 07.02.2024

После доработки: 09.03.2024

Принята к публикации: 18.03.2024

Аннотация. В данной статье описан разработанный с использованием языка программирования MATLAB/GNU Octave прототип программного средства для расчета (экстракции) первичных погонных параметров (матриц коэффициентов электростатической и электромагнитной индукции) многопроводных линий передачи с произвольным поперечным сечением, любым числом диэлектрических слоев и проводников на каждом слое. В созданном прототипе используется численный метод моментов. Реализованы равномерная и два способа неравномерной сегментации границ поперечных сечений линий передачи (сегментация с учащением в углах и проекционная), а также три способа ее итерационного учащения (учащение всех сегментов, только тех сегментов, длина которых превышает установленное пороговое значение, и сегментов с наибольшей плотностью заряда). Реализованы четыре метода решения получаемых матричных уравнений: LU-разложение, BI-BiCGStab, BI-IDR(S) и BI-GMRES. После экстракции параметров выполняется их проверка на физичность. Разработанный прототип также способен осуществлять многовариантный анализ, то есть автоматическую итерационную экстракцию параметров линии передачи при изменении по определенному правилу входных данных (геометрических и электрофизических параметров). На основе результатов многовариантного анализа может быть выполнен расчет основных статистических характеристик, таких как математическое ожидание, дисперсия, среднеквадратическое отклонение и доверительный интервал. Выполнена верификация прототипа, показавшая, что экстрагированные параметры согласуются с аналогичными из других программных средств. Большая гибкость конфигурирования, модульность, различные методы сегментации и решения матричного уравнения позволяют использовать как прототип в целом, так и его отдельные модули для исследовательских и академических задач. Применяя вычисленные первичные параметры линии, можно легко рассчитать ее параметры рассеяния.

Ключевые слова: программное средство, расчет первичных погонных параметров, многопроводные линии передачи, квазистатическое приближение, метод моментов

Благодарности. Разработка прототипа программного средства выполнена за счет гранта РФФИ № 23-79-10165 (<https://rscf.ru/project/23-79-10165/>) в ТУСУР. Моделирование многопроводных линий передачи выполнено в рамках проекта FEWM-2024-0005 Минобрнауки России

Введение. Для учета требований по целостности сигнала в межсоединениях радиоэлектронных средств при их производстве необходимо предварительное моделирование межсоединений с помощью специализированных программных средств. Поскольку одним из основных типов межсоединений являются различные линии передачи, повышение скорости передачи данных и плотности монтажа приводит к необходимости учета электромагнитных связей между всеми близко расположенными проводниками, а также их моделирования как *многопроводных линий передачи* (МПЛП) [1, 2]. Моделирование МПЛП при помощи аналитических методов из-за сложности их структуры на практике часто неприменимо, поэтому обычно используется квазистатическое приближе-

ние, в ходе которого выполняется расчет (экстракция) параметров (матриц погонных коэффициентов электростатической \mathbf{L} и электромагнитной \mathbf{C} индукции, сопротивлений \mathbf{R} и проводимостей \mathbf{G}) [1] с применением численных методов. Один из наиболее популярных и универсальных из них – метод моментов [3, 4]. В двумерном случае моделируется поперечное сечение исследуемой линии, в котором границы раздела сред (проводник–диэлектрик и диэлектрик–диэлектрик) разбиваются на сегменты для последующего расчета взаимовлияний. При этом некорректный выбор сегментации может привести к существенной погрешности при расчете параметров МПЛП. Так, малое число сегментов приводит к низкой точности аппроксимации распределения зарядов

на поверхности проводников, например, к возникновению скачкообразных переходов, а излишне большое – к существенному увеличению вычислительных затрат из-за того, что размерность итогового матричного уравнения равна числу сегментов (N), а вычислительная сложность составляет $O(N^2)$.

Подходы к сегментации в разных программных средствах различаются. Так, в ряде средств на основе метода моментов, например, в программах PathWave ADS и FasterCap, доступна только автоматическая сегментация, в других, таких как CONMTL и LINPAR, по умолчанию используется автоматическая сегментация с возможностью ее ограниченной модернизации. Поскольку автоматическая сегментация часто может быть неоптимальной, для эффективного анализа программное средство должно поддерживать как различные виды адаптивной сегментации, так и ручную сегментацию вплоть до указания числа сегментов для каждой границы. Кроме того, в популярных программных средствах, как правило, отсутствует проверка корректности элементов матриц. Тогда вычисленные матрицы \mathbf{L} , \mathbf{C} , \mathbf{R} и \mathbf{G} выводятся как есть, без проверки их физичности, которая возлагается на пользователя (о чем, например, сообщается в документации программы PathWave ADS). При этом критерии физичности хорошо известны, а способы программной проверки достаточно тривиальны.

Технологические вариации геометрических параметров МПЛП, анализ температурных зависимостей и прочее требуют многократной экстракции погонных параметров, то есть многовариантного анализа. Для последующей же обработки множества экстрагированных параметров требуется статистический анализ.

Цель данной работы – представить прототип программного средства расчета (экстракции) погонных параметров МПЛП, отличающегося наличием модулей равномерной и неравномерной сегментации, проверки проектных норм экстрагированных параметров, а также многовариантного и статистического анализа.

Метод моментов

При анализе МПЛП с помощью метода моментов рассматривается уравнение Пуассона–Лапласа [5] в интегральном виде:

$$\varphi(\mathbf{r}) = \frac{1}{\varepsilon} \int \sigma(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') d\Gamma,$$

где σ – поверхностная плотность заряда; \mathbf{r} и \mathbf{r}' – точки наблюдения и источника соответственно; $d\Gamma$ – дифференциал по поверхности МПЛП;

ε – диэлектрическая проницаемость; $G(\mathbf{r}, \mathbf{r}')$ – функция Грина.

В данной постановке задачи считаются заданными граничные условия по приложенному напряжению φ , при этом требуется найти поверхностную плотность заряда σ . Для двумерного случая функция Грина имеет вид

$$G(\mathbf{r}, \mathbf{r}') = \frac{1}{2\pi} \ln |\mathbf{r} - \mathbf{r}'|. \quad (1)$$

При отсутствии в структуре анализируемой МПЛП бесконечной плоскости земли для корректного моделирования в ней должен присутствовать как минимум один опорный проводник. В этом случае для проводниковых границ (1) записывается как

$$\varphi(\mathbf{r}) = \frac{1}{2\pi\varepsilon_0} \int_{L_C} \sigma(\mathbf{r}') \ln |\mathbf{r} - \mathbf{r}'| dl', \quad \mathbf{r} \in L_C, \quad (2)$$

где L_C – контур проводниковых границ; dl' – элемент этого контура, а для диэлектрических границ получается

$$0 = \frac{\varepsilon_1 + \varepsilon_2}{\varepsilon_1 - \varepsilon_2} \frac{\sigma(\mathbf{r})}{2\varepsilon_0} + \frac{1}{2\pi\varepsilon_0} \int_{L_D} \sigma(\mathbf{r}') \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^2} \mathbf{n} dl', \quad \mathbf{r} \in L_D, \quad (3)$$

где L_D – контур диэлектрических границ; \mathbf{n} – вектор внешней нормали; ε_1 и ε_2 – диэлектрические проницаемости с одной (на которую направлен \mathbf{n}) и другой (от которой направлен \mathbf{n}) сторон границы.

С использованием аппарата базисных и тестовых функций выражения (2) и (3) сводятся к матричному уравнению вида

$$\mathbf{Z}\boldsymbol{\Sigma} = \boldsymbol{\Phi}, \quad (4)$$

где \mathbf{Z} – матрица размера $N \times N$, описывающая МПЛП через связи сегментов ее границ между собой; $\boldsymbol{\Phi}$ – матрица размера $N \times M$, содержащая известные потенциалы на проводниках; $\boldsymbol{\Sigma}$ – матрица размера $N \times M$, содержащая неизвестные общие (соответственно свободные и поляризационные) плотности заряда на этих сегментах; N – число базисных и тестовых функций, число сегментов расчетной сетки; M – число проводников МПЛП, не считая опорного.

Матричное уравнение (4) может быть решено прямым или итерационным методом, использующим последовательные приближения для получения более точного решения на каждой следующей итерации. Из решения (4) плотность свободных зарядов $\boldsymbol{\Sigma}_F$ вычисляется как

$$\boldsymbol{\Sigma}_F = \varepsilon_r \boldsymbol{\Sigma}, \quad (5)$$

где ε_r – относительная диэлектрическая проницаемость диэлектрика, к которому направлен вектор нормали от границы проводника.

Матрица \mathbf{C} , содержащая коэффициенты электростатической индукции, вычисляется из $\Sigma_F = \mathbf{C}\Phi$. (6)

Для получения значения c_{ij} нужно заземлить все проводники на опорный, кроме j -го, на который подается напряжение v_j , и определить заряд q_i на проводнике i . При этом на остальных проводниках будет минус q_i . В результате процесс вычисления элементов матрицы может быть представлен как M расчетов двухпроводной емкости, где остальные $(M - 1)$ проводники полагаются под одним потенциалом с опорным, а (4) служит для расчета зарядов на проводниках по заданным потенциалам.

При моделировании без учета потерь, кроме матрицы \mathbf{C} , также вычисляется матрица коэффициентов электромагнитной индукции \mathbf{L} как

$$\mathbf{L} = \mu_0 \epsilon_0 \mathbf{C}_0^{-1}, \quad (7)$$

где \mathbf{C}_0 – матрица, полученная по тому же алгоритму, что и \mathbf{C} , но при удалении из структуры диэлектрических границ (линия с воздушным заполнением), а μ_0 и ϵ_0 – магнитная и электрическая постоянные.

Программная реализация

Прототип программного средства расчета первичных погонных параметров МПЛП по (1)–(7) разработан с использованием синтаксиса интерпретируемого языка программирования MATLAB, совместимого с GNU Octave. В его составе шесть независимых модулей:

- построение геометрической модели поперечного сечения МПЛП (M1);
- сегментация границ поперечного сечения МПЛП (M2);
- формирование и решение матричного уравнения (M3);
- расчет (экстракция) параметров (матрицы \mathbf{C} и \mathbf{L}) (M4);
- проверка корректности элементов матриц (M5);
- многовариантный и статистический анализ (M6).

Каждый модуль получает данные и возвращает результат в заданных форматах. Схема взаимодействия модулей представлена на рисунке 1.

Модель поперечного сечения МПЛП задается в виде комбинации геометрических примитивов: прямоугольник, многоугольник и эллипс. Для их построения на вход модуля M1 подаются массивы **rect**, **poly** и **ellip** соответственно. Каждая замкнутая фигура образует область (электрически- и магнитнооднородную площадь).

Массив **rect** имеет размерность $M \times 4$, где M – число прямоугольников, каждый из которых определяется четырьмя координатами: минимальная и максимальная абсциссы, минимальная и максимальная ординаты. Трехмерный массив **poly** имеет размерность $2 \times L \times M$, где M – число многоугольников, а L – число линейных сегментов (граней) M -го многоугольника. Для каждой грани первой строкой массива указываются все абсциссы точек, а второй – ординаты. Массив **ellip** имеет размерность $M \times 5$, где M – число эллипсов. Для каждого из них последовательно указываются координаты центра (абсцисса и ордината), длины большой и малой полуосей и вращательный угол (в радианах). Кроме того, на входе модуля M1 для каждой области задается значение относительной диэлектрической проницаемости ϵ_r внутри нее. После окончания работы модуля M1 реализующая его функция возвращает структуру данных *str* (размерностью $10 \times N$, где N – общее число границ), содержащую следующие поля: *s* – номер начальной вершины границы (поперечного сечения); *e* – номер конечной вершины границы; *d* – номер области, к которой относится граница; *sx* – абсцисса центра границы; *sy* – ордината центра границы; *h* – длина границы; *nx* – горизонтальная компонента единичного вектора нормали к границе; *ny* – вертикальная компонента единичного вектора нормали к границе; *ein* – значение ϵ_r внутри области (против вектора нормали); *eout* – значение ϵ_r снаружи области (по вектору нормали).

Структура данных *str* передается на вход модуля M2, где реализуется сегментация границ поперечного сечения МПЛП. Так, доступна равномерная сегментация, при которой пользователь может выбрать только максимальный размер сегмента, и ряд неравномерных, у которых доступны дополнительные параметры. При использовании равномерной сегментации каждая граница разбивается на сегменты равной длины. Их число определяется путем округления в большую сторону частного от деления длины границы на заданный пользователем максимальный размер сегмента. Реализованы два вида неравномерной сегментации: с учащением в углах [6] и проекционная. Также в модуле M2 реализована возможность автоматического итерационного учащения сегментации. Доступны три варианта учащения сегментов на каждой итерации: всех (разбиение каждого сегмента пополам); длина которых превышает установленное пороговое значение [7]; с наибольшей плотно-

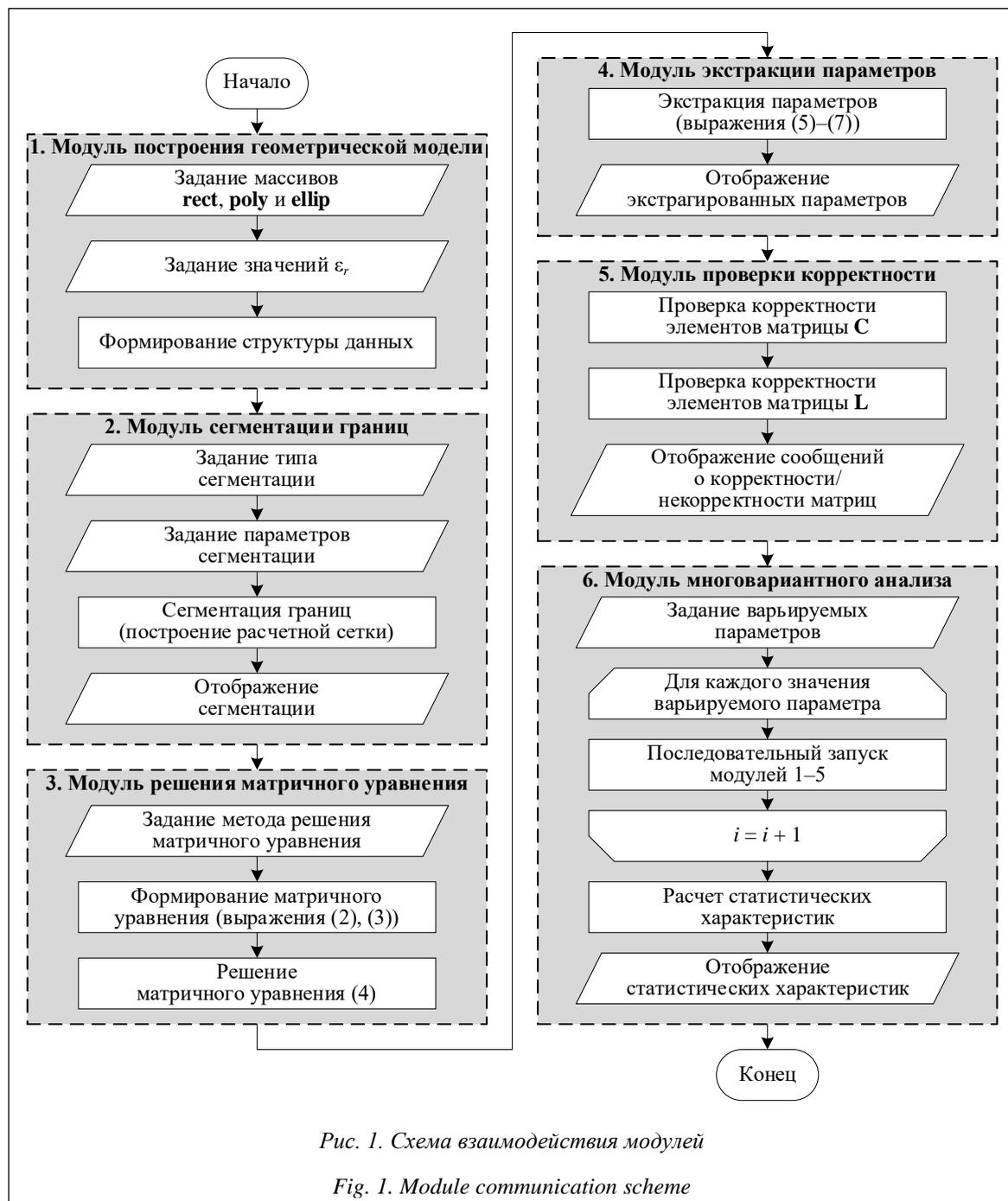


Рис. 1. Схема взаимодействия модулей

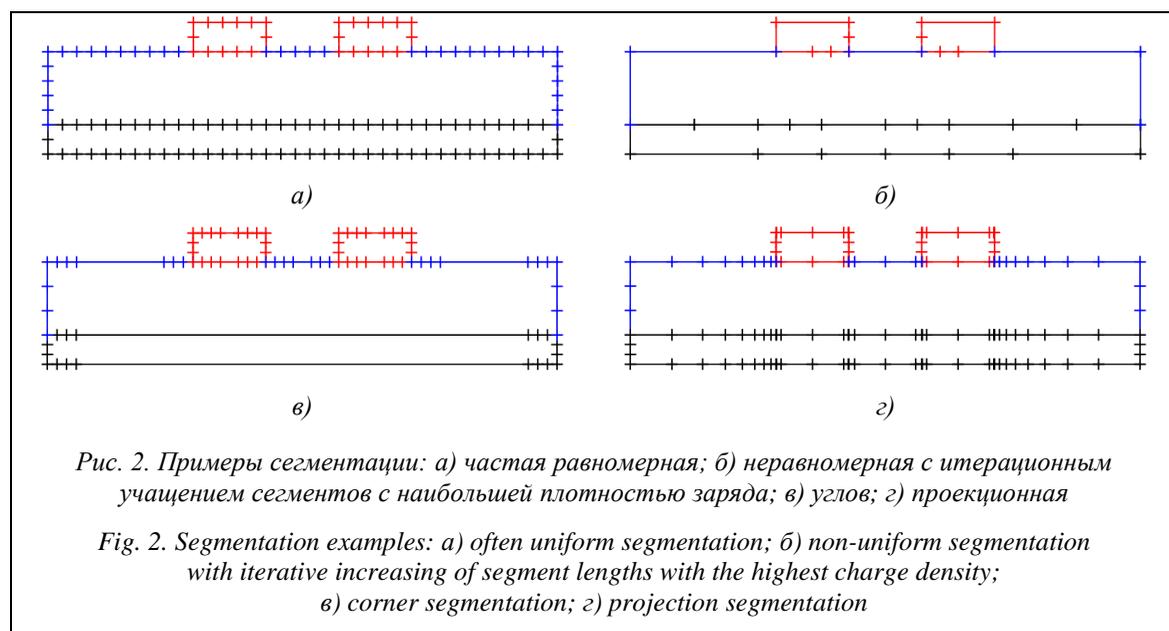
Fig. 1. Module communication scheme

стью заряда [5], используя информацию из модуля M4.

Различные методы сегментации и ее учащения могут быть скомбинированы. На выходе модуля M2 используется та же структура данных, что и у M1, но в ней описаны не границы, а их сегменты. Сегментация границ поперечного сечения может быть отображена графически. Около каждого сегмента по желанию пользователя отображаются номер сегмента, номер области, к которой относится сегмент,

относительные диэлектрические проницаемости сред слева и справа от сегмента, направление вектора нормали. Примеры итоговой сегментации приведены на рисунке 2.

В модуле M3 формируется и решается матричное уравнение (4). При решении уравнений небольшого размера и с плотной матрицей, как правило, используются прямые методы, такие как метод Гаусса или его модифицированная версия, основанная на LU-разложении [8]. Однако для больших матричных уравнений пря-



мые методы вычислительно затратны, поэтому применяются итерационные. Для прямых методов затраты времени на решение матричного уравнения соответствуют $O(N^3)$, а машинной памяти – $O(N^2)$. Для итерационных методов эти затраты, как правило, ниже и зависят от требуемой точности. Еще одной особенностью итерационных методов является то, что получаемая погрешность решения из-за конечного числа разрядов намного меньше, чем в методе Гаусса, так как она не накапливается, а определяется только последней итерацией и не зависит от их числа. Поэтому решение с заданной точностью при росте числа обусловленности матрицы достигается просто увеличением числа итераций.

Для больших матричных уравнений с несимметричными матрицами часто используют методы крыловского типа [9]. В основном они сводятся к генерации подходящего базиса векторного пространства, называемого подпространством Крылова, и к выбору фактической итерации в пределах этого пространства. Достаточно популярными методами крыловского типа являются стабилизированный метод бисопряженных градиентов (BiCGStab), метод обобщенных минимальных невязок (GMRES), метод индуцированного уменьшения размерности (IDR) и его модификация IDR(S) [9, 10].

Для решения последовательности матричных уравнений, что актуально при использовании многовариантного анализа (модуль М6), целесообразно использовать блочные версии итерационных методов, иначе возникает необ-

ходимость последовательных отдельных вычислений с каждой правой частью. Это дополнительно увеличивает вычислительные затраты [9]. Поэтому в модуле М3 реализованы метод LU-разложения (встроенная в MATLAB/GNU Octave версия) и блочные версии методов BiCGStab, IDR(S) и GMRES [11, 12] (авторские реализации).

Модуль М4 получает на вход матрицу Σ , вычисленную в модуле М3, реализует вычисления по (5)–(7) и возвращает (передает модулю М5) рассчитанные матрицы \mathbf{C} и \mathbf{L} .

Модуль М5 проверяет корректность элементов матриц в соответствии с блок-схемами, приведенными на рисунках 3 и 4, и выводит соответствующие сообщения.

Модуль М6 позволяет выполнять многовариантный анализ параметров МПЛП. При варьировании любого параметра МПЛП модуль автоматически итерационно запускает остальные модули, позволяя получать на выходе набор матриц \mathbf{C} и \mathbf{L} . Далее над этим набором может быть выполнен статистический анализ с расчетом основных статистических характеристик.

Верификация программного средства расчета первичных погонных параметров

Верификация разработанного прототипа программного средства выполнена на примере двухпроводной микрополосковой линии передачи (далее – ЛПП) со следующими параметрами [13]: ширина проводников $w = 2,35$ мм;

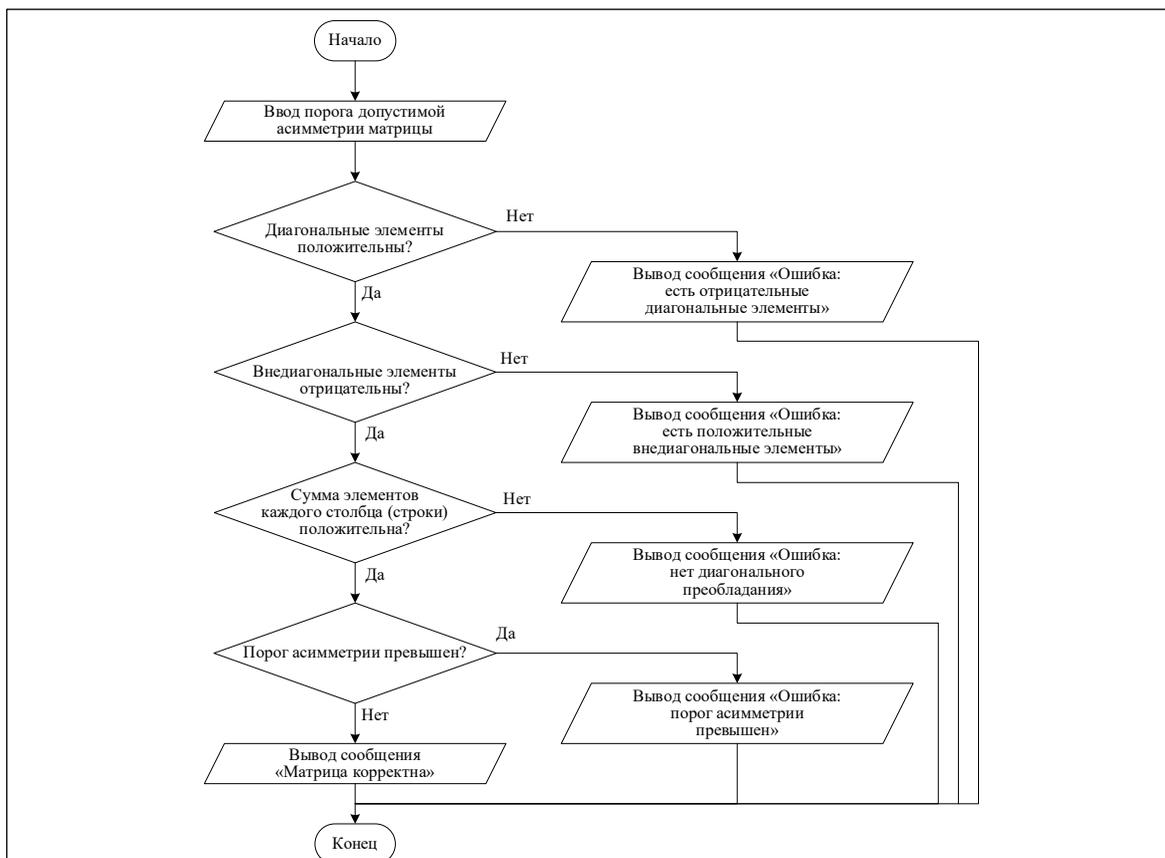


Рис. 3. Блок-схема алгоритма проверки корректности элементов матрицы C

Fig. 3. Block diagram of the algorithm for checking the correctness of matrix C entries

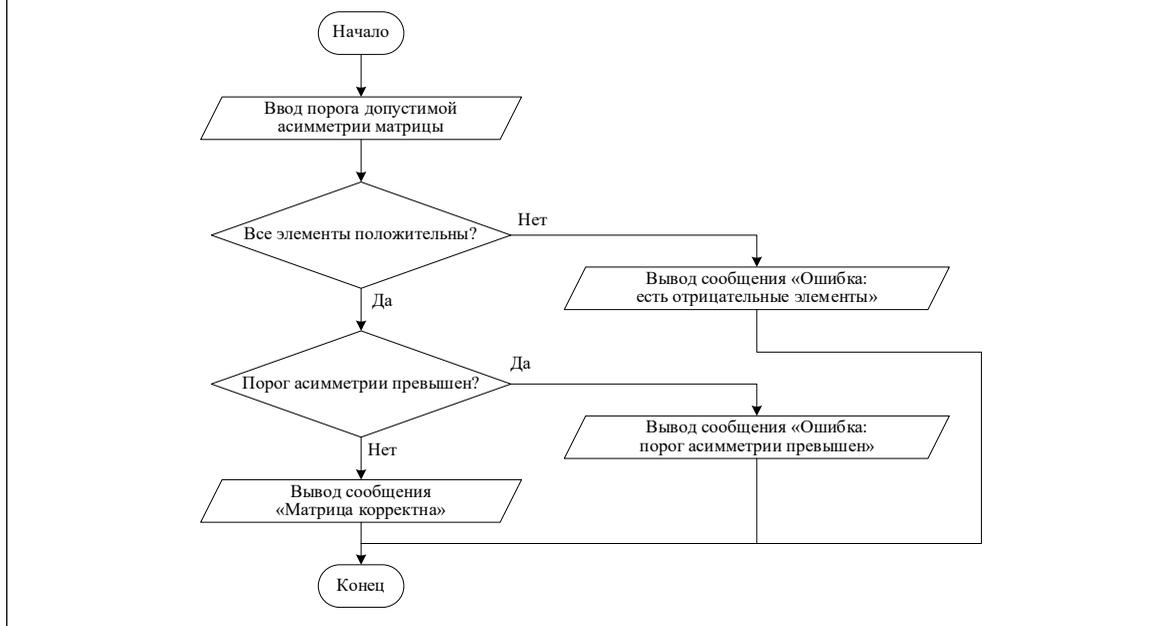


Рис. 4. Блок-схема алгоритма проверки корректности элементов матрицы L

Fig. 4. Block diagram of the algorithm for checking the correctness of matrix L entries

расстояние между проводниками $s = 0,65$ мм; толщина проводников $t = 35$ мкм; толщина ди-

электрического слоя $h = 1,5$ мм; расстояние от проводников до границ структуры $d = 9,825$ мм;

$\epsilon_r = 5,18$. Ее поперечное сечение представлено на рисунке 5. Для вычисления матриц **C** и **L** использованы программные средства PathWave ADS, CONMTL, Ansys Q2D Extractor, а также численно-аналитический метод конформных отображений (МКО) с помощью интеграла Кристоффеля–Шварца [13]. Полученные матрицы отображены в таблице 1. Также в ней указаны параметры сегментации.

Дополнительно выполнена верификация на примере более сложной МПЛП (далее – ЛП2), поперечное сечение которой представлено на рисунке 6. Ее параметры [14]: $w = 0,2$ мм, $t = 0,1$ мм, $h = 0,1$ мм, $h_1 = 0,5$ мм, $h_2 = 0,4$ мм, $h_3 = 9,1$ мм, $d = 1,5$ мм, $\epsilon_{r1} = 6,8$, $\epsilon_{r2} = 4,5$, $\epsilon_{r3} = 1$. В таблице 2 представлены экстрагированные матрицы.

Как видно из таблиц 1 и 2, параметры, экстрагированные в разработанном прототипе

Таблица 1

Матрицы **C** и **L** для ЛП1, полученные в различных программных средствах

Table 1

Matrices **C** and **L** for TL1 obtained in different programs

Программное средство	Сегментация	C, пФ/м			L, мкГн/м				
МКО [13]	Нет данных	131,30	-23,60	-23,60	131,30	0,3245	0,0916	0,0916	0,3245
PathWave ADS	Нет данных	131,85	-23,28	-23,28	131,85	0,3255	0,0915	0,0915	0,3255
CONMTL	Равномерная, $N = 2\ 320$	130,41	-22,83	-22,83	130,41	0,3262	0,0931	0,0931	0,3262
	Неравномерная, $N = 2\ 320$	130,96	-23,06	-23,06	130,96	0,3260	0,0931	0,0931	0,3260
Ansys Q2D Extractor (метод конечных элементов)	Неравномерная, число треугольных элементов $N_t = 2\ 248$	132,23	-23,56	-23,56	132,17	0,3408	0,0904	0,0904	0,3408
	Неравномерная с итерационным учащением, $N_t = 1\ 1986$	132,02	-23,48	-23,48	132,02	0,3408	0,0904	0,0904	0,3408
Разработанное программное средство	Равномерная, $N = 2\ 369$	132,65	-22,37	-22,37	132,65	0,3254	0,0929	0,0929	0,3254
	Неравномерная с итерационным учащением, $N = 1\ 262$	132,10	-22,19	-22,19	132,10	0,3252	0,0928	0,0928	0,3252

Таблица 2

Матрицы **C** и **L** для ЛП2, полученные в различных программных средствах

Table 2

Matrices **C** and **L** for TL2 obtained in different programs

Программное средство	Сегментация	C, пФ/м			L, мкГн/м												
Ansys Q2D Extractor (метод конечных элементов)	$N_t = 6\ 715$	126,61	-13,39	-13,39	33,82	-72,26	-7,67	376,50	0,5041	0,2178	0,1205	0,2178	0,6487	0,0835	0,1205	0,0835	0,2231
[14]	Нет данных	125,86	-13,13	-13,13	34,10	-69,56	-7,18	357,62	0,4919	0,1989	0,1775	0,1989	0,6128	0,0768	0,1775	0,0768	0,2299
Разработанное программное средство	$N = 576$	125,47	-10,00	-9,97	34,17	-69,37	-8,50	356,04	0,4669	0,1413	0,1101	0,1413	0,5522	0,0650	0,1101	0,0650	0,2267

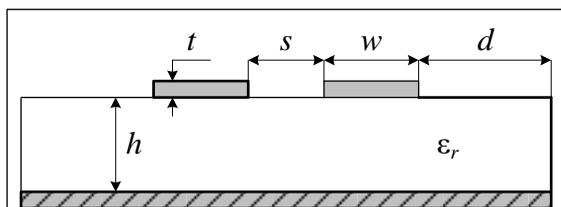


Рис. 5. Поперечное сечение двухпроводной микрополосковой линии передачи (ЛП1)

Fig. 5. Cross section of a two-conductor microstrip transmission line (TL1)

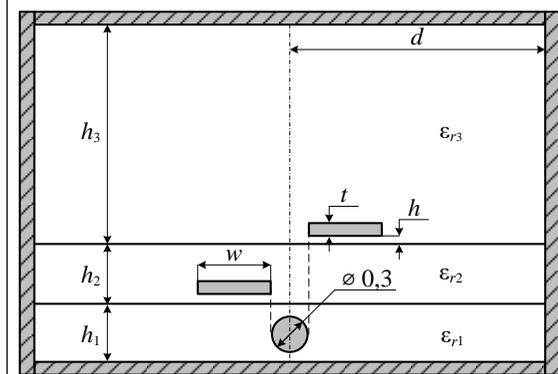


Рис. 6. Поперечное сечение МПЛП (ЛП2)

Fig. 6. Cross section of a multiconductor transmission line (TL2)

программного средства, хорошо согласуются с аналогичными параметрами в других программных средствах. Так, относительное различие по элементу c_{11} для ЛП1 не превышает 2 %, для ЛП2 – 1 %.

Выполним многовариантный анализ для ЛП1. Для этого последовательно варьируются геометрические параметры w и t в диапазоне от –7 % до 7 % с шагом 1 % (всего по 15 вариаций). Общее число вычислений (число экстрагированных матриц **C** и **L**) составило 225. По экстрагированным матрицам выполнен статистический анализ (вычислены характеристики: математическое ожидание E , дисперсия D , среднееквадратическое отклонение σ и довери-

тельный интервал δ), представленный в таблице 3. Как видно, все среднееквадратические отклонения находятся в диапазоне 1–3 % от среднего значения (E), при этом наименьшее значение σ наблюдается у c_{12} (1,2 %).

Таблица 3

Статистические характеристики для элементов матриц **C и **L** ЛП1**

Table 3

Statistical characteristics for the entries of TL1 matrices **C and **L****

Элемент матрицы	E	D	σ	δ
c_{11}	130,60, пФ/м	11,492, (пФ/м) ²	3,390, пФ/м	0,445, пФ/м
c_{12}	–22,84, пФ/м	0,069, (пФ/м) ²	0,264, пФ/м	0,035, пФ/м
l_{11}	0,3250, мкГн/м	56,160, (нГн/м) ²	7,494, нГн/м	0,979, нГн/м
l_{12}	0,0911, мкГн/м	7,885, (нГн/м) ²	2,808, нГн/м	0,369, нГн/м

Заключение

Таким образом, с использованием синтаксиса интерпретируемого языка программирования MATLAB, совместимого с GNU Octave, разработан прототип программного средства расчета первичных погонных параметров МПЛП (матриц **L** и **C**) и проверки корректности их элементов. Разработанный прототип состоит из шести модулей, первые пять из которых являются полностью самостоятельными, а шестой выполняет их последовательный запуск. Разработанный прототип верифицирован с использованием ряда существующих программных средств. В дальнейшем на его основе планируется создание программного комплекса экстракции параметров на низкоуровневом языке программирования. Кроме того, модульность прототипа позволяет использовать его отдельные модули, в частности, для академических задач.

Список литературы

1. Paul C.R. Analysis of multiconductor transmission lines. Hoboken, New Jersey, John Wiley & Sons Publ., 2007, 803 p.
2. Belousov A.O., Chernikova E.B., Samoylichenko M.A. et al. From symmetry to asymmetry: The use of additional pulses to improve protection against ultrashort pulses based on modal filtration. Symmetry, 2020, vol. 12, no. 7, art. 1117. doi: 10.3390/sym12071117.
3. Makarov S.N., Noetscher G.M., Nazarian A. Low-frequency electromagnetic modeling for electrical and biological systems using MATLAB. Hoboken, New Jersey, John Wiley & Sons Publ., 2015, 616 p.
4. Gibson W.C. The method of moments in electromagnetics. Boca Raton, Chapman & Hall/CRC Publ., 2021, 272 p.
5. Maksimov A.E., Kuksenko S.P. Accurate capacitance matrices for multiconductor transmission lines. IEEE Transactions on Electromagnetic Compatibility, 2022, vol. 64, no. 5, pp. 1514–1521. doi: 10.1109/TEMC.2022.3175717.
6. Maksimov A., Kuksenko S. Adaptive segmentation of multiconductor transmission lines in quasi-static analysis by the method of moments. JPCS, 2021, vol. 1862, no. 1, art. 012020. doi: 10.1088/1742-6596/1862/1/012020.

7. Максимов А.Е., Куksenko С.П. Адаптивный итерационный выбор оптимальной сегментации при решении электростатических задач методом моментов // СИБРЕСУРС-26-2020: матер. Междунар. науч.-практич. конф. 2020. С. 112–116.
8. Gu Z., Wu H., Zhao X., Zhang Y. New LU decomposition accelerating parallel method of moments and its application. Proc. IEEE MTT-S IWS, 2018, pp. 1–3. doi: 10.1109/IEEE-IWS.2018.8400942.
9. Meurant G., Tebbens J.D. Krylov methods for nonsymmetric linear systems: From theory to computations. Berlin, Heidelberg, Springer Publ., 2020, 691 p. doi: 10.1007/978-3-030-55251-0.
10. Kehl R., Nabben R., Szlyd D.B. Adaptive multilevel Krylov methods. ETNA, 2017, vol. 51, pp. 512–528. doi: 10.1553/etna_vol51s512.
11. Daas H., Grigori L., Hénon P., Ricoux P. Enlarged GMRES for solving linear systems with one or multiple right-hand sides. IMA J. of Numerical Analysis, 2019, vol. 39, no. 4, pp. 1924–1956. doi: 10.1093/imanum/dry054.
12. Sun D.L., Huang T.Z., Jing Y.F., Carpentieri B. A block GMRES method with deflated restarting for solving linear systems with multiple shifts and multiple right-hand sides. Numerical Linear Algebra with Applications, 2018, vol. 25, no. 5, art. e2148. doi: 10.1002/nla.2148.
13. Стручков С.М., Сычев А.Н. Методика измерения погонных параметров симметричных связанных линий // Электронные средства и системы управления: матер. Междунар. науч.-практич. конф. 2015. № 1-1. С. 159–163.
14. Musa S.M., Sadiku M.N.O. Application of the finite element method in calculating the capacitance and inductance of multiconductor transmission lines. Proc. IEEE SoutheastCon, 2008, pp. 300–304. doi: 10.1109/SECON.2008.4494308.

Software & Systems

doi: 10.15827/0236-235X.142.344-353

2024, 37(3), pp. 344–353

MATLAB/GNU Octave for calculating per-unit-length parameters of multiconductor transmission lines using the method of moments

Aleksandr E. Maksimov ¹✉, Pavel P. Snetkov ¹,
Anton A. Ivanov ¹, Sergey P. Kuksenko ¹

¹ Tomsk State University of Control Systems and Radioelectronics (TUSUR) –
Research Laboratory “Safety and Electromagnetic Compatibility of Radioelectronic Equipment”,
Tomsk, 634050, Russian Federation

For citation

Maksimov, A.E., Snetkov, P.P., Ivanov, A.A., Kuksenko, S.P. (2024) ‘MATLAB/GNU Octave for calculating per-unit-length parameters of multiconductor transmission lines using the method of moments’, *Software & Systems*, 37(3), pp. 344–353 (in Russ.). doi: 10.15827/0236-235X.142.344-353

Article info

Received: 07.02.2024

After revision: 09.03.2024

Accepted: 18.03.2024

Abstract. The paper describes a prototype of software for calculating (extraction) primary per-unit-length parameters (electrostatic and electromagnetic induction coefficient matrices) developed using MATLAB/GNU Octave programming language. The program extracts parameters of multiconductor transmission lines with arbitrary cross-section, any number of dielectric layers and conductors on each layer. It uses the numerical method of moments. There are implemented uniform and 2 methods of non-uniform segmentation of transmission line cross-section boundaries (corner and projection segmentation). There are also 3 methods of iterative segmentation (segmentation of all segments, only those segments whose length exceeds the set threshold value and segments with the highest charge density). Four methods of solving the matrix equations are implemented: LU-decomposition, BI-BiCGStab, BI-IDR(S) and BI-GMRES. Parameter extracting is followed by their physicality check. The developed prototype is also able to perform multivariate analysis, i.e. automatic iterative extraction of transmission line parameters when the input data (geometric and electrophysical parameters) change according to a certain rule. The results of multivariate analysis can become a base for statistical analysis (calculation of basic statistical characteristics such as mathematical expectation, variance, standard deviation and confidence interval). Verification of the prototype showed that the extracted parameters agree with similar ones from other software. Great flexibility of configuration, modularity, different methods of segmentation and matrix equation solution allow using both the prototype as a whole and its separate modules for research and academic tasks. The calculated primary per-unit-length parameters of the line help to calculate its scattering parameters easily.

Keywords: program, primary per-unit-length parameter calculation, multiconductor transmission lines, quasi-static approximation, method of moments

Acknowledgements. The development of the software prototype was funded by the Russian Science Foundation 23-79-10165 (<https://rscf.ru/en/project/23-79-10165/>) at TUSUR. The modeling of multiconductor transmission lines was supported by the Ministry of Science and Higher Education of the Russian Federation, project FEWM-2024-0005

References

1. Paul, C.R. (2007) *Analysis of Multiconductor Transmission Lines*. Hoboken, New Jersey: John Wiley & Sons Publ., 803 p.
2. Belousov, A.O., Chernikova, E.B., Samoylichenko, M.A. et al. (2020) ‘From symmetry to asymmetry: The use of additional pulses to improve protection against ultrashort pulses based on modal filtration’, *Symmetry*, 12(7), art. 1117. doi: 10.3390/sym12071117.
3. Makarov, S.N., Noetscher, G.M., Nazarian, A. (2015) *Low-frequency Electromagnetic Modeling for Electrical and Biological Systems Using MATLAB*. Hoboken, New Jersey: John Wiley & Sons Publ., 616 p.
4. Gibson, W.C. (2021) *The Method of Moments in Electromagnetics*. Boca Raton, Chapman & Hall/CRC Publ., 272 p.
5. Maksimov, A.E., Kuksenko, S.P. (2022) ‘Accurate capacitance matrices for multiconductor transmission lines’, *IEEE Transactions on Electromagnetic Compatibility*, 64(5), pp. 1514–1521. doi: 10.1109/TEMC.2022.3175717.
6. Maksimov, A., Kuksenko, S. (2021) ‘Adaptive segmentation of multiconductor transmission lines in quasi-static analysis by the method of moments’, *JPCS*, 1862(1), art. 012020. doi: 10.1088/1742-6596/1862/1/012020.
7. Maksimov, A.E., Kuksenko, S.P. (2020) ‘Adaptive iterative selection of optimal segmentation in solving electrostatic problems by the method of moments’, *Proc. Int. Sci. and Pract. Conf. Natural and Intellectual Resources of Siberia*, pp. 112–116 (in Russ.).
8. Gu, Z., Wu, H., Zhao, X., Zhang, Y. (2018) ‘New LU decomposition accelerating parallel method of moments and its application’, *Proc. IEEE MTT-S IWS*, pp. 1–3. doi: 10.1109/IEEE-IWS.2018.8400942.
9. Meurant, G., Tebbens, J.D. (2020) *Krylov Methods for Nonsymmetric Linear Systems: From Theory to Computations*. Berlin, Heidelberg: Springer Publ., 691 p. doi: 10.1007/978-3-030-55251-0.
10. Kehl, R., Nabben, R., Szyld, D.B. (2017) ‘Adaptive multilevel Krylov methods’, *ETNA*, 51, pp. 512–528. doi: 10.1553/etna_vol51s512.
11. Daas, H., Grigori, L., Hénon, P., Ricoux, P. (2019) ‘Enlarged GMRES for solving linear systems with one or multiple right-hand sides’, *IMA J. of Numerical Analysis*, 39(4), pp. 1924–1956. doi: 10.1093/imanum/dry054.
12. Sun, D.L., Huang, T.Z., Jing, Y.F., Carpentieri, B. (2018) ‘A block GMRES method with deflated restarting for solving linear systems with multiple shifts and multiple right-hand sides’, *Numerical Linear Algebra with Applications*, 25(5), art. e2148. doi: 10.1002/nla.2148.
13. Struchkov, S.M., Sychev, A.N. (2015) ‘Methodology for measuring linear parameters of symmetrical coupled lines’, *Proc. Int. Sci. and Pract. Conf. Electronic Means and Control Systems*, (1-1), pp. 159–163 (in Russ.).
14. Musa, S.M., Sadiku, M.N.O. (2008) ‘Application of the finite element method in calculating the capacitance and inductance of multiconductor transmission lines’, *Proc. IEEE SoutheastCon*, pp. 300–304. doi: 10.1109/SECON.2008.4494308.

Авторы

Максимов Александр Евгеньевич¹,
инженер, mae@tusur.ru
Снетков Павел Павлович¹,
техник, spp@tu.tusur.ru
Иванов Антон Андреевич¹, к.т.н.,
старший научный сотрудник, anton.ivvv@gmail.com
Куксенко Сергей Петрович¹,
д.т.н., доцент, зав. лабораторией,
sergei.p.kuksenko@tusur.ru

¹ Томский государственный университет систем
управления и радиоэлектроники (ТУСУР) –
Научно-исследовательская лаборатория
«Безопасность и электромагнитная совместимость
радиоэлектронных средств»,
г. Томск, 634050, Россия

Authors

Aleksandr E. Maksimov¹,
Engineer, mae@tusur.ru
Pavel P. Snetkov¹,
Technical Fellow, spp@tu.tusur.ru
Anton A. Ivanov¹, Cand. of Sci. (Engineering),
Senior Researcher, anton.ivvv@gmail.com
Sergey P. Kuksenko¹, Dr.Sci. (Engineering),
Associate Professor, Head of Laboratory,
sergei.p.kuksenko@tusur.ru

¹ Tomsk State University of Control Systems
and Radioelectronics (TUSUR) –
Research Laboratory
“Safety and Electromagnetic
Compatibility of Radioelectronic Equipment”,
Tomsk, 634050, Russian Federation

Модификация алгоритма поиска кратчайших путей ГН-графа для анализа сложных технических систем

Е.Р. Зяблова ¹✉

¹ Южный федеральный университет, г. Таганрог, 347900, Россия

Ссылка для цитирования

Зяблова Е.Р. Модификация алгоритма поиска кратчайших путей ГН-графа для анализа сложных технических систем // Программные продукты и системы. 2024. Т. 37. № 3. С. 354–363. doi: 10.15827/0236-235X.142.354-363

Информация о статье

Группа специальностей ВАК: 2.3.8

Поступила в редакцию: 25.02.2024

После доработки: 29.03.2024

Принята к публикации: 09.04.2024

Аннотация. В работе предложен один из подходов к моделированию сложных технических систем на примере решения задачи формирования зон влияния объектов системы охраны протяженного периметра. Подход основан на использовании так называемого ГН-графа (нечеткого графа с разнотипными вершинами и множественными и разнотипными связями) и определенных алгоритмических средств. В рамках данного исследования выполнен синтез модифицированного алгоритма Форда–Беллмана поиска кратчайших путей ГН-графа. Модифицированный алгоритм позволяет анализировать разнотипные информационные потоки в сложных технических системах. От других его отличает возможность поиска расстояний для полных или усеченных множеств вершин и/или связей графа, например, для связей заданного типа. Предложено списковое представление множественных и разнотипных связей ГН-графа. Сформулированы критерии поиска кратчайших путей в ГН-графе: критерий 1 – выбор типа (типов) вершин, участвующих в работе алгоритма, критерий 2 – выбор типов (векторов) связей. При этом имеется возможность оставить полные множества вершин и связей модели в списковом представлении графа. Определено значение вычислительной сложности модифицированного алгоритма, которое не превышает значение сложности исходного алгоритма Форда–Беллмана. Время работы предложенного алгоритма снижается за счет использования в ГН-графе множественных связей в виде вектора, позволяющих объединять ряд разнотипных связей. Результатом работы модифицированного алгоритма является матрица расстояний с учетом полных или усеченных множеств вершин и/или связей для последующего вычисления метрических характеристик графовой модели средствами, определенными ранними исследованиями. Предложенные подходы к моделированию сложных технических систем для решения поставленной задачи подробно рассмотрены на примере определения зоны влияния технических устройств (квадрокоптеров) на объекты системы и выбора подходящей модели квадрокоптера. Приводится описание программной реализации модуля вычисления характеристик графа.

Ключевые слова: алгоритм поиска кратчайших путей графа, алгоритм Форда–Беллмана, нечеткий граф, ГН-граф, радиус, диаметр, разнотипные связи, множественные связи, система охраны, программная реализация

Введение. Проблемы исследования *сложных технических систем* (СТС) с разнотипными информационными потоками остаются актуальными. Так, на примере задачи обеспечения охраны протяженного периметра они рассмотрены в работе [1]. Особое внимание уделяется решению практической задачи планирования совместных действий объектов таких систем, для моделирования которых целесообразно использование ГН-графов (Graph-Hypergraph) – нечетких графов с множественными и разнотипными связями [2]) и определенных алгоритмических средств, в том числе синтеза алгоритма пропорционального разделения ГН-графа и средств вычисления его метрик. Традиционно ориентированные связи графа названы дугами, неориентированные – ребрами, а в общем случае использован термин «связи».

Для вычисления метрических характеристик графа (подграфа), таких как радиус, диа-

метр и другие, необходимо сформировать матрицу расстояний между парами вершин графа. Сложности вычисления расстояний на ГН-графах обуславливают синтез новых или модификацию известных алгоритмов поиска кратчайших путей. Решению этой задачи в основном и посвящена данная работа.

Известно большое количество алгоритмов поиска кратчайшего пути в графе, в том числе алгоритм Ли, алгоритм Дейкстры, алгоритмы Форда–Беллмана и Флойда–Уоршелла.

Оригинальный алгоритм Ли не может быть применен к взвешенным графам. Для решения этой проблемы в работе [3] предлагаются средства перехода от взвешенного к невзвешенному графу, что существенно увеличивает количество вершин графа, а следовательно, приводит к росту времени вычислений.

Алгоритм Дейкстры не может работать с отрицательными весами на ребрах графа, к тому же вычислительная сложность такого алгорит-

ма и большей части его модификаций составляет $O(n^2 + m)$, где n и m – количество вершин и ребер графа соответственно [4, 5]. Вычислительная сложность алгоритма Флойда–Уоршелла имеет более высокое численное значение – $O(n^3)$ [6].

Алгоритм Форда–Беллмана [7, 8] имеет более низкую оценку вычислительной сложности по сравнению с рассмотренными алгоритмами, а именно – $O(n \times m)$ [6], и позволяет работать с ребрами отрицательного веса в ориентированных или неориентированных графах. Хотя в нечетких графах отрицательные веса ребер не используются, эта функция может быть полезна в будущем.

Для поиска кратчайших путей в графах с кратными односторонними ребрами известен алгоритм, предложенный в работе [9], но использование его для решения поставленной в данном исследовании задачи вызывает затруднения ввиду проблемы разнотипности ребер.

В представленной статье рассматривается проблема моделирования больших и сложных технических систем, а потому их графовая модель может быть довольно объемной по количеству вершин и связей и, кроме того, динамически изменяющейся во времени. Все это обосновано анализом существующих алгоритмов поиска кратчайших путей на больших и динамически изменяющихся графах.

Примеры таких алгоритмов рассмотрены в работах [10–12]. В общем случае описанные в них подходы можно сформулировать так: большой граф случайным образом разбивается сначала на кластеры, потом в случае необходимости – на подкластеры, а затем выполняется перекомпоновка полученных кластеров (подкластеров) с учетом определенных требований, например, в зависимости от компоненты связности графа или других параметров [10]. Вычисление кратчайших путей происходит внутри кластеров (подкластеров), а затем между ними, что приводит к получению итоговой матрицы расстояний. Однако предложенные подходы не учитывают разнотипные связи в ГН-графах, а также характеризуются избыточностью действий по перекомпоновке кластеров, что влечет за собой рост временных затрат на вычисления. Тем не менее подходы по разделению графа на части могут быть полезны в данном исследовании.

Обзор известных алгоритмов поиска кратчайшего пути в графе позволяет сделать следующие выводы. Наиболее подходящим алгоритмом поиска кратчайших путей для использования

в ГН-графе является алгоритм Форда–Беллмана, для которого имеются реализации матричным и списковым способами представления графа. Однако использование смешанных односторонних, разнотипных и множественных связей в виде вектора в ГН-графе [13], а также возможность использования графов больших размеров обуславливают необходимость модификации оригинального алгоритма Форда–Беллмана.

В рамках данной статьи для планирования совместных действий объектов СТС предложено развивать подходы, основанные на использовании графовых моделей и алгоритмических средств.

Подходы к моделированию СТС для решения практических задач

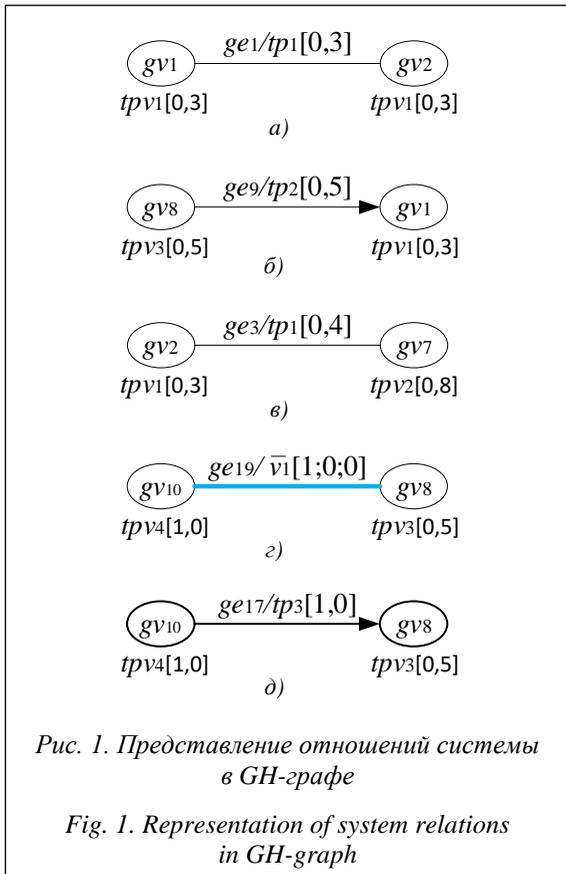
Для моделирования СТС предполагается использование графовой модели и ряда алгоритмов. Рассмотрим предложенные подходы на примере решения задачи обеспечения безопасности системы охраны протяженного периметра (далее – система). Предлагается последовательность следующих действий.

1. Разработка графовой модели.

В работах [1, 2] обосновано использование в качестве модели системы охраны ГН-графа, который позволяет представить разнотипные объекты, а также возможные разнотипные и сложные отношения в системе.

Объектами системы являются: наземные объекты охраны, включая стационарные объекты (здания и др.) и мобильные (транспорт, люди, роботы); потенциальные нарушители охраняемого периметра, в том числе люди, наземные роботизированные платформы; квадрокоптеры или *беспилотные летательные аппараты* (БПЛА); ЛПП – лицо, принимающее решение, в качестве которого используется главный компьютер или вычислительная система, предназначенные для приема и обработки информации о состоянии объектов охраны, квадрокоптеров и возможных действиях нарушителей территории.

В ГН-графе для представления объектов системы используются разнотипные вершины, а для представления отношений между объектами системы – разнотипные и множественные связи между вершинами. В работе [1] рассмотрен пример модели на основе ГН-графа, который задает часть реальной системы. В рамках данного исследования продемонстрирован пример представления на ГН-графе отдельных возможных отношений в системе (рис. 1).



На рисунке 1а ребро ge_1 типа tp_1 представляет отношение «удаленность объектов охраны», вес ребра $\mu_1 = 0,3$ соответствует степени удаленности объектов (низкий уровень). Объекты охраны представлены в ГН-графе вершинами gv_1, gv_2 типа tp_{v1} с весами $\eta_1 = \eta_2 = 0,3$, что соответствует низкому уровню значимости объектов системы.

На рисунке 1б вершина gv_8 типа tp_{v3} с весом $\eta_8 = 0,5$ представляет квадрокоптер, а ее вес соответствует среднему уровню значимости. Дуга ge_9 типа tp_2 соответствует отношению «наблюдать», а вес дуги $\mu_9 = 0,5$ – степени удаленности квадрокоптера и объекта охраны (средний уровень).

На рисунке 1в отношение «удаленность объектов охраны» задано ребром ge_3 типа tp_1 ГН-графа. Вершина gv_7 типа tp_{v2} с весом $\eta_7 = 0,8$ представляет потенциального нарушителя, а ее вес соответствует высокому уровню значимости.

В ГН-графе множественная связь в виде вектора \bar{v}_1 – ребро ge_{19} с весами $\mu_{19,1} = 1, \mu_{19,2} = 0, \mu_{19,3} = 0$ (рис. 1г) соответствует отношению «передача информации по каналам», в качестве каналов используются $v_{1,1}$ – Wi-fi, $v_{1,2}$ – Bluetooth, $v_{1,3}$ – УКВ. При этом передача информа-

ции в системе происходит между квадрокоптером и ЛПП (соответственно вершины графа gv_8 и gv_{10}). Вес вершины gv_{10} ($\eta_{10} = 1,0$) соответствует высокому уровню значимости объекта. Применение множественных связей в виде вектора в ГН-графе позволяет уменьшить время вычисления расстояний между вершинами не менее, чем в 1,2 раза, в зависимости от размера модели (в работе [13] проведено сравнение моделей одинаковой размерности (от 100 до 1 000 вершин) на основе ГН-графа и графов, допускающих использование только разнотипных или однотипных связей).

На рисунке 1д дуга ge_{17} типа tp_3 ГН-графа представляет отношение «управлять движением квадрокоптера», а вес дуги $\mu_{17} = 1,0$ – степень соответствия заданному направлению движения (максимальное значение).

Таким образом, на рассмотренных примерах можно сделать вывод об эффективности применения ГН-графа для моделирования СТС, математический аппарат которого позволяет задать возможные отношения в системе и уменьшить время анализа системы с разнотипными информационными потоками.

После представления системы в виде ГН-графа для решения задачи формирования зон влияния объектов системы (в данном случае – квадрокоптеров) необходимо применить один за другим два алгоритма: алгоритм пропорционального разделения ГН-графа и модифицированный алгоритм Форда–Беллмана поиска кратчайших путей ГН-графа. Особенностью применения данных алгоритмов является предварительная подготовка графовой модели, которая заключается в возможности фильтрации вершин и связей ГН-графа по типам, что соответствует критериям использования алгоритмов.

2. Использование алгоритма пропорционального разделения ГН-графа.

В работе [1] подробно рассмотрен сам алгоритм пропорционального разделения ГН-графа и показано его применение для решения практической задачи формирования зон влияния объектов системы. Используя на начальном этапе в данном алгоритме критерии выбора типов вершин и связей, можно оставить вершины типов tp_{v1} и tp_{v2} , что соответствует объектам охраны и потенциальным нарушителям системы, и связи между такими вершинами, представляющие отношения «удаленность объектов», веса которых соответствуют степени удаленности объектов охраны и/или потенциальных нарушителей системы.

Полученная таким образом графовая модель фактически представляет собой карту, на которой отмечены объекты охраны, потенциальные нарушители и расстояния между ними. Теперь такой граф в результате применения данного алгоритма делится на k частей (подграфов), где k – количество квадрокоптеров.

3. *Использование модифицированного алгоритма Форда–Беллмана поиска кратчайших путей ГН-графа и вычисление метрических характеристик графа.*

Данный модифицированный алгоритм применяется к каждому полученному подграфу поиска кратчайших путей ГН-графа, в результате чего формируются k матриц расстояний между вершинами. На основе полученных матриц вычисляются метрические характеристики подграфов. После этого выполняется сравнение на соответствие метрик подграфов и технических характеристик квадрокоптеров, в результате чего принимается решение о формировании или неформировании зон влияния объектов. Если хотя бы одна зона влияния объекта не может быть сформирована, принимается решение о повторном применении обоих алгоритмов с измененным значением k .

Рассмотрим подробнее предлагаемый в работе алгоритм.

Данный алгоритм предназначен для формирования матрицы расстояний ГН-графа. Модификация алгоритма заключается в возможности поиска расстояний для полных или усеченных множеств вершин и/или связей графа, например, для связей заданного типа. Это составляет научную новизну модифицированного алгоритма и позволяет выполнить анализ разнотипных информационных потоков в СТС. Вычислительная сложность модифицированного алгоритма оценивается величиной $O(n \times m)$, как и исходного алгоритма Форда–Беллмана. Однако при этом снижается время работы алгоритма ГН-графа за счет использования множественных связей в виде вектора, объединяющего ряд разнотипных связей [13].

Критерием вычисления кратчайших путей графа является учет полного или усеченного множества связей и вершин в соответствии с требованиями практической задачи.

Модифицированный алгоритм Форда–Беллмана поиска кратчайших путей ГН-графа основан на следующих формальных соотношениях.

1. Формально ГН-граф

$$G = (Gv, Ge) \tag{1}$$

задается нечеткими конечными множествами вершин $Gv = \{gv_i \mid i = 1, 2, \dots, n\}$ (gv – graph

vertex) и связей $Ge = \{ge_l \mid l = 1, 2, \dots, m\}$, соединяющих пары вершин (ge – graph edge) [14, 15].

Множество вершин графа содержит подмножества вершин по типу $Gv = \{Gv(tpv_p)\}$.

Множество связей графа содержит подмножества однотипных связей $Ge(o)$, разнотипных связей $Ge(tp_k)$ и связей в виде вектора $Ge(v_e)$, позволяющих объединить несколько разнотипных связей: $Ge = \{Ge(o), Ge(tp_k), Ge(v_e)\}$.

2. Форма представления множественных и разнотипных связей ГН-графа в виде списка $Le(G)$:

$$Le(G) = Cons(\langle gv_i \rangle : \langle gv_j \rangle, [ige]). \tag{2}$$

Здесь в качестве головы списка выступает вершина $\langle gv_i \rangle$, от которой идет связь. Хвост списка – вершина $\langle gv_j \rangle$, к которой идет связь от вершины $\langle gv_i \rangle$, и инцидентная им связь $[ige]$. Все элементы списка содержат свои атрибуты. Атрибутами вершины $Av = \langle tpv, \eta \rangle$ являются тип tv и вес η , а атрибутами связи – $Ae = \langle Por, Ptp, Pv, \mu \rangle$, где Por – признак ориентации; Pt – признак типа; Pv – признак вектора; μ – вес связи.

Признак ориентации Por :

$$Por = \begin{cases} 0, & \text{если неориентированная связь,} \\ 1, & \text{если ориентированная связь.} \end{cases} \tag{3}$$

Признак типа Pt :

$$Pt = \begin{cases} 0, & \text{если однотипная связь,} \\ tp_i, & \text{если разнотипная связь.} \end{cases} \tag{4}$$

Признак вектора Pv :

$$Pv = \begin{cases} 0, & \text{если немножественная связь,} \\ v_i, & \text{если множественная связь.} \end{cases} \tag{5}$$

3. Определение критерия 1 поиска кратчайших путей в графе $Kv(G)$:

$$Kv(G) = \begin{cases} 0, & \text{если полное множество} \\ & \text{вершин графа } Gv, \\ tpv_i, & \text{если учитывается} \\ & \text{определенный тип вершин} \\ & \text{графа.} \end{cases} \tag{6}$$

4. Определение критерия 2 поиска кратчайших путей в графе: $Ke(G) = 0$, если учитывается полное множество связей графа Ge .

В случае усеченного множества связей графа критерий $Ke(G)$ вычисляется как

$$Ke(G) = \begin{cases} 0, & \text{если однотипная связь,} \\ tp_i, & \text{если разнотипная связь,} \\ v_i, & \text{если множественная связь} \\ & \text{в виде вектора.} \end{cases} \tag{7}$$

5. Используется функция $GetOrAlg$, реализующая выполнение оригинального алгоритма Форда–Беллмана [3, 4].

6. В результате работы модифицированного алгоритма Форда–Беллмана поиска кратчайших путей GH-графа формируется матрица расстояний в зависимости от выбранных критериев: $Ad(G) = \{d(gv_i, gv_j)\}$, где $i, j = 1, 2, \dots, (n - 1)$.

На рисунке 2 приведена блок-схема модифицированного алгоритма Форда–Беллмана поиска кратчайших путей GH-графа с учетом соотношений (1)–(7).

Опишем алгоритм по шагам.

Шаг 1. Загрузка исходного графа G , представленного в виде списка ребер $Le(G)$.

Шаг 2. Формирование множества вершин и связей графов Gv и Ge и значения переменной n .

Шаг 3. Ввод пользователем значения $Kv(G)$ (см. (6)).

Шаг 4. Проверка на определение критерия 1 поиска кратчайших путей в графе. Если $Kv(G) = 0$,

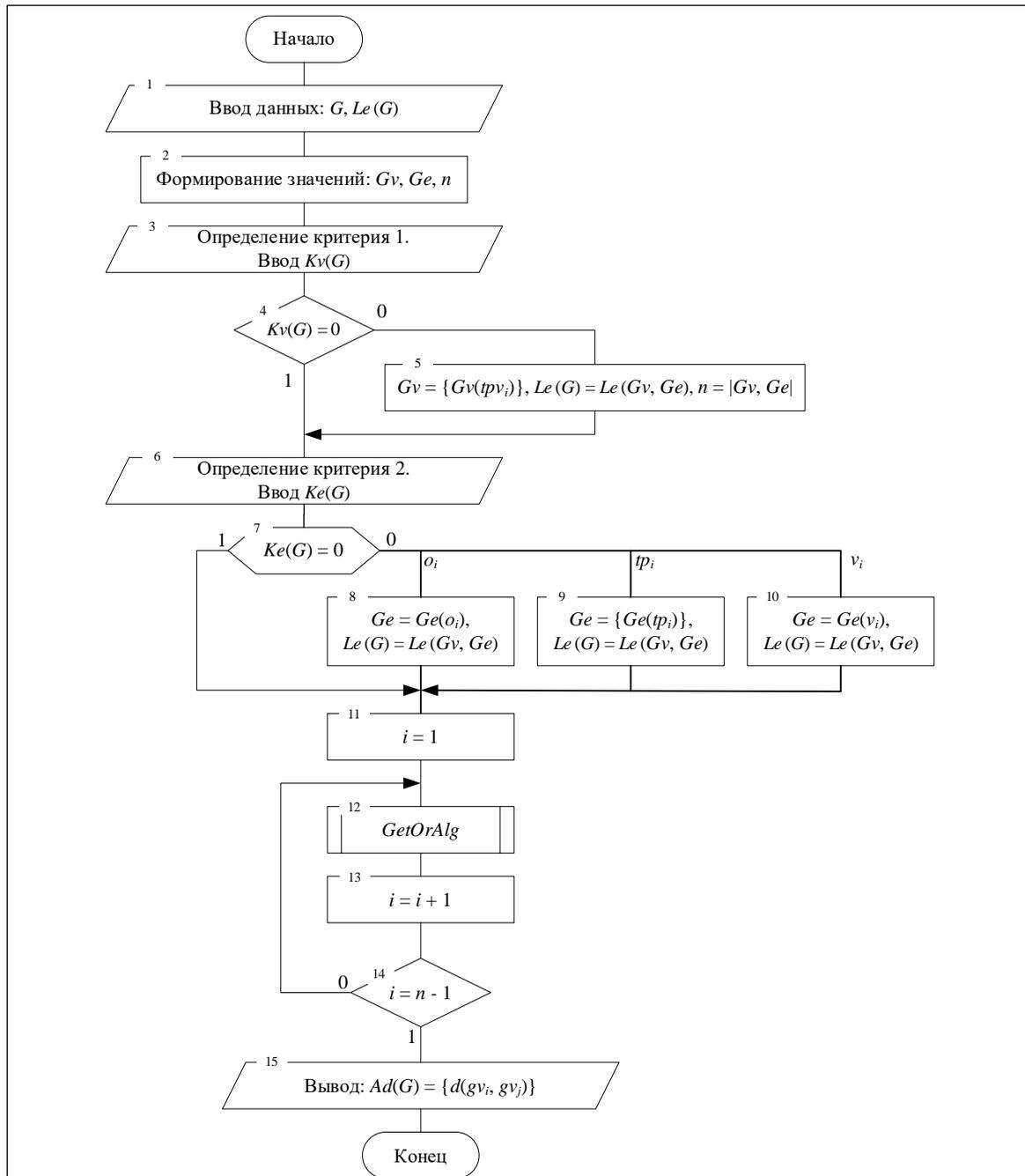


Рис. 2. Модифицированный алгоритм поиска кратчайших путей GH-графа

Fig. 2. Modified algorithm for finding GH-graph shortest ways

то выполняется переход к шагу 6 для определения критерия 2, иначе – переход к шагу 5.

Шаг 5. Если $Kv(G) = \{tpv_i\}$, то в списке ребер и во множестве вершин исходного графа остаются только вершины заданного типа tpv_i : $Gv = \{Gv(tpv_i)\}$, $Le(G) = Le(Gv, Ge)$. Формируется новое значение переменной $n = |Gv|$. Затем выполняется переход к шагу 6.

Шаг 6. Определение критерия 2 поиска кратчайших путей в графе путем ввода значения $Ke(G)$ (см. (7)).

Шаг 7. Если $Ke(G) = 0$, то учет полного множества связей графа Ge . Иначе – переход к шагам 8–10.

Шаги 8–10. В случае усеченного множества связей графа определение критерия $Ke(G)$ по формуле (7), таким образом, в списке остаются только связи выбранного подмножества связей, например, только однотипные связи $Ge(o_j)$, или связи выбранного типа $Ge(tp_j)$, или связи в виде заданного вектора конкретного индекса $Ge(v_j)$.

Шаги 11–14. Выполнение функции $GetOrAlg$, реализующей оригинальный алгоритм Форда–Беллмана.

Шаг 15. Вывод матрицы расстояний ГН-графа $Ad(G)$ в зависимости от выбранных критериев 1 и 2.

Таким образом, предложенный алгоритм обладает элементами научной новизны, сохраняет вычислительную сложность исходного алгоритма Форда–Беллмана и позволяет снизить время вычислений.

Формирование зон влияния объектов системы при помощи алгоритмов

Продемонстрируем предложенный подход к решению задачи формирования зоны влияния квадрокоптера K_1 . Примем, что в наличии имеются модели с учетом их технических характеристик, приведенных в таблице.

Характеристики квадрокоптеров

Characteristics of the quadcopter

Модель квадрокоптера	Диаметр обзора (км)	Время работы аккумулятора (мин.)
DJI Matrice 300 RTK	0,5	55
DJI Mavic 2 Enterprise Advanced	0,95	150
DJI Mavic 3 Cine Premium Combo	0,4	46

Поскольку ГН-модель является нечетким графом, переведем диаметры из км в условные единицы (1 км = 1 ед.), то есть полагаем, что 1 км – максимальная степень удаленности объектов в рассматриваемой системе.

Требуется определить наиболее подходящую модель квадрокоптера для обеспечения функций охраны участка периметра. Для этого исходную графовую модель [1] представим с учетом соотношений (6) и (7) $Kv(G) = \{tpv_1, tpv_2\}$, $Ke(G) = \{tp_1\}$, как показано на рисунке 3. Тогда в списке $Le(G)$ (соотношение (2)) остаются вершины и связи выбранного типа.

Полученный граф разделим на подграфы при помощи алгоритма пропорционального разделения ГН-графа с учетом выбранных критериев 1 и 2 (соотношения (6) и (7)). Полученный подграф G_1 изображен на рисунке 4. На рисунке 4а вершины подграфа gv_1-gv_3 типа tpv_1 с весами $\eta_i = 0,3$ представляют объекты охраны OO_1-OO_3 , а вершина gv_7 типа tpv_2 с весом $\eta_i = 0,8$ – потенциального нарушителя ПН₁. Степень удаленности объектов характеризуется весами ребер ge_1-ge_5 типа tp_1 ($\mu_1 = 0,3$, $\mu_2 = 0,5$, $\mu_3 = 0,4$, $\mu_4 = 0,7$, $\mu_5 = 0,6$).

Используя предложенный модифицированный алгоритм Форда–Беллмана поиска кратчайших путей ГН-графа, для подграфа G_1 сформи-

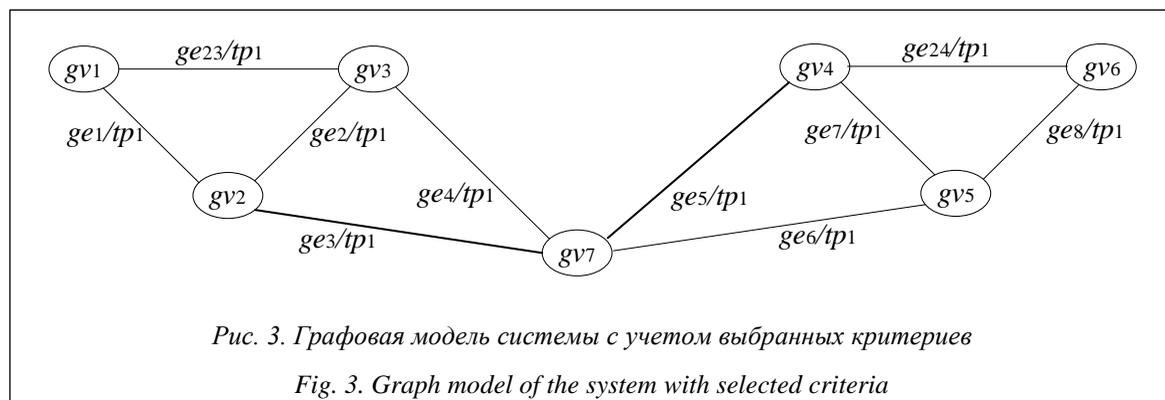
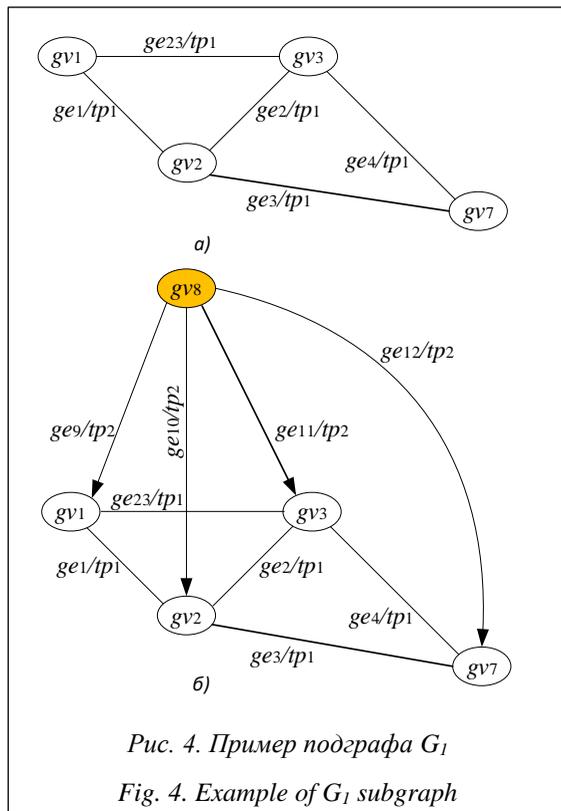


Рис. 3. Графовая модель системы с учетом выбранных критериев

Fig. 3. Graph model of the system with selected criteria



руем матрицу расстояний $Ad(G_1)$. Затем с помощью стандартных средств вычисления [16] определим метрические характеристики подграфа – радиус $r(G_1) = 0,5$ и диаметр $d(G_1) = 0,7$.

С учетом характеристик квадрокоптеров для обеспечения функций охраны участка периметра подходит только модель DJI Mavic 2 Enterprise Advanced с максимальным диаметром обзора 0,95. На подграфе такой квадрокоптер K_1 может быть представлен вершиной gv_8 типа tp_3 с весом $\eta_8 = 0,5$, как показано на рисунке 4б. Здесь все дуги $ge_9 - ge_{12}$ типа tp_2 (отношение «наблюдать») с соответствующими весами.

Таким образом, предложенный подход, основанный на применении ГН-графа и двух алгоритмов (алгоритма пропорционального разделения графа и модифицированного алгоритма Форда–Беллмана), позволяет уменьшить время анализа разнотипных данных в СТС и может быть использован в общем случае для решения задачи классификации на графах, пример которой рассмотрен на задаче формирования зон влияния объектов системы.

Программная реализация

Структура программного комплекса, позволяющего реализовать модели на основе ГН-

графов, представлена и подробно описана в [1]. Разработанные в рамках данного исследования алгоритмические средства составляют блок вычисления характеристик ГН-графа в программном комплексе [17]. Обновление предложенного ранее программного комплекса заключается в модификации графовой модели и алгоритмов за счет допустимости разнотипных вершин. Структура блока вычисления характеристик обновленного программного комплекса моделирования системы на основе ГН-графа изображена на рисунке 5.

Здесь модуль ввода/вывода данных служит для организации ввода исходного графа G , представленного в виде списка ребер $Le(G)$, и вывода численных значений характеристик графа, в том числе радиуса, диаметра, подмножества центральных и периферийных вершин и др. Для хранения входных, выходных данных и промежуточных результатов работы блока служит модуль хранения. В качестве промежуточных результатов данный модуль позволяет сохранить таблицу расстояний и эксцентриситеты вершин графа.

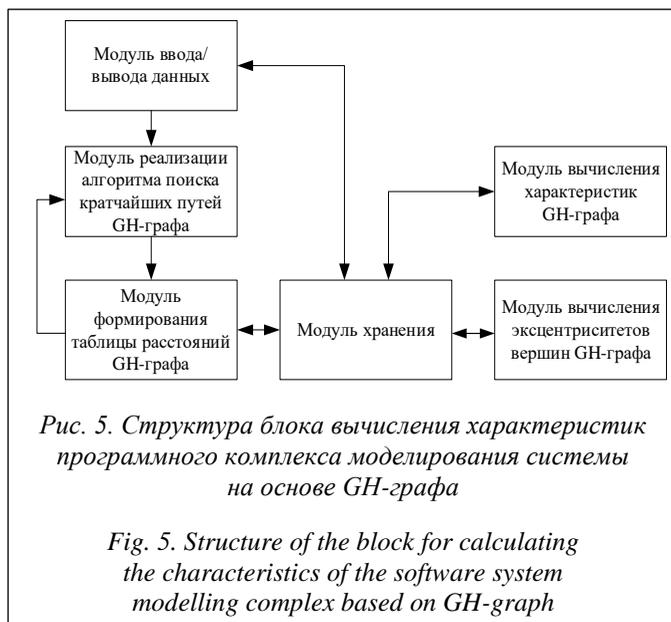
Последовательность вычисления характеристик ГН-графа:

- реализация модифицированного алгоритма Форда–Беллмана поиска кратчайших путей ГН-графа, в результате чего формируется таблица расстояний;
- на основании полученной таблицы расстояний вычисление численных значений эксцентриситетов вершин графа;
- в зависимости от полученных значений эксцентриситетов вершин вычисление метрических характеристик и определение подмножеств центральных и периферийных вершин графа.

Для реализации программного комплекса использован язык программирования C++ (включая библиотеку Qt). Комплекс может функционировать в различных операционных системах, и планируется его использование с применением SQL или NoSQL БД, например, графовой СУБД Neo4j и т.д. [18].

Заключение

В статье описан подход моделирования СТС, основанный на использовании ГН-графа и двух алгоритмов (алгоритма пропорционального разделения графа и модифицированного алгоритма Форда–Беллмана) и позволяющий выполнять анализ разнотипных данных. Предложен модифицированный алгоритм поиска



кратчайших путей Форда–Беллмана, приведены блок-схема алгоритма и его пошаговое описание. Показаны преимущества предложенного подхода, заключающиеся в следующем: вычислительная сложность модифицированного алгоритма не отличается от сложности

исходного алгоритма Форда–Беллмана; время анализа разнотипных данных СТС снижается за счет использования множественных связей в виде вектора в ГН-графе.

Для ГН-модели, разработанной ранее, рассмотрены возможности применения алгоритмов для решения задачи формирования зон влияния объектов в системе на примере системы охраны протяженного периметра. Показаны возможности использования предложенного подхода для решения данной задачи, в том числе выбор конкретной модели квадрокоптера подсистемы и определение количества квадрокоптеров с заданными техническими характеристиками, необходимые для организации видеонаблюдения системы охраны. Использование предложенных алгоритмов на ГН-графе позволяет решить задачу классификации, что продемонстрировано на рассмотренном примере. Реализованы блок вычисления характеристик графа в ПК, а также соответствующее ПО для поддержки предложенного подхода на основе графовой модели.

Список литературы

1. Мунтян Е.Р. Разработка алгоритма пропорционального разделения ГН-графа для формирования зон влияния объектов в сложных технических системах // Программные продукты и системы. 2023. Т. 36. № 3. С. 378–387. doi: 10.15827/0236-235X.143.378-387.
2. Еремеев А.П., Мунтян Е.Р. Разработка онтологии на основе графов с множественными и разнотипными связями // Искусственный интеллект и принятие решений. 2021. № 3. С. 3–18. doi: 10.14357/20718594210301.
3. Сысов В.В. Итерационный алгоритм поиска кратчайшего пути в невзвешенном неориентированном графе // Современные информационные технологии и ИТ-образование. 2021. Т. 17. № 3. С. 585–592.
4. Близнякова Е.А., Куликов А.А., Куликов А.В. Сравнительный анализ методов поиска кратчайшего пути в графе // Архитектура, строительство, транспорт. 2022. № 1. С. 80–87. doi: 10.31660/2782-232X-2022-1-80-87.
5. Хренов В.В. Модификации алгоритма Дейкстры для поиска кратчайшего пути // АИ. 2023. № 13. С. 12–17.
6. Cormen Th.H., Leiserson Ch.E., Rives R.L., Stein C. Introduction to Algorithms. MIT Press and McGraw-Hill Publ., 2022, 1312 p.
7. Bellman R. On a routing problem. Quart. Appl. Math., 1958, vol. 16, pp. 87–90. doi: 10.1090/QAM/102435.
8. Ford L.R., Fulkerson D.R. Flows in Networks. Princeton University Press, 1962, 216 p.
9. Смирнов А.В. Задача о кратчайшем пути в кратном графе // Моделирование и анализ информационных систем. 2017. Т. 24. № 6. С. 788–801. doi: 10.18255/1818-1015-2017-6-788-801.
10. Zhang J., Li W., Yuan L., Qin L., Zhang Y., Chang L. Shortest-path queries on complex networks: Experiments, analyses, and improvement. PVLDB, 2022, vol. 15, no. 11, pp. 2640–2652. doi: 10.14778/3551793.3551820.
11. Ураков А.Р., Тимирязев Т.В. Алгоритм решения динамической задачи поиска кратчайших расстояний в графе // Управление большими системами. 2017. № 65. С. 60–86.
12. Агафонов А.А., Мясников В.В. Метод определения надежного кратчайшего пути в стохастической сети с использованием параметрически заданных устойчивых распределений вероятностей // Тр. СПИИРАН. 2019. Т. 3. № 18. С. 558–582. doi: 10.15622/sp.2019.18.3.557-581.
13. Мунтян Е.Р. Реализация нечеткой модели взаимодействия объектов сложных технических систем на основе графов // Программные продукты и системы. 2019. Т. 32. № 3. С. 411–418. doi: 10.15827/0236-235X.127.411-418.
14. Minaev Y.N., Filimonova O.Y., Minaeva J.I., Filimonov A. Fuzzy mathematics with limited possibilities for assigning membership functions. Cybernetics and Systems Analysis, 2020, vol. 56, pp. 29–39. doi: 10.1007/s10559-020-00218-9.
15. Дандыбаев С.Т. Нечеткие множества с нечеткими функциями принадлежности // Теория и практика современной науки. 2021. № 1. С. 130–133.
16. Muntyan E.R., Melnik E.V. The graph-based analysis of structural delays in distributed multiprogram systems of information processing. JPCS, 2020, vol. 1661, no. 1, art. 012061. doi: 10.1088/1742-6596/1661/1/012061.

17. Мунтян Е.Р. Программный модуль для моделирования взаимодействия акторов и групп акторов на основе графов: Свид. о регистр. ПрЭВМ № 2018665593. Рос. Федерация, 2018.

18. Еремеев А.П., Панявин Н.А. Унификация модели представления данных и преобразование форматов на основе нереляционной СУБД Neo4j // Программные продукты и системы. 2022. Т. 35. № 4. С. 549–556. doi: 10.15827/0236-235X.140.549-556.

Software & Systems

doi: 10.15827/0236-235X.142.354-363

2024, 37(3), pp. 354–363

Modifying a GH-graph shortest path search algorithm for analyzing complex technical systems

Evgeniya R. Zyablova ¹✉¹ Southern Federal University, Taganrog, 347900, Russian Federation**For citation**

Zyablova, E.R. (2024) 'Modifying a GH-graph shortest path search algorithm for analyzing complex technical systems', *Software & Systems*, 37(3), pp. 354–363 (in Russ.). doi: 10.15827/0236-235X.142.354-363

Article info

Received: 25.02.2024

After revision: 29.03.2024

Accepted: 09.04.2024

Abstract. The paper proposes one of the approaches to modelling complex technical systems on the example of solving the problem of forming zones of influence of the extended perimeter security system objects. The approach uses the so-called GH-graph (fuzzy graph with different types of vertices and multiple and different types of links) and certain algorithmic tools. This paper follows author's previous works that detail the possible graph model of the system (or its part), the algorithm of GH-graph proportional partitioning, and its application to solve the problem. This study includes synthesis of a modified Ford-Bellman algorithm for finding the shortest paths of a GH-graph. The modified algorithm allows analyzing different types of information flows in complex technical systems. It is characterized by the ability to find distances for complete or truncated sets of vertices and/or graph links, e.g. for given type links. There is a list representation of multiple and different types of GH-graph links. The formulated criteria for finding shortest paths in the GH-graph are the following: criterion 1 – selecting a type (types) of vertices involved in the algorithm; criterion 2 – selecting the types (vectors) of edges involved in the algorithm. To this end, it is possible to leave the complete sets of vertices and edges of the model in the graph list representation. The determined value of the computational complexity of the modified algorithm does not exceed the complexity value of the original Ford-Bellman algorithm. The operating time of the proposed algorithm is reduced due to using multiple edges in the GH-graph as a vector, which allows combining a number of different types of edges. The result of the modified algorithm is a distance matrix considering complete or truncated sets of vertices and/or edges for subsequent calculation of graph model metric characteristics by means determined by early research. The authors consider the proposed approaches to modeling complex technical systems in detail by determining the zones of influence of technical devices (quadrotors) on system objects and selecting a suitable quadrotor model. There is a brief description of the software implementation of the graph characteristics calculation module.

Keywords: graph shortest path search algorithm, Ford-Bellman algorithm, fuzzy graph, GH-graph, radius, diameter, dissimilar links, multiple links, security system, software implementation

References

1. Muntyan, E.R. (2023) 'Developing a GH-graph proportional separation algorithm to form of object influence zones in complex technical systems', *Software & Systems*, 36(3), pp. 378–387 (in Russ.). doi: 10.15827/0236-235X.143.378-387.
2. Eremeev, A.P., Muntyan, E.R. (2021) 'Development of an ontology based on graphs with multiple and different types of edges', *Artificial Intelligence and Decision Making*, (3), pp. 3–18 (in Russ.). doi: 10.14357/20718594210301.
3. Sysoev, V.V. (2021) 'Iterative algorithm for finding the shortest ways in an unweighted undirected graph', *Modern Information Technologies and IT-Education*, 17(3), pp. 585–592 (in Russ.).
4. Bliznyakova, E.A., Kulikov, A.A., Kulikov, A.V. (2022) 'Comparative analysis of methods for finding the shortest distance in a graph', *Architecture, Construction, Transport*, (1), pp. 80-87 (in Russ.). doi: 10.31660/2782-232X-2022-1-80-87.
5. Khrenov, V.V. (2023) 'Modifications of the Dijkstra algorithm for finding the shortest path', *Actual Research*, (13), pp. 12–17 (in Russ.).
6. Cormen, Th.H., Leiserson, Ch.E., Rives, R.L., Stein, C. (2022) *Introduction to Algorithms*. MIT Press and McGraw-Hill Publ., 1312 p.

7. Bellman, R. (1958) 'On a routing problem', *Quart. Appl. Math.*, (16), pp. 87–90. doi: 10.1090/QAM/102435.
8. Ford, L.R., Fulkerson, D.R. (1962) *Flows in Networks*. Princeton University Press, 216 p.
9. Smirnov, A.V. (2017) 'The shortest path problem for a multiple graph', *Modeling and Analysis of Information Systems*, 24(6), pp. 788–801 (in Russ.). doi: 10.18255/1818-1015-2017-6-788-801.
10. Zhang, J., Li, W., Yuan, L., Qin, L., Zhang, Y., Chang, L. (2022) 'Shortest-path queries on complex networks: Experiments, analyses, and improvement', *PVLDB*, 15(11), pp. 2640–2652. doi: 10.14778/3551793.3551820.
11. Urakov, A.R., Timiryayev, T.V. (2017) 'Algorithm for solving the dynamic problem of finding the shortest distances in a graph', *Large-Scale Systems Control*, (65), pp. 60–86 (in Russ.).
12. Agafonov, A.A., Myasnikov, V.V. (2019) 'Method for reliable shortest path determination in stochastic networks using parametrically defined stable probability distributions', *SPIIRAS Proc.*, 3(18), pp. 558–582 (in Russ.). doi: 10.15622/sp.2019.18.3.557-581.
13. Muntyan, E.R. (2019) 'Implementation of a fuzzy model of interaction between objects in complex technical systems based on graphs', *Software & Systems*, 32(3), pp. 411–418 (in Russ.). doi: 10.15827/0236-235X.127.411-418.
14. Minaev, Y.N., Filimonova, O.Y., Minaeva, J.I., Filimonov, A. (2020) 'Fuzzy mathematics with limited possibilities for assigning membership functions', *Cybernetics and Systems Analysis*, 56, pp. 29–39. doi: 10.1007/s10559-020-00218-9.
15. Dandybaev, S.T. (2021) 'Fuzzy sets with fuzzy membership functions', *Theory and Practice of Modern Sci.*, (1), pp. 130–133 (in Russ.).
16. Muntyan, E.R., Melnik, E.V. (2020) 'The graph-based analysis of structural delays in distributed multiprogram systems of information processing', *JPCS*, 1661(1), art. 012061. doi: 10.1088/1742-6596/1661/1/012061.
17. Muntyan, E.R. (2018) *A Software Module for Modeling the Interaction of Actors and Actors Groups Based on Graphs*. Pat. RF, № 2018665593.
18. Ereemeev, A.P., Paniavin, N.A. (2022) 'Unification of a data presentation model and format conversion based on a non-relational Neo4j DBMS', *Software & Systems*, 35(4), pp. 549–556 (in Russ.). doi: 10.15827/0236-235X.140.549-556.

Авторы

Зяблова Евгения Ростиславна¹, к.т.н.,
доцент, ermuntyan@sfedu.ru

¹ Южный федеральный университет,
г. Таганрог, 347900, Россия

Authors

Evgeniya R. Zyablova¹, Cand. of Sci. (Engineering),
Associate Professor, ermuntyan@sfedu.ru

¹ Southern Federal University,
Taganrog, 347900, Russian Federation

Основные принципы работы обобщенной регрессионной нейронной сети при заполнении пропущенных значений в наборах данных

Т.М. Татарникова ¹✉, В.В. Боженко ¹

¹ Санкт-Петербургский государственный университет аэрокосмического приборостроения, г. Санкт-Петербург, 190000, Россия

Ссылка для цитирования

Татарникова Т.М., Боженко В.В. Основные принципы работы обобщенной регрессионной нейронной сети при заполнении пропущенных значений в наборах данных // Программные продукты и системы. 2024. Т. 37. № 3. С. 364–368. doi: 10.15827/0236-235X.142.364-368

Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 09.04.2024

После доработки: 25.04.2024

Принята к публикации: 14.05.2024

Аннотация. В статье обсуждается актуальность заполнения пропущенных значений в исходном наборе данных на этапе их предобработки при решении задач анализа данных и машинного обучения. Предложено применение обобщенной регрессионной нейронной сети для решения задачи заполнения пропущенных значений в наборе исходных данных, что в сравнении со статистическим методом на основе среднего или медианного значения по столбцу предполагает учет возможных зависимостей между данными. Рассмотрены основные принципы работы обобщенной регрессионной нейронной сети, особенности ее архитектуры, преимущества и недостатки. Показано, что преимуществами обобщенной регрессионной нейронной сети являются быстрое обучение на небольшом объеме входных данных и прогнозирование пропущенных значений благодаря возможности аппроксимации сложных функций. Приведен алгоритм использования обобщенной регрессионной нейронной сети для восстановления пропусков. Алгоритм обучения нейронной сети является однопроходным, во время которого настраиваются веса связей между слоями сети, параметр радиальной базисной функции и скорость обучения. Целью обучения нейронной сети является минимизация ошибки прогнозирования, в качестве которой выбрана среднеквадратичная ошибка. Предложена схема заполнения пропущенных значений статистическим методом. Приведен алгоритм применения схемы заполнения пропусков, основанный на определении среднего по имеющимся значениям признака, то есть по данным, расположенным выше заполняемой ячейки столбца-признака. Прогнозирование пропущенных значений статистическим методом также оценивалось с помощью среднеквадратической ошибки. Продемонстрированы результаты обучения модели обобщенной регрессионной нейронной сети и применения статистического метода на валидационном наборе данных. Сравнение результатов заполнения пропущенных значений двумя методами показало преимущество обобщенной регрессионной нейронной сети на значительном (большом) наборе данных.

Ключевые слова: предварительная обработка данных, пропущенные значения, обобщенная регрессионная нейронная сеть, математическое ожидание, проверка ошибки заполнения пропущенных данных, валидационные данные

Введение. Заполнение пропущенных значений в наборах данных является важным этапом предобработки и может оказать значительное влияние на результаты анализа. Современные исследователи часто сталкиваются с проблемой пропущенных данных, которые необходимо обрабатывать, потому что отсутствие некоторых значений или их некорректное заполнение приводит к невозможности применения методов машинного обучения и принятия решения на основе имеющихся данных, делая анализ таких данных бесполезным [1]. Таким образом, нахождение эффективных методов для правильной обработки пропусков является актуальной задачей. Существуют различные методы заполнения таких значений, к наиболее простым относится заполнение средним или медианным значением по столбцу [1]. Однако такой подход к заполнению отсутствующих данных является сугубо статистическим и не

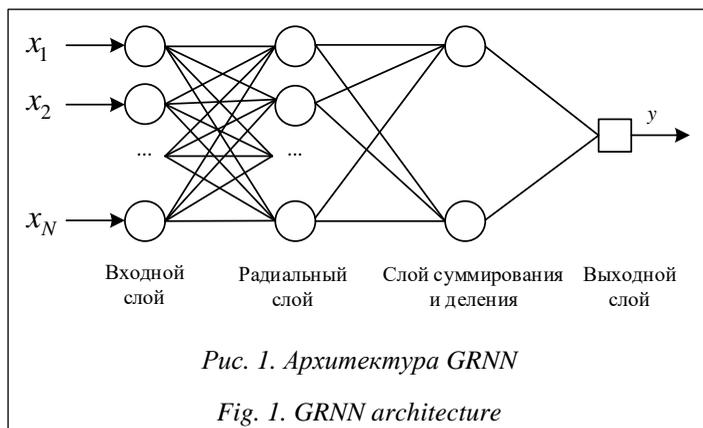
учитывает возможных зависимостей в данных, принадлежащих разным столбцам [2].

Если задачу восстановления пропусков числовых данных свести к задаче прогнозирования, то одним из перспективных подходов к ее решению является применение нейронных сетей. В данной работе рассматривается обобщенная регрессионная нейронная сеть GRNN (Generalized Regression Neural Network), которая показывает лучшие результаты прогнозирования в сравнении, например, с сетями прямого распространения [3, 4].

GRNN является разновидностью нейронных сетей с радиальным базисом, ее архитектура приведена на рисунке 1.

GRNN состоит из четырех слоев: входного, радиального, суммирования и деления, выходного.

На входной (первый) слой поступают наборы данных, функция активации нейронов входного слоя линейная:



$$f_i = (x_i), \quad i = \overline{1, N}, \tag{1}$$

где x_i – сигнал, поступающий на вход нейронов первого слоя.

Второй слой называется радиальным, каждый его нейрон воспроизводит гауссову поверхность отклика:

$$f_i = \exp\left(-\frac{x_i^2}{2\sigma^2}\right), \tag{2}$$

где x_i – сигнал, поступающий на вход нейронов второго слоя GRNN; $\sigma \in [0, 1]$ – параметр, определяющий радиус влияния каждой базисной функции и быстроту стремления к нулю при удалении от центра (рис. 2).

Слой суммирования и деления передает на первый нейрон этого слоя числитель – сумму произведений значений сигналов нейронов второго слоя на значения их функций активации:

$$\sum_{j=1}^J x_{2j} f_j, \quad j = \overline{1, J}, \tag{3}$$

и знаменатель – сумму значений функций активации:

$$\sum_{j=1}^J f_j, \quad j = \overline{1, J}. \tag{4}$$

Выходной слой содержит один нейрон, он вычисляет выходные данные путем деления части числителя на часть знаменателя и предназначен для оценки взвешенного среднего [5, 6]:

$$\hat{y}(\mathbf{X}) = \frac{\sum_{i=1}^n x_i \exp\left(-\frac{x_i^2}{2\sigma^2}\right)}{\sum_{i=1}^n \exp\left(-\frac{x_i^2}{2\sigma^2}\right)}. \tag{5}$$

Алгоритм обучения GRRN является однопроходным. Во время обучения настраиваются веса связей между слоями GRRN и параметры, такие как σ и скорость обучения. Цель обучения GRRN – минимизировать ошибку прогнозирования, в качестве которой, как правило, вы-

ступает среднеквадратичная ошибка (Mean Squared Error, MSE) [7, 8].

Достоинством сети GRNN можно считать определенность структуры: сеть вмещает в себя все обучающие данные. С другой стороны, такая структура нейронной сети является ее основным недостатком, поскольку при большом объеме обучающих данных скорость работы сети падает. Однако при небольшом объеме входных данных сеть способна быстро обучаться, например, в сравнении с сетью прямого распространения.

Также важно отметить, что GRNN позволяет учитывать нелинейные зависимости в данных. Это обстоятельство делает GRNN эффективным инструментом для аппроксимации сложных функций, а значит, и предсказания пропущенных значений [9, 10].

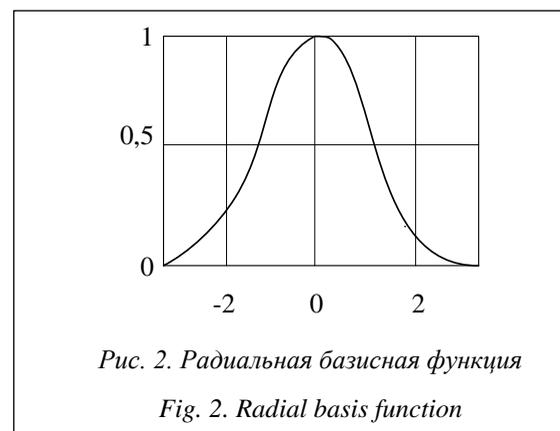
Таким образом, учитывая достоинства GRNN для решения задачи пропущенных значений в наборах данных, воспользуемся этим инструментом и сравним его со статистическим методом – заполнение средним значением по столбцу.

Описание эксперимента

В качестве входных наборов данных использованы клинично-лабораторные показатели – 142 000 наборов данных по общему анализу крови, в которых присутствуют показатели обмена железа, маркеры воспаления, ретикулоцитарные показатели и другие. Для некоторых показателей данные либо отсутствуют, либо некорректны (около 15 %).

Алгоритм заполнения пропущенных значений с применением GRNN будет следующим:

- проверить наличие пропусков в данных на этапе предварительной обработки данных;



- определить признак или признаки, в которых необходимо заполнить пропуски, – y ;
- выделить из полного набора данных X часть наборов \hat{X} , в которых значения в строках для этого признака полностью заполнены;
- обучить GRNN на наборе данных \hat{X} ;
- оценить MSE на валидационных данных из набора \hat{X} ;
- восстановить значения целевого столбца y в наборе данных $(X - \hat{X})$ с помощью обученной модели GRNN.

Для разработки модели GRNN использованы язык программирования Python, библиотеки Keras и PyGRNN. Перед обучением сети выполнена предварительная обработка данных, которая включала их нормализацию и удаление выбросов для более точного прогноза. Строки, в которых отсутствовали значения целевого столбца, в обучении не участвовали. Оставшиеся данные были поделены на тренировочную и валидационную выборки, на валидацию приходилась 1/4 часть.

На рисунке 3 представлены результаты MSE для 125 эпох обучения: на обучающей выборке ошибка обучения на последней эпохе составила 0.0144, на валидационной – 0.0377.

В ходе работы также был проанализирован набор данных, в котором количество заполненных данных для обучения составляло 150 строк. Следует отметить, что обучить модель GRNN для эффективного выявления закономерностей в данных не удалось. Можно предположить, что такой результат связан с сокращением количества наборов данных для обучения.

Схема заполнения пропущенных значений статистическим методом приведена на рисунке 4.

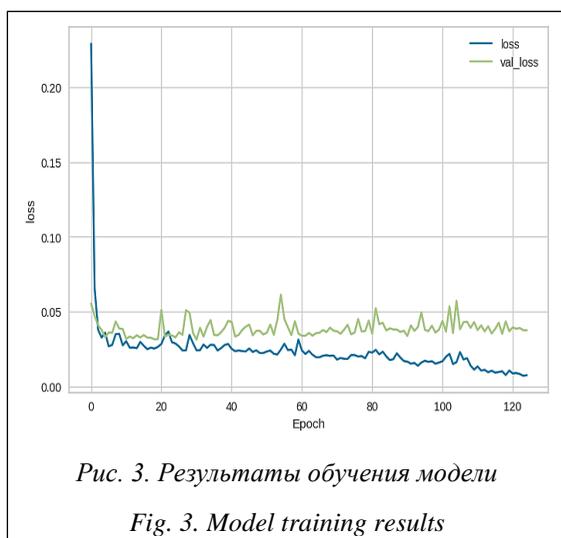


Рис. 3. Результаты обучения модели

Fig. 3. Model training results

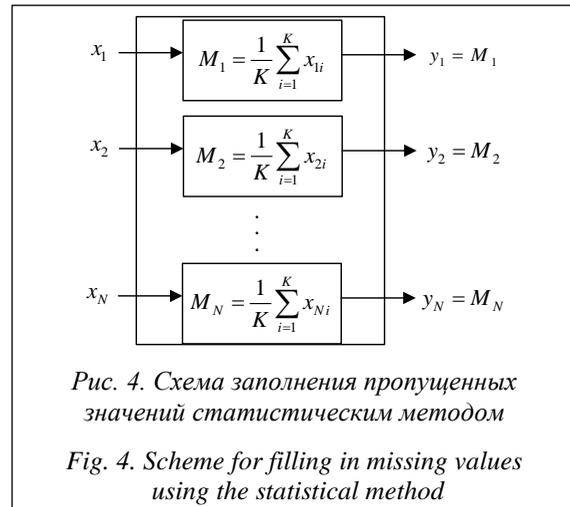


Рис. 4. Схема заполнения пропущенных значений статистическим методом

Fig. 4. Scheme for filling in missing values using the statistical method

Алгоритм заполнения пропущенных значений статистическим методом:

- проверить наличие пропусков в данных на этапе их предварительной обработки;
- определить номера столбцов-признаков, требующих заполнения, и для каждого такого столбца: а) найти следующую незаполненную ячейку a_{ij} , где i – номер строки, j – номер столбца-признака; б) найти математическое ожидание для заполненных данных столбца j , находящихся выше строки i : $M_j = \sum_{k=1}^i x_k$, и принять значение признака y_j равным M_j ; в) заполнить ячейку a_{ij} значением y_j ;
- оценить MSE на валидационных данных из набора \hat{X} ;
- восстановить значения целевого столбца y в наборе данных $(X - \hat{X})$ с помощью статистического метода.

В таблице приведены результаты MSE, полученные на валидационных данных после применения GRNN и статистического метода соответственно.

Значения MSE на разных объемах наборов данных

MSE values on different volumes of datasets

Число записей	GRNN	Статистический метод
150	0.401	0.127
1 000	0.309	0.109
5 000	0.101	0.102
50 000	0.072	0.091
100 000	0.039	0.088
142 000	0.038	0.072

Выводы

В работе представлены методы заполнения пропущенных значений в исходных наборах данных с использованием обобщенной регрессионной нейронной сети GRNN и статистического подхода.

Как показывают результаты эксперимента, GRNN представляет собой эффективный инструмент, который можно использовать для заполнения пропущенных значений на больших наборах данных, так как он позволяет автоматически извлекать зависимости из данных и выполнять прогнозы высокой точности.

В ходе работы проанализированы небольшие наборы данных с числом заполненных данных для обучения 1 000 строк и менее, для которых не удалось обучить модель GRNN с допустимым значением $MSE = 0,1$. Причина высокой ошибки обучения может быть связана с небольшим количеством данных для обучения. В то же время для небольшого объема наборов данных статистический метод показывает лучшие в сравнении с GRNN результаты.

Очевидно, что представляют интерес дальнейшее исследование зависимости эффективности данного метода от количества объектов для обучения и сравнение данного подхода с другими методами.

Список литературы

1. Bozhenko V.V., Tatarnikova T.M. Application of data preprocessing in medical research. Proc. WECONF, 2023, pp. 1–4. doi: 10.1109/WECONF57201.2023.10148004.
2. Сташкова О.В., Шестопа О.В. Использование искусственных нейронных сетей для восстановления пропусков в массиве исходных данных // Изв. вузов. Северо-Кавказский регион. Технич. науки. 2017. № 1. С. 37–42.
3. Андреев П.Г., Андреева Т.В., Юрков Н.К. Использование искусственной нейронной сети типа GRNN в задачах прогнозирования // Международная конференция по мягким вычислениям и измерениям. 2017. Т. 2. С. 63–66.
4. Шпаков А.В., Лавина Т.А. Применение нейронных сетей для аппроксимации экспериментальных данных // Тенденции развития науки и образования. 2022. № 84-2. С. 60–64. doi: 10.18411/trnio-04-2022-61.
5. Sharma S., Sharma S., Athaiya A. Activation functions in neural networks. IJEAST, 2020, vol. 4, no. 12, pp. 310–316. doi: 10.33564/ijeast.2020.v04i12.054.
6. Богданов П.Ю., Пойманова Е.Д., Красва Е.В., Веревкин С.А., Татарникова Т.М. Программные среды для изучения основ нейронных сетей // Программные продукты и системы. 2021. Т. 31. № 1. С. 145–150. doi: 10.15827/0236-235X.133.145-150.
7. Тарик Р. Создаем нейронную сеть. М.: Диалектика, 2017. 272 с.
8. Macpherson T., Matsumoto M., Gomi H., Morimoto J., Uchibe E., Hikida T. Parallel and hierarchical neural mechanisms for adaptive and predictive behavioral control. Neural Networks, 2021, vol. 144, pp. 507–521. doi: 10.1016/j.neunet.2021.09.009.
9. Yu S., Jiang F., Li L., Xie Y. CNN-GRNN for image sharpness assessment. In: LNIP. Proc. CNN, 2017, vol. 10116, pp. 50–61. doi: 10.1007/978-3-319-54407-6_4.
10. Крутиков А.К. Прогнозирование спортивных результатов в индивидуальных видах спорта с помощью обобщенно-регрессионной нейронной сети // Молодой ученый. 2018. № 12. С. 22–26.

Basic principles of generalized regression neural network when filling missing values in datasets

Tatiana M. Tatarnikova , Viktoriya V. Bozhenko ¹

¹ Saint Petersburg State University of Aerospace Instrumentation, Saint Petersburg, 190000, Russian Federation

For citation

Tatarnikova, T.M., Bozhenko, V.V. (2024) 'Basic principles of generalized regression neural network in filling missing values in datasets', *Software & Systems*, 37(3), pp. 364–368 (in Russ.). doi: 10.15827/0236-235X.142.364-368

Article info

Received: 09.04.2024

After revision: 25.04.2024

Accepted: 14.05.2024

Abstract. The paper discusses the relevance of filling missing values in the initial data set at the preprocessing stage when solving problems of data analysis and machine learning. The authors of the paper propose to use a generalized regression neural network to solve the problem of filling missing values in the initial data set. In comparison with the statistical method based on the mean or median value per column, it implies taking into account possible dependencies between data.

The paper considers the basic principles of the generalized regression neural network, its architecture features, advantages and disadvantages. It also shows that the advantages of the generalized regression neural network include fast training on a small amount of input data and the ability to predict missing values due to its capability to approximate complex functions. The authors also give an algorithm for using a generalized regression neural network for gap recovery. The algorithm is one-pass; it adjusts the weights of links between network layers, a radial basis function parameter, and a learning rate during one-pass training of the neural network. Training the neural network aims to minimize the prediction error, which is RMS error. There is a scheme for filling in the missing values using a statistical method. The paper presents an algorithm for applying the omission filling scheme based on determining the average feature according to the available values, that is the data located above the feature column cell to be filled in. The prediction of missing values by the statistical method was also evaluated using the mean square error. The authors demonstrate the results of training the generalized regression neural network model and applying the statistical method on a validation dataset. Comparison of the results of filling in missing values by two methods showed the advantage of the generalized regression neural network on a significant (large) dataset. **Keywords:** data preprocessing, missing values, generalized regression neural network, mathematical expectation, missing data filling error check, validation data

References

1. Bozhenko, V.V., Tatarnikova, T.M. (2023) 'Application of data preprocessing in medical research', *Proc. WECONF*, pp. 1–4. doi: 10.1109/WECONF57201.2023.10148004.
2. Stashkova, O.V., Shestopal, O.V. (2017) 'Using artificial neural networks to restore gaps in the source data array', *Bull. of Higher Educational Institutions. North Caucasus Region. Tech. Sci.*, (1), pp. 37–42 (in Russ.).
3. Andreev, P.G., Andreeva, T.V., Yurkov, N.K. (2017) 'Using an artificial neural network of the GRNN type in forecasting problems', *Proc. Int. Conf. SCM-2017*, 2, pp. 63–66 (in Russ.).
4. Shpakov, A.V., Lavina, T.A. (2022) 'Application of neural networks for approximation of experimental data', *Trends in the Development of Sci. and Education*, (84-2), pp. 60–64 (in Russ.). doi: 10.18411/trnio-04-2022-61.
5. Sharma, S., Sharma, S., Athaiya, A. (2020) 'Activation functions in neural networks', *IJEAST*, 4(12), pp. 310–316. doi: 10.33564/ijeast.2020.v04i12.054.
6. Bogdanov, P.Yu., Kraeva, E.V., Verevkin, S.A., Poymanova, E.D., Tatarnikova, T.M. (2021) 'Software environments for studying the basics of neural networks', *Software & Systems*, 34(1), pp. 145–150 (in Russ.). doi: 10.15827/0236-235X.133.145-150.
7. Tariq, R. (2016) *Make Your Own Neural Network*. CreateSpace Publ., 223 p. (Russ. ed.: (2017) Moscow, 272 p.).
8. Macpherson, T., Matsumoto, M., Gomi, H., Morimoto, J., Uchibe, E., Hikida, T. (2021) 'Parallel and hierarchical neural mechanisms for adaptive and predictive behavioral control', *Neural Networks*, 144, pp. 507–521. doi: 10.1016/j.neunet.2021.09.009.
9. Yu, S., Jiang, F., Li, L., Xie, Y. (2017) 'CNN-GRNN for image sharpness assessment', in *LNIP. Proc. CNN*, 10116, pp. 50–61. doi: 10.1007/978-3-319-54407-6_4.
10. Krutikov, A.K. (2018) 'Forecasting sports results in individual sports using a generalized regression neural network', *Young Scientist*, (12), pp. 22–26 (in Russ.).

Авторы

Татарникова Татьяна Михайловна¹, д.т.н.,
профессор, директор института,
tm-tatarn@yandex.ru
Боженко Виктория Вячеславовна¹,
старший преподаватель, vibozhenko@yandex.ru

Authors

Tatiana M. Tatarnikova¹, Dr.Sci. (Engineering),
Professor, Director University,
tm-tatarn@yandex.ru
Viktoriya V. Bozhenko¹,
Senior Lecturer, vibozhenko@yandex.ru

Санкт-Петербургский государственный
университет аэрокосмического приборостроения,
г. Санкт-Петербург, 190000, Россия

¹ Saint Petersburg State University
of Aerospace Instrumentation,
Saint Petersburg, 190000, Russian Federation

Функциональное моделирование двух сложных конкурирующих систем для обоснования выбора рациональной структуры исследуемой системы

Н.В. Долгов ¹✉, В.А. Ильин ¹

¹ Военный институт дополнительного профессионального образования
ВУНЦ ВМФ «Военно-морская академия»,
г. Санкт-Петербург, 195112, Россия

Ссылка для цитирования

Долгов Н.В., Ильин В.А. Функциональное моделирование двух сложных конкурирующих систем для обоснования выбора рациональной структуры исследуемой системы // Программные продукты и системы. 2024. Т. 37. № 3. С. 369–373. doi: 10.15827/0236-235X.142.369-373

Информация о статье

Группа специальностей ВАК: 2.3.7

Поступила в редакцию: 03.05.2024

После доработки: 10.06.2024

Принята к публикации: 17.06.2024

Аннотация. Классическое функциональное моделирование на качественном уровне выполняется на основе анализа структуры связей между отдельными функциями моделируемого процесса без учета их взаимодействия с другими системами, в том числе взаимного воздействия отдельных функций двух и более систем. В представленной статье дается общий анализ функционального моделирования двух сложных взаимодействующих систем с антагонистическими целями. При взаимодействии этих систем одна из них целенаправленно атакует другую, а противоположная сторона защищается, пассивно или активно противодействуя этим атакам, то есть существует определенная конкуренция между системами. В качестве примера рассмотрен некий абстрактный процесс, представленный в нотации технологии IDEF0 и являющийся функциональной моделью двух сложных взаимодействующих систем с антагонистическими целями. Интерпретируя эту модель как оргграф, осуществляется формирование матрицы связей функций моделируемого процесса, на основе которой выводятся математические зависимости определения эффективности функциональной модели двух сложных взаимодействующих систем с антагонистическими целями.

Ключевые слова: функциональная модель, сложные системы с антагонистическими целями, эффективность функциональной модели, матрица связей, процесс функционирования, технология IDEF0, воздействие, атака, защита

Введение. Неизбежное усложнение систем, и в первую очередь искусственных организационно-технических, создаваемых человеком в разных областях деятельности, обусловило необходимость предвидеть результаты длительных и дорогостоящих работ по их созданию и поиску инструментария для этого [1]. Цель поиска – увидеть и оценить функционирование создаваемой системы раньше, чем она будет создана физически [2].

Общая методология функционального моделирования IDEF включает в себя ряд частных методологий. Одна из них, IDEF0, используется для создания функциональных моделей, отображающих структуру и функции системы, а также потоки информации и материальных объектов, преобразуемые этими функциями. С помощью наглядного графического языка IDEF0 изучаемая система предстает перед разработчиками и аналитиками в виде набора взаимосвязанных функций (в терминах IDEF0 – функциональных блоков).

Известно, что любая система создается для того, чтобы функционировать. Результатом работы системы является осуществление ею ма-

териально-информационных обменов, то есть преобразований ее входов в выходы [3]. Система обеспечивается материальной инфраструктурой, составляющей ее устройство и морфологию. Характер (режим) функционирования системы в пределах ее морфологических возможностей определяется управлением работой системы. Именно особенности функционирования обуславливают и морфологию, и материально-информационные потоки системы [4].

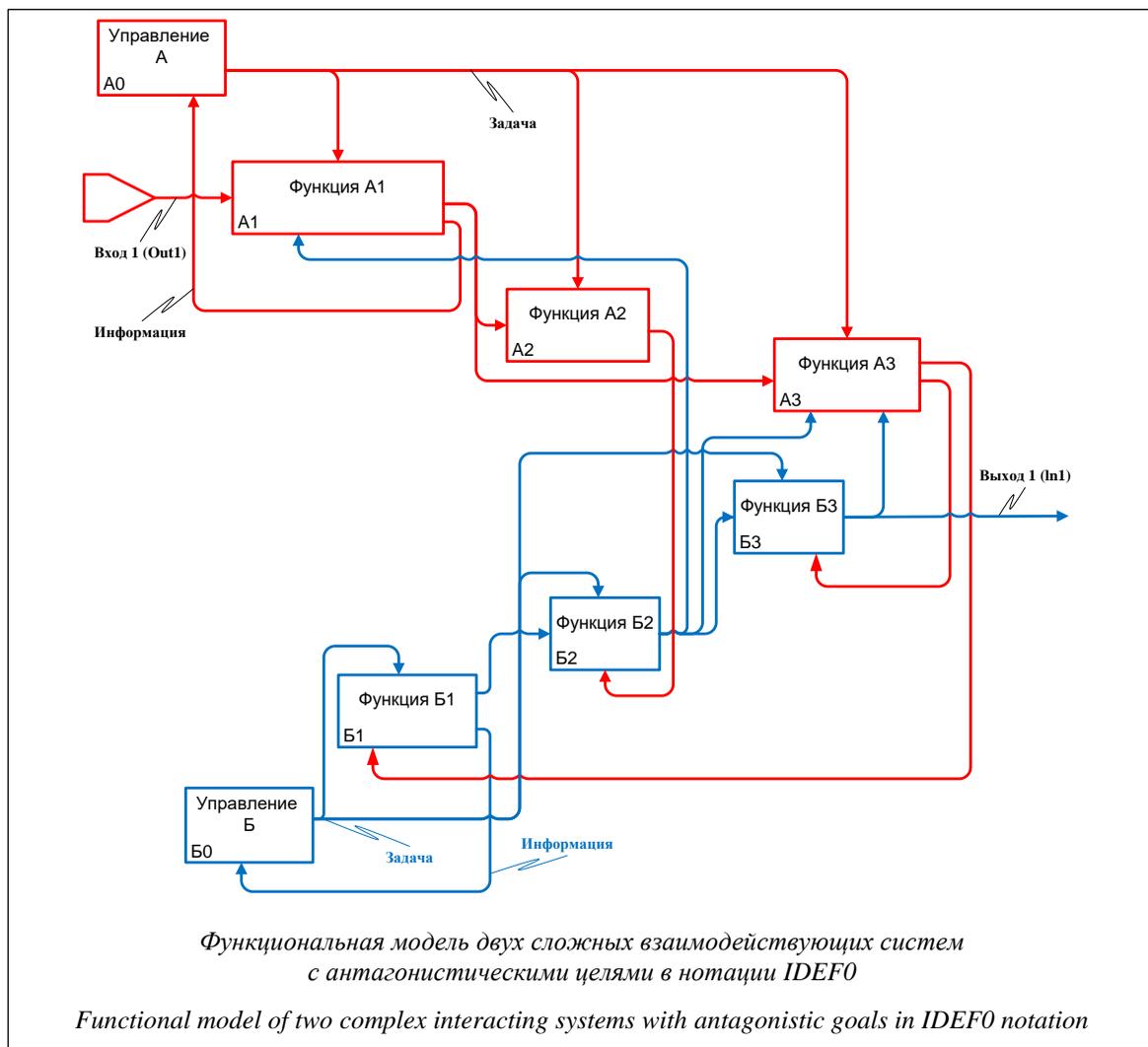
Функциональная модель есть графическое представление функционирования систем и объектов [5]. Она может использоваться как для анализа процессов, протекающих в существующих системах (их функционирование), так и при проектировании новых систем для синтеза их структур [6]. Однако анализ функциональных моделей в настоящее время выполняется на качественном уровне путем анализа структуры связей между отдельными функциями моделируемого процесса, а также сравнения различных структур процесса и различных процессов между собой. В незначительном количестве имеющихся источников и документов по методологии функционального

моделирования нет информации о методах аналитического количественного анализа функциональных моделей и процессах, описываемых ими [7–9].

Функциональная модель двух сложных взаимодействующих систем с антагонистическими целями представляет собой противостояние двух сложных, например, конкурирующих систем. В общем случае такой процесс может содержать в себе два варианта процессов: с одной стороны – воздействие, атака на систему, с другой – защита от воздействия и противодействие, а также одновременно воздействие и защита от него каждой сложной системой.

В первом случае процесс будет иметь один выход – из атакуемой системы, во втором два – из каждой системы. В статье рассматривается только первый вариант, поскольку второй является частным случаем и принципиально не отличается от первого.

Процесс действий двух сложных систем с антагонистическими целями в нотации технологии IDEF0 представлен на рисунке. Он включает в себя шесть процедурных функций (A1–A3, B1–B3), выполняющих, скажем, обработку информации, и две специфические функции управления процессом (A0, B0), связанные между собой прямыми и обратными функциональными связями. Процесс имеет один вход из внешней среды (может быть из другого процесса – Out1) и один выход во внешнюю среду (во внешние процессы – In1). Как видно, вход из внешней среды выполняется через функцию A1, обеспечивающую обработку входной информации и ее передачу другим функциям. Выход во внешнюю среду осуществляется функцией B3, которая, очевидно, решает основные задачи процесса, а остальные – обеспечивающие. Следует заметить, что количество входов и выходов в общем случае может быть любым и определяется только структурой исследуемого процесса [10].



Интерпретируя функциональную модель как орграф, сформируем матрицу связей функций моделируемого процесса, представленную в таблице. В матрице по строкам единицами и нулями записаны входы в функции, по столбцам, соответственно, – выходы из них. Суммы входов по строкам и выходов по столбцам позволяют ранжировать функции по степени важности в исследуемом процессе. Общая сумма входов и выходов позволяет сравнивать структуры одного процесса и разные процессы по их сложности [10].

Матрица связей функциональной модели двух сложных взаимодействующих систем с антагонистическими целями

Relationship matrix for a functional model of two complex interacting systems with antagonistic goals

	Ou tl	A0	A1	A2	A3	B0	B1	B2	B3	In1	Σin
Ou tl	0	0	1	0	0	0	0	0	0	0	1
A0	0	0	1	1	1	0	0	0	0	0	3
A1	0	1	0	1	1	0	0	0	0	0	3
A2	0	0	0	0	0	0	0	1	0	0	1
A3	0	0	0	0	0	0	1	0	1	0	2
B0	0	0	0	0	0	0	1	1	1	0	3
B1	0	0	0	0	0	1	0	1	0	0	2
B2	0	0	1	0	1	0	0	0	1	0	3
B3	0	0	0	0	1	0	0	0	0	0	1
In1	0	0	0	0	0	0	0	0	1	0	1
Σo ut	0	1	3	2	4	1	2	3	4	0	20

Эффективность процесса может быть определена как среднее арифметическое четырех коэффициентов и рассчитывается по формуле

$$\Theta_n = \frac{K_m + K_c + K_p + K_n}{4} \tag{1}$$

K_m – масштабный коэффициент процесса, оценивающий количество функций в нем. Чем больше коэффициент, тем сложнее процесс. Расчет осуществляется по формуле

$$K_m = 1 - e^{-\left(1 - \frac{2}{f_\Sigma}\right)} \tag{2}$$

где f_Σ – общее количество функций исследуемого процесса.

K_c – коэффициент связанности функций процесса, оценивающий коммуникативность функций в процессе. Более сложный процесс имеет больший коэффициент связанности. Расчет осуществляется по формуле

$$K_c = e^{-\left[1 - \frac{S_\Sigma}{f_\Sigma(f_\Sigma - 1)}\right]} \tag{3}$$

где S_Σ – общее число связей, то есть число единиц в матрице.

K_p – коэффициент равномерности распределения функций в процессе. Чем меньше коэффициент равномерности, тем лучше структурирован процесс. Коэффициент рассчитывается по формуле

$$K_p = e^{-\frac{\sum_j |\Sigma in - \bar{in}| + \sum_j |\Sigma out - \bar{out}|}{2S_\Sigma}} \tag{4}$$

где Σin – сумма единиц в строке, то есть сумма входных связей функции процесса; \bar{in} – среднее число входных связей; Σout – сумма единиц в столбце, то есть сумма выходных связей функции процесса; \bar{out} – сумма единиц в столбце, то есть среднее число выходных связей.

K_n – коэффициент изолированности функций, определяющий общее число несвязанных функций, а также не имеющих входов или выходов. Изолированные функции не влияют на качество процесса и рассчитываются по формуле

$$K_n = e^{-\frac{n_{стр} + m_{стол}}{2f_\Sigma}} \tag{5}$$

где $n_{стр}$ – число пустых строк (число функций без входов); $m_{стол}$ – число пустых столбцов (функций без выходов).

Числовые значения матрицы связей функциональной модели действий двух сложных взаимодействующих систем с антагонистическими целями из рассматриваемого примера подставим в формулы (1)–(5) и произведем расчет коэффициентов и эффективности процесса.

Масштабный коэффициент процесса.

Общее количество функций $f_\Sigma = 6$, тогда

$$K_m = 1 - e^{-\left(1 - \frac{2}{6}\right)} = 1 - e^{-\left(1 - \frac{2}{6}\right)} = 1 - e^{-0,666} = 0,4865.$$

Коэффициент связанности функций процесса.

Общее число связей, то есть число единиц в матрице $S_\Sigma = 20$, тогда

$$K_c = e^{-\left[1 - \frac{S_\Sigma}{f_\Sigma(f_\Sigma - 1)}\right]} = e^{-\left[1 - \frac{20}{6(6-1)}\right]} = e^{-0,333} = 0,7165.$$

Коэффициент равномерности распределения функций в процессе.

Среднее число входных связей $\bar{in} = 10$.

Среднее число выходных связей $\bar{out} = 12$.

$$K_p = e^{\frac{\sum_f |\Sigma_{in-\bar{in}}| - \sum_f |\Sigma_{out-\bar{out}}|}{2S_{\Sigma}}} =$$

$$= e^{\frac{|20-10| - |20-12|}{2 \cdot 20}} = e^{-0,05} = 0,9512.$$

Коэффициент изолированности функций.

Число пустых строк $n_{стр} = 0$, число пустых столбцов $n_{сто} = 2$, тогда

$$K_{и} = e^{\frac{n_{стр} + m_{сто}}{2f_{\Sigma}}} = e^{\frac{0+2}{2 \cdot 6}} = e^{-0,1666} = 0,8464.$$

Эффективность процесса.

$$\Theta_n = \frac{K_m + K_c + K_p + K_{и}}{4} =$$

$$= \frac{0,4865 + 0,7165 + 0,9512 + 0,8464}{4} = 0,7502.$$

Таким образом, получена оценка эффективности функциональной модели при воздействии на нее другой системы.

Заключение

В статье сделан общий анализ функционального моделирования двух сложных взаимодействующих систем с антагонистическими целями. Выполнен расчет эффективности действий функциональной модели двух этих систем.

Предлагаемый метод позволяет оценить эффективность сложной системы при преднамеренном или непреднамеренном воздействии на функции исследуемой системы другой системы, имеющей, как правило, противоположные, антагонистические, цели.

Список литературы

1. Харари Ф. Теория графов. М.: Книжный дом «ЛИБРОКОМ», 2009. 300 с.
2. Ильин В.А., Козлов И.Л. К проблеме автоматизации управления силами и средствами, функциональное моделирование // Автоматизация процессов управления. 2009. № 4. С. 53–57.
3. Дубейковский В.И. Практика функционального моделирования с AllFusion Process Modeler. М.: Диалог-МИФИ, 2004. 364 с.
4. Шпак В.Ф., Директоров Н.Ф. и др. Информационные технологии в системе управления силами ВМФ. СПб: Элмор, 2005. 827 с.
5. Дрешер Д. Основы блокчейна: вводный курс для начинающих в 25 небольших главах. М.: ДМК Пресс, 2018. 320 с.
6. Карпычев В.Ю. Функциональное моделирование (IDEF0) как метод исследования блокчейн-технологии // Тр. НГТУ им. Р.Е. Алексеева. 2018. № 4. С. 22–32.
7. Тебекин А.В. Методология функционального моделирования сложных технических систем модульного типа // Журнал технических исследований. 2021. Т. 7. № 2. С. 3–12.
8. Тебекин А.В., Тебекин П.А., Егорова А.А. Технологические трансформации XXI века как индуцирующий вектор перехода к новому качеству производства // Теоретическая экономика. 2021. № 1. С. 42–53.
9. Тебекин А.В., Петров В.С. Использование методологии моделирования IDEF при формировании структурно-параметрической модели реализации технологий обеспечения эффективного развития промышленных предприятий в условиях постиндустриальной экономики // Транспортное дело России. 2017. № 4. С. 43–49.
10. Ильин В.А., Янча С.П. Методы анализа функциональных моделей // Программные продукты и системы. 2009. Т. 22. № 4. С. 32–34.

Justifying the choice of a rational structure of the system under study: Functional modeling of two complex competing systems

Nikolay V. Dolgov ¹✉, Viktor A. Ilin ¹

¹ Military Institute of Continuing Professional Education of the VUNTS Navy "Naval Academy",
St. Petersburg, 195112, Russian Federation

For citation

Dolgov, N.V., Ilin, V.A. (2024) 'Justifying the choice of a rational structure of the system under study: Functional modeling of two complex competing systems', *Software & Systems*, 37(3), pp. 369–373 (in Russ.). doi: 10.15827/0236-235X.142.369-373

Article info

Received: 03.05.2024

After revision: 10.06.2024

Accepted: 17.06.2024

Abstract. Classical functional modelling is based on the structure analysis of links between individual functions of the modelled process. It does not take into account their interaction with other systems, including mutual influence of separate

functions of two or more systems. The paper shows a general analysis of functional modelling of two complex interacting systems with antagonistic goals. When these systems interact, one of them purposefully attacks the other and the opposite side defends itself. It passively or actively counteracts these attacks, i.e. there is a certain competition between the systems. As an example, the paper considers a certain abstract process presented in the IDEF0 technology notation, which is a functional model of two complex inter-acting systems with antagonistic goals. Interpreting this model as a digraph forms a matrix of relations of the modelled process functions. This matrix is a basis for deriving mathematical dependencies for determining the effectiveness of the functional model of two complex interacting systems with antagonistic goals.

Keywords: functional model, complex systems with antagonistic goals, efficiency of a functional model, relationship matrix, functioning process, IDEF0 technology, impact, attack, defense

References

1. Harari, F. (1969) *Graph Theory*. Addison-Wesley Publ., 274 p. (Russ. ed.: (2009) Moscow, 300 p.).
2. Ilin, V.A., Kozlov, I.L. (2009) 'On the problem of automation of control of forces and means, functional modeling', *Automation of Control Processes*, (4), pp. 53–57 (in Russ.).
3. Dubeykovskiy, V.I. (2004) *Practice Functional Modeling with AllFusion Process Modeler*. Moscow, 364 p. (in Russ.).
4. Shpak, V.F., Direktorov, N.F. (2005) *Information Technologies in the Navy Force Management System*. St. Petersburg, 827 p. (in Russ.).
5. Drescher, D. (2017) *Blockchain Basics: A Non-technical Introduction in 25 Steps*. 255 p. (Russ. ed.: (2018) Moscow, 320 p.).
6. Karpuychev, V.Y. (2018) 'Functional modeling (IDEF0) as a method of research on blockchain technology', *Transactions of NNSTU n.a. R.E. Alekseev*, (4), pp. 22–32 (in Russ.).
7. Tebekin, A.V. (2021) 'Methodology for functional modeling of complex technical systems of modular type', *J. of Tech. Research*, 7(2), pp. 3–12 (in Russ.).
8. Tebekin, A.V., Tebekin, P.A., Egorova, A.A. (2021) 'Technological transformations of the 21st century as an inducing vector of transition to a new quality of production', *Theoretical Economics*, (1), pp. 42–53 (in Russ.).
9. Tebekin, A.V., Petrov, V.S. (2017) 'Using the modeling methodology IDEF in the formation of structural-parametric model for the implementation of technology ensure the efficient development of industrial enterprises in the condition of postindustrial economy', *Transport Business of Russia*, (4), pp. 43–49 (in Russ.).
10. Ilin, V.A., Yancha, S.P. (2009) 'Methods analysis of functional models', *Software & Systems*, 22(4), pp. 32–34 (in Russ.).

Авторы

Долгов Николай Валерьевич ¹,
адъюнкт, d_n_v92@mail.ru

Ильин Виктор Алексеевич ¹, д.в.н.,
профессор, профессор кафедры,
vikil.45@yandex.ru

Authors

Nikolay V. Dolgov ¹, Adjunct,
d_n_v92@mail.ru

Viktor A. Ilin ¹, Dr.Sci. (Military Sciences),
Professor, Professor of the Department,
vikil.45@yandex.ru

¹ Военный институт дополнительного профессионального образования ВУНЦ ВМФ «Военно-морская академия», г. Санкт-Петербург, 195112, Россия

¹ Military Institute of Continuing Professional Education of the VUNTS Navy "Naval Academy", St. Petersburg, 195112, Russian Federation

Реберная раскраска кубического графа в задаче распараллеливания расчетов на неструктурированной поверхностной расчетной сетке

А.А. Гуличева¹, А.А. Рыбаков^{2,3}✉

¹ МИРЭА – Российский технологический университет, г. Москва, 119454, Россия

² Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

³ МСЦ РАН – филиал ФГУ ФНЦ НИИСИ РАН, г. Москва, 119334, Россия

Ссылка для цитирования

Гуличева А.А., Рыбаков А.А. Реберная раскраска кубического графа в задаче распараллеливания расчетов на неструктурированной поверхностной расчетной сетке // Программные продукты и системы. 2024. Т. 37. № 3. С. 374–383. doi: 10.15827/0236-235X.142.374-383

Информация о статье

Группа специальностей ВАК: 2.3.5

Поступила в редакцию: 06.05.2024

После доработки: 10.06.2024

Принята к публикации: 14.06.2024

Аннотация. В статье рассмотрен вопрос повышения производительности работы конечно-объемных численных методов на вычислительных системах с общей памятью. В этих методах на этапе расчета перетекания потоков консервативных величин через границы расчетных ячеек возможны конфликты по данным, что приводит к снижению производительности, особенно при большом количестве одновременно работающих потоков. Для устранения конфликтов по данным предлагается решение, основанное на разбиении множества обрабатываемых границ ячеек расчетной сетки на подмножества без конфликтов и на обработке этих множеств по отдельности. Решение рассматривается на примере расчетов на поверхностных неструктурированных расчетных сетках, для которых поставленная задача сводится к задаче построения реберной раскраски кубического графа. Для построения реберной раскраски кубического графа применены два алгоритма: тривиальный линейный алгоритм раскраски в пять цветов и алгоритм построения раскраски Тейта в три цвета. Проводится сравнение двух алгоритмов раскраски, а также замеряется влияние предложенного подхода на эффективность распараллеливания расчетов на поверхностной расчетной сетке. Использование подхода избавления от зависимостей по данным с помощью реберной раскраски кубического графа протестировано на численной задаче расчета обледенения поверхности тела. Запуски выполнялись на микропроцессоре Intel Xeon Phi с большим количеством параллельных потоков. Результаты показали, что при возрастании количества потоков до 144 и более эффективность распараллеливания при использовании реберной раскраски вдвое выше, чем при обычном подходе по устранению зависимостей с помощью директив OpenMP.

Ключевые слова: численные методы, распараллеливание, векторизация, расчетная сетка, дуальный граф, реберная раскраска, раскраска Тейта

Благодарности. Работа выполнена в рамках госзадания НИЦ «Курчатовский институт» по теме FNEF 2024 0016

Введение. При изучении природных явлений и проектировании сложных технических систем используется компьютерное моделирование физических процессов, описываемых в виде систем дифференциальных уравнений в частных производных. Одним из наиболее распространенных видов численных методов, которые применяются в компьютерном моделировании для решения таких систем уравнений, является метод конечных объемов. Конечно-объемные численные методы активно используются при численном решении задач газовой динамики, теории мелкой воды и многих других [1–3]. При этом постоянно ведутся исследования, направленные на максимальное использование возможностей массивно-параллельных вычислительных систем для решения таких задач [4–6]. Отдельно следует отметить исследования, направленные на максимальное использование возможностей векторизации решателей различного вида [7]. Векторизация

вычислений позволяеткратноувеличиватьпроизводительностьпрограммногокодапутемобъединенияоднотипныхскалярныхоперацийввекторныеинструкции.СовременнымикропроцессорыIntelподдерживаютнаборвекторныхинструкцийAVX-512,спомощьюкоторыхможновекторизоватьнетолькобезусловныеоперации,ноипрограммныйкод,содержащийразветвленноеуправление,командыпереходовивызовыфункций.Дляуспешнойвекторизациипрограммногокодаважноизбавитьсяотзависимостеймеждуотдельнымивекторизуемымиэлементами(например,привекторизациикцикласледуетминимизироватьзависимостимеждуегоитерациями).Вработе[8]рассматриваетсявекторизациятрехмерногогазодинамического решателя, и можно заметить, что наименьшая эффективность векторизации наблюдается для функции пересчета потоков консервативных величин между расчетными ячейками.

Данная работа посвящена задаче, решаемой на неструктурированной поверхностной расчетной сетке, – моделированию обледенения поверхности обтекаемого тела [9].

Моделирование процесса обледенения поверхности тела выполняется итерационно. На каждой итерации расчетов в каждой расчетной ячейке решается система уравнений массового и теплового балансов, из которой получаются основные данные состояния расчетной ячейки (температура, количество жидкости и накопленного льда). Между расчетными итерациями выполняется моделирование протекания потоков жидкости в соседних ячейках (массы и тепла через границы ячеек) [10–12]. Также в зависимости от настроек решателя через заданные промежутки времени выполняется пересчет геометрии поверхности тела за счет накопленной в ее ячейках массы льда. Будем считать, что расчеты обледенения выполняются на неструктурированной поверхностной расчетной сетке с треугольными ячейками. Пересчет состояния ячейки происходит независимо от других ячеек, данные вычисления могут производиться параллельно. Пересчет потоков массы и тепла осуществляется для каждого ребра домена расчетной сетки, при этом поток перетекает из одной инцидентной ребру ячейки в другую.

Общая схема пересчета потоков для одного ребра может выглядеть так:

```
void
Edge::calc_flows ()
{
    // Получение пары ячеек.
    Cell* fst = ...;
    Cell* sec = ...;

    // Вычисление потоков.
    double flow_w = ...;
    double flow_q = ...;

    // Корректировка потоков.
    fst->flow_out_w += flow_w;
    fst->flow_out_q += flow_q;
    sec->flow_out_w -= flow_w;
    sec->flow_out_q -= flow_q;
}
```

Для выполнения пересчета потоков для всех ребер домена расчетной сетки необходимо обработать все ребра в цикле:

```
for (auto e : own_edges)
{
    e->calc_flows ();
}
```

При обработке ребер домена расчетной сетки в цикле возникает желание распараллели-

лить или векторизовать данный цикл. При распараллеливании вычислений с помощью OpenMP следует учитывать конфликты по данным, которые могут возникнуть при корректировке потоков (если несколько потоков начнут одновременно изменять значение одной области памяти). Для устранения этих конфликтов достаточно выполнять операции корректировки потоков в атомарном режиме:

```
#pragma omp atomic
fst->flow_out_w += flow_w;
#pragma omp atomic
fst->flow_out_q += flow_q;
#pragma omp atomic
sec->flow_out_w -= flow_w;
#pragma omp atomic
sec->flow_out_q -= flow_q;
```

Использование `#pragma omp atomic` гарантирует, что указанная команда одновременно будет обрабатываться только одним потоком (то есть между чтением старого значения переменной и записью нового значения не вклинется другой поток). При большом количестве используемых потоков это может приводить к потерям производительности. Отдельно следует отметить, что такой подход не сработает при попытке векторизации рассматриваемого цикла. Зачастую вычисление самих потоков через ребро не содержит сложного управления и может быть легко векторизовано, но векторизации всего цикла мешают зависимости по данным при корректировке потоков. Возникает желание разбить исходное множество ребер сетки на такие подмножества, чтобы параллельная обработка ребер каждого отдельного подмножества не приводила к возникновению конфликтов при корректировке потоков.

Сведение задачи распараллеливания вычислений к реберной раскраске

Будем решать задачу разрешения конфликтов между ребрами расчетной сетки с помощью реберной раскраски графа конфликтов. Такой подход является достаточно естественным [13], однако в рассматриваемом случае вершины графа конфликтов будут иметь небольшую степень, что свидетельствует о допустимости реберной раскраски в небольшое количество цветов.

Рассмотрим ситуацию, при которой возможно возникновение конфликта при корректировке потоков во время параллельной обработки двух ребер расчетной сетки. Такой конфликт возможен в том случае, когда оба обрабатыва-

емых ребра являются инцидентными одной и той же ячейке. Рассмотрим дуальный граф расчетной сетки – то есть граф, вершины которого соответствуют ячейкам расчетной сетки, а ребро проведено между двумя вершинами только в том случае, когда две соответствующие ячейки расчетной сетки являются соседними по ребру. Без ограничения общности будем считать, что рассматриваемая расчетная сетка не имеет края, то есть каждая ее ячейка имеет ровно трех соседей. В этом случае ее дуальный граф будет кубическим. При этом задача разбиения ребер расчетной сетки на неконфликтующие множества сводится к построению реберной раскраски дуального графа. В процессе нахождения реберной раскраски построенного дуального графа возникает вопрос о том, в какое минимальное количество цветов можно раскрасить этот граф.

Сразу следует отметить, что ребра произвольного кубического графа можно тривиальным способом раскрасить в пять цветов, поскольку у любого ребра есть ровно четыре смежных ребра, которые должны отличаться от него по цвету, поэтому для него всегда можно выбрать пятый цвет. Данный алгоритм является линейным и интереса не представляет. С другой стороны, очевидно, что двумя цветами обойтись не удастся, так как в каждой вершине сходятся по три разноцветных ребра. Если кубический граф допускает правильную реберную раскраску в три цвета, то такая раскраска называется раскраской Тейта. Следует заметить, что не все кубические графы допускают раскраску Тейта. В работе [14] предлагается способ построения раскраски Тейта для плоских кубических графов. Авторы данной работы будут использовать его в качестве основы. Кубический граф, порожденный поверхностной расчетной сеткой, не всегда является плоским, но попытка поиска для него раскраски Тейта может быть оправданна. К тому же для целей практического использования не будет большой проблемой отсутствие раскраски Тейта, в этом случае всегда можно воспользоваться тривиальной раскраской в пять цветов.

Редуцирование кубического графа по ребру и его восстановление

Рассмотрим принцип и реализацию алгоритма построения раскраски Тейта для кубического графа, являющегося дуальным графом для замкнутой поверхностной неструктуриро-

ванной расчетной сетки. В работах [14, 15] рассматривается подход к построению раскраски Тейта путем удаления ребер из исходного плоского кубического графа. Ребра удаляются до тех пор, пока не будет получен кубический граф с уже известной раскраской. После этого ребра возвращаются в граф в обратном порядке с соответствующей коррекцией раскраски. Операцию удаления ребра из графа будем называть редуцированием кубического графа по ребру. Чтобы редуцирование кубического графа по ребру можно было использовать для построения раскраски Тейта, необходимо уметь провести последовательность редукиций до достижения простого по структуре кубического графа, раскраска которого не представляет сложности. При этом в получающихся графах допустимо наличие параллельных ребер, однако запрещено появление петель, так как раскраска Тейта для кубического графа с петлями невозможна. Опишем операцию редуцирования графа более подробно.

Сначала рассмотрим выполнение редуцирования кубического графа по ребру e с концами v_1 и v_2 , где инцидентными ребрами вершины v_1 являются ребра $e, e_1(v_1), e_2(v_1)$, а инцидентными ребрами вершины v_2 являются ребра $e, e_1(v_2), e_2(v_2)$, а также среди ребер $e, e_1(v_1), e_2(v_1), e_1(v_2), e_2(v_2)$ нет параллельных (рис. 1, слева). Из этого следует, что существует только одно ребро, соединяющее вершины v_1 и v_2 . В этом случае будем говорить, что редуцирование выполняется по уникальному ребру e .

При выполнении редуцирования по ребру e само ребро e удаляется, также удаляются вершины v_1 и v_2 , ребра $e_1(v_1)$ и $e_2(v_1)$ склеиваются в результирующее ребро re_1 , ребра $e_1(v_2)$ и $e_2(v_2)$ склеиваются в ребро re_2 (рис. 1, справа).

Другим случаем редуцирования является вариант, при котором концами рассматриваемого ребра e также являются вершины v_1 и v_2 , однако между ними проходит еще одно ребро, без ограничения общности будем считать, что это ребро $e_2(v_1) = e_2(v_2)$ (рис. 2, слева). В этом случае при редуцировании удаляются ребро e ,

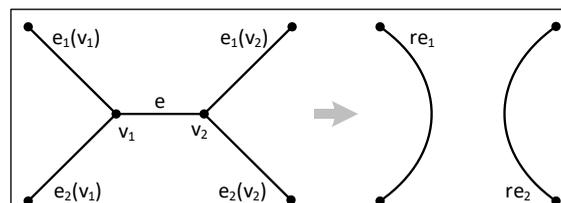
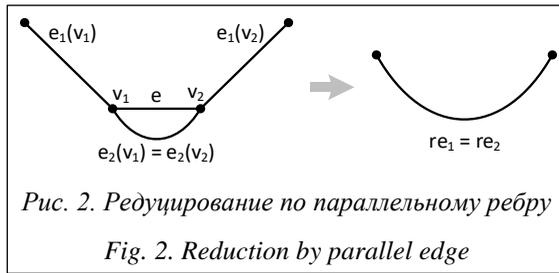


Рис. 1. Редуцирование по уникальному ребру

Fig. 1. Reduction by unique edge

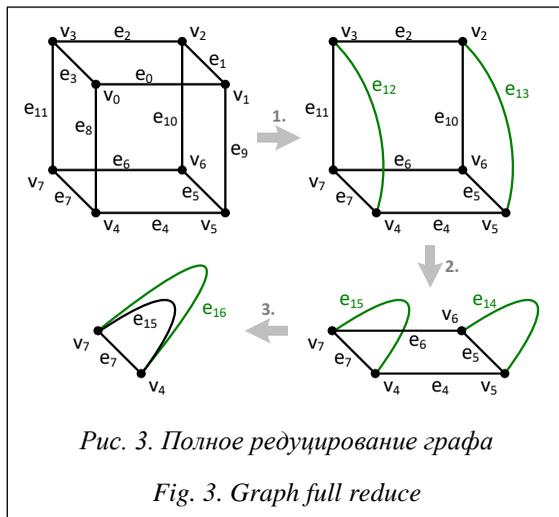


вершины v_1 и v_2 , а все три ребра $e_1(v_1)$, $e_2(v_1) = e_2(v_2)$, $e_1(v_2)$ склеиваются в единое ребро $re_1 = re_2$ (рис. 2, справа). Такой шаг редуцирования будем называть редуцированием по параллельному ребру.

Отдельно отметим случай, когда между двумя вершинами проходят три параллельных ребра. Если мы достигли такого графа, то это и есть минимальный кубический граф, построение раскраски для которого очевидно и от которого нужно двигаться в обратную сторону, постепенно восстанавливая исходный граф. Также следует рассмотреть случай, при котором после редуцирования по уникальному ребру граф перестает быть связным. Это значит, что в исходном графе ребро e было мостом, такие графы рассматривать не будем. В других случаях в графе найдется либо уникальное, либо параллельное ребро, по которому можно осуществить следующий шаг редуцирования.

В качестве примера рассмотрим редуцирование графа, представляющего собой куб, как показано на рисунке 3.

Кубический граф содержит 8 вершин v_0-v_7 и 12 ребер e_0-e_{11} . Будем считать, что нижние индексы в именах вершин и ребер являются также их идентификаторами. Для полного редуцирования указанного графа требуется вы-



полнить три шага, которые могут быть записаны в историю редуцирования следующим образом:

```

e0 [(v0 : e3, e8 -> re12),
(v1 : e1, e9 -> re13)]
e2 [(v2 : e13, e10 -> re14),
(v3 : e12, e11 -> re15)]
e5 [(v5 : e4, e14 -> re16),
(v6 : e6, e14 -> re16)]
    
```

По такой записи истории редуцирования можно идентифицировать каждый шаг, определить его тип (редуцирование по уникальному ребру или по параллельному) и восстановить граф.

Опираясь на указанные операции редуцирования, можно описать алгоритм построения раскраски Тейта во время восстановления графа.

Алгоритм реберной раскраски кубического графа в три цвета

Прежде чем перейти к описанию алгоритма построения раскраски, рассмотрим центральный объект, который будет использован в данном построении. Без ограничения общности будем считать, что выполняется раскраска в красный, синий и зеленый цвета. Эти же цвета используются и на иллюстрациях. Пусть ребра некоторого кубического графа правильным образом раскрашены в три цвета. Возьмем два произвольных цвета, например, красный и синий. Если рассмотреть все покрашенные в них ребра, а также все инцидентные им вершины, то получим граф порядка 2, у которого в каждой вершине сходятся разноцветные ребра. Очевидно, что такой граф является объединением простых циклов четной длины (двухцветных циклов). На рисунке 4 приведен пример такого двухцветного красно-синего цикла.

С таким двухцветным циклом можно выполнять следующие операции. Во-первых, для всех ребер двухцветного цикла можно заменить цвет на противоположный, после чего раскраска в исходном графе останется правильной (рис. 4). Также можно применить перекраску в противоположный цвет не всех ребер цикла, а только расположенных между двумя фиксированными ребрами e_1 и e_2 (на рисунке 5 приведены два варианта такой перекраски в порядке обхода цикла по часовой стрелке и против). При такой перекраске ребер в цикле возникают два конфликта по цветам: между ребрами e_1 и e_2 и их перекрашенными соседями.

Смысл частичной замены цветов в двухцветном цикле становится понятным, если возникает необходимость поместить в граф новое ребро. На рисунке 6 продемонстрирована операция, при которой выполняется частичная перекраска цикла между ребрами e_1 и e_2 , а затем на ребра e_1 и e_2 добавляются новые вершины v_1 и v_2 соответственно, между которыми проводится ребро. Полученные после разбиения ребер e_1 и e_2 более мелкие ребра перекрашиваются для устранения конфликтов. Новое проведенное ребро при этом перекрашивается в третий, свободный, цвет, что приводит к сохранению правильной раскраски во всем графе. Таким образом, выбрав два произвольных ребра на любом двухцветном цикле, можно добавить новое ребро с концами на выбранных ребрах, а затем перекрасить ребра и сохранить правильную реберную раскраску. Такую операцию будем называть восстановлением ребра по двухцветному циклу.

Также понадобится операция поиска двухцветного цикла, начиная с произвольного ребра e , содержащего ребра цветов $color(e)$ и $color_2 \neq color(e)$. Такой цикл всегда существует, и он ровно один.

Теперь, имея в своем распоряжении три простые операции: поиск двухцветного цикла по ребру и второму цвету, перекраска двухцветного цикла и восстановление ребра по

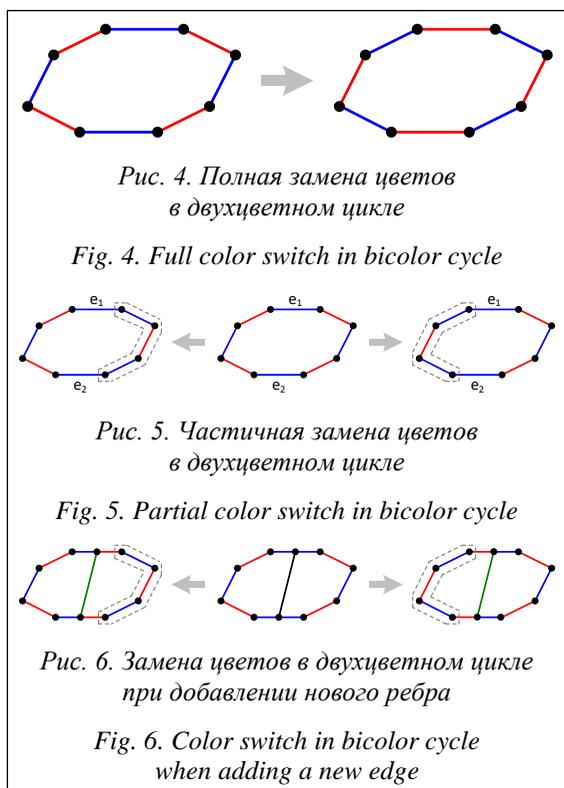
двухцветному циклу, – можно описать алгоритм восстановления одного шага редуцирования графа с сохранением правильной реберной раскраски (рис. 7).

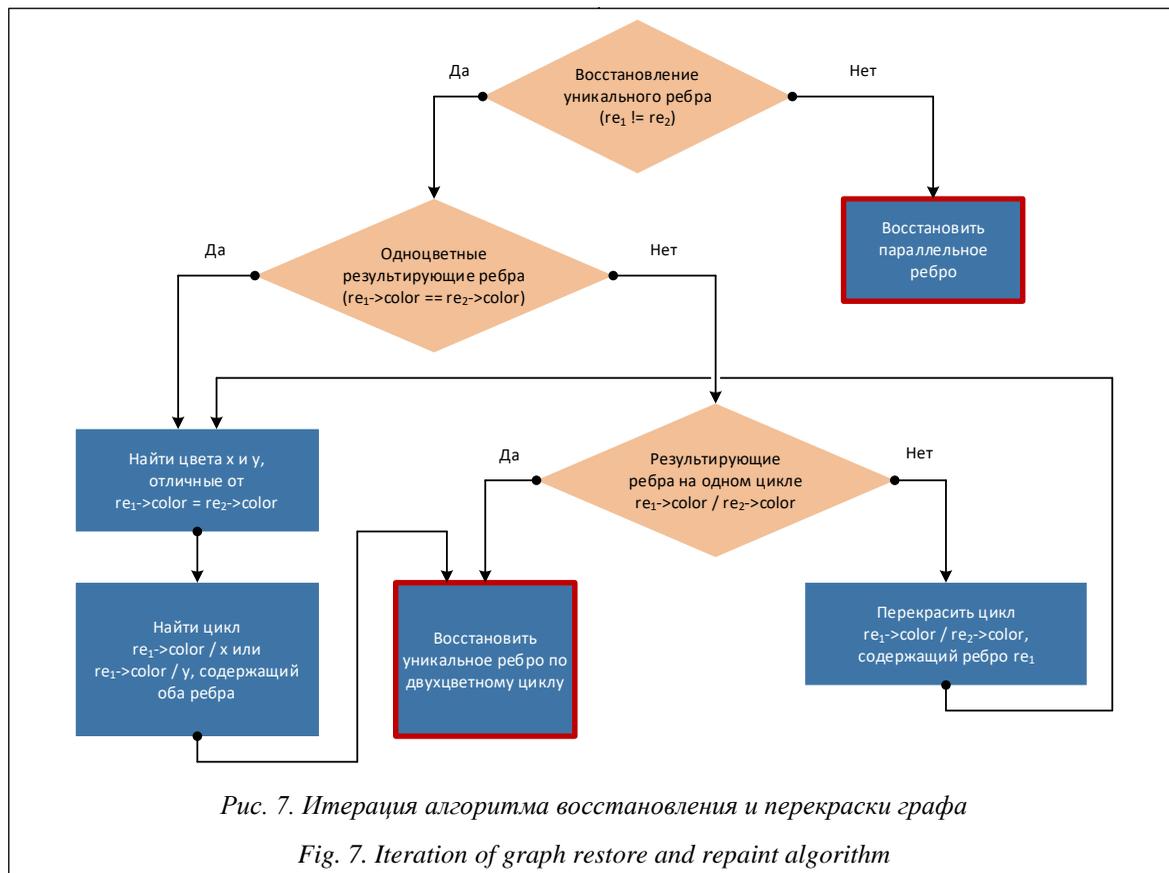
Дополнительно следует отметить, что восстановление шага редуцирования графа по параллельному ребру не представляет сложности, так как это локальная операция, которая затрагивает единственное результирующее ребро. Поэтому рассмотрим подробно только восстановление шага редуцирования по уникальному ребру.

Пусть имеется шаг редуцирования графа, результирующими ребрами после выполнения которого являются различные ребра re_1 и re_2 . В процессе восстановления нужно поместить на эти ребра новые вершины v_1 и v_2 соответственно, провести между ними ребро и выполнить коррекцию раскраски. Опишем последовательность действий в данном случае.

Вариант 1. Если результирующие ребра re_1 и re_2 имеют разные цвета, следует найти двухцветный цикл, начиная с ребра re_1 , включающий в себя ребра цветов $color(re_1)$ и $color(re_2)$. Если найденный цикл содержит также и ребро re_2 , то можно восстановить исходное ребро по найденному двухцветному циклу. В противном случае найденный двухцветный цикл перекрашивается. Поскольку перекрашивание затронет только ребро re_1 , после этой операции получится ситуация, в которой ребра re_1 и re_2 имеют один цвет. В этом случае осуществляется переход ко второму варианту.

Вариант 2. Если результирующие ребра re_1 и re_2 имеют один и тот же цвет, то рассмотрим два оставшихся цвета – x и y , а также два двухцветных цикла, начиная с ребра re_1 . Причем в первый цикл будем включать ребра с цветами $color(re_1)$ и x , а во второй ребра с цветами $color(re_1)$ и y . Если один из найденных двухцветных циклов будет содержать также и ребро re_2 , то по этому циклу и нужно выполнить восстановление исходного ребра, которое было удалено из графа при редуцировании. Отметим, что в общем случае нет гарантии, что один из найденных двухцветных циклов будет содержать ребро re_2 (исследования в работе [14] касаются только плоских кубических графов, а дуальный граф поверхностной сетки необязательно будет плоским), в этом случае корректировка реберной раскраски невозможна и исходный граф не допускает раскраски Тейта. Однако в ходе экспериментов по раскраске кубических дуальных графов, построенных по замкнутым поверхностным расчетным сеткам,





такая ситуация не наблюдалась и для всех таких графов раскраска Гейта была найдена.

На рисунке 8 приведена последовательность восстановления графа, редуцирование которого продемонстрировано на рисунке 3. Приведем еще раз историю редуцирования графа:

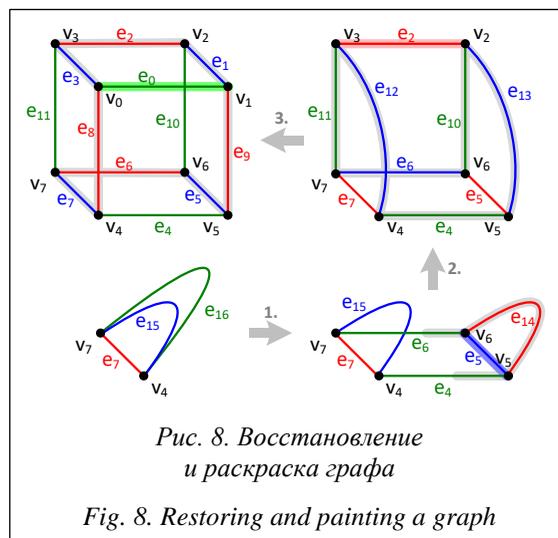
```
e0 [(v0 : e3, e8 -> re12),
(v1 : e1, e9 -> re13)]
e2 [(v2 : e13, e10 -> re14),
(v3 : e12, e11 -> re15)]
e5 [(v5 : e4, e14 -> re16),
(v6 : e6, e14 -> re16)]
```

Первый шаг восстановления относится к редуцированию по параллельному ребру. Ребро e_{16} разбивается на ребра e_6, e_{14}, e_4 с помощью вершин v_6 и v_5 . Между вершинами v_6 и v_5 восстанавливается ребро e_5 . Ребра e_6 и e_4 наследуют зеленый цвет ребра e_{16} , а ребра e_{14} и e_5 раскрашиваются в два оставшихся цвета.

Второй шаг восстановления относится к редуцированию по уникальному ребру. Сначала ищется красно-синий цикл, содержащий ребро e_{14} . Так как он не содержит ребро e_{15} , цикл из двух ребер $e_{14}-e_5$ перекрашивается. После этого ищется сине-зеленый цикл, содержащий ребро e_{14} . Это цикл $e_{14}-e_6-e_{15}-e_4$, и он содержит ребро e_{15} .

Происходит восстановление ребра по найденному двухцветному циклу.

Третий шаг восстановления также относится к редуцированию по уникальному ребру. Так как существует сине-красный цикл $e_{12}-e_2-e_{13}-e_5-e_6-e_7$, сразу можно выполнить восстановление ребра по этому циклу (в данном случае можно выполнить восстановление и по сине-зеленому циклу, что приведет к другой раскраске).



Реализацию данного алгоритма можно найти в репозитории <https://github.com/r-aax/caesar> (метод `graph::Graph::edges_coloring_for_cubic_graph_with_bicolor_cycles_algorithm`).

Очевидно, что описанный алгоритм имеет квадратичную сложность по порядку графа, так как количество шагов восстановления пропорционально порядку исходного графа, а на каждом шаге восстановления уникального ребра необходимо выполнять поиск и перекраску двухцветных циклов, что в худшем случае пропорционально порядку текущего графа.

Анализ полученных результатов

Если рассматривать кубический граф порядка n , то он содержит $3n/2$ ребер, причем n четно. Если данный граф допускает правильную реберную раскраску в три цвета, то в каждый из цветов раскрашено ровно $n/2$ ребер. Таким образом, множество ребер распадается на три одинаковых по размеру подмножества ребер без конфликтов. Вызывает интерес вопрос о процентном соотношении ребер, раскрашенных в разные цвета, для жадной раскраски в пять цветов при порядке графа, стремящемся в бесконечность.

Для экспериментального получения такого распределения будем строить искусственные кубические графы следующим образом. В качестве нулевого графа возьмем кубический граф K_4 , далее последовательно в текущем кубическом графе будем случайным образом выбирать вершину и заменять ее на треугольную конструкцию, как показано на рисунке 9. Вполне очевидно, что получающиеся таким образом графы будут кубическими и допускают правильную реберную раскраску в три цвета.

К полученным сгенерированным описанным способом кубическим графам порядка более 10^5 был применен жадный линейный алгоритм реберной раскраски в пять цветов. Получившееся распределение цветов показано на рисунке 10.

На рисунке видно, что распределение цветов жадной раскраски устроено примерно следующим образом. Большинство ребер кубического графа практически равномерно раскрасились в три цвета, однако для окрашивания около 10 % ребер потребовалось использование четвертого цвета, а для окрашивания около 1,5 % ребер пришлось задействовать пятый цвет.

Для оценки целесообразности применения реберной раскраски для устранения конфлик-

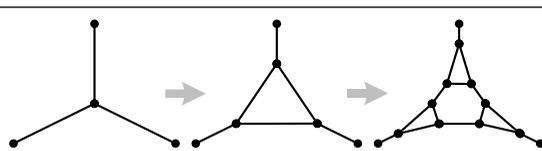


Рис. 9. Генерация кубического графа

Fig. 9. Cubic graph generation

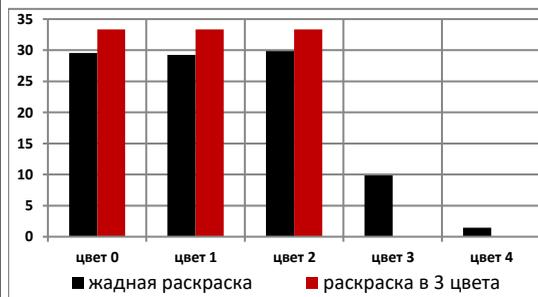


Рис. 10. Распределение процентной доли цветов для разных раскрасок

Fig. 10. Colors percentage distribution for different colorings

тов по данным при расчете потоков в задаче моделирования процесса обледенения поверхности были выполнены запуски модельной задачи на одном вычислительном узле на базе микропроцессора Intel Xeon Phi Knights Landing 7 290. Данный микропроцессор имеет 72 ядра, на каждом из которых может быть запущено до четырех потоков. Таким образом, на этом микропроцессоре возможно распараллеливание запуска на 288 потоков [16].

Для сравнения эффективности методов устранения конфликтов при параллельном расчете потоков через границы ячеек замерялась эффективность распараллеливания данного кода. Она считалась следующим образом. В качестве ускорения, достигаемого при распараллеливании на i потоков, бралась величина $s(i) = t(1)/t(i)$, где $t(1)$ – эталонное время вычислений при использовании одного потока, $t(i)$ – время вычислений при использовании i потоков. Под эффективностью распараллеливания понималась величина, вычисляемая как $e(i) = s(i)/i$. Смысл данной величины заключается в следующем. Можно предположить, что при идеальном распараллеливании вычислений при увеличении количества потоков ровно в k раз время выполнения уменьшается ровно в k раз. Таким образом, в случае идеального распараллеливания $s(i) = i$, а $e(i) = 1$. Эффективность распараллеливания вычислений является удобным показателем качества исполняемого

параллельного кода и сравнения различных вычислительных систем между собой.

На рисунке 11 можно заметить, что при использовании распараллеливания на относительно небольшое количество потоков нет существенной разницы между описанными методами устранения конфликтов. Однако при возрастании количества используемых потоков метод реберной раскраски гораздо эффективнее использования директивы `omp atomic` с точки зрения эффективности распараллеливания.

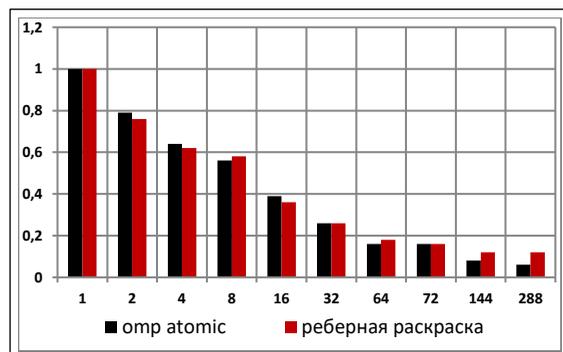


Рис. 11. Эффективность распараллеливания пересчета потоков для разных способов устранения конфликтов по данным

Fig. 11. Flows calculation parallelization efficiency for different data conflict elimination methods

Заключение

В процессе работы над оптимизацией ПО компьютерного моделирования обледенения поверхности выявлено узкое место для распараллеливания вычислений на общей памяти, связанное с возникновением конфликтов по данным. Эти конфликты приводят к необходимости использования режима атомарного исполнения отдельных команд внутри распараллеленного цикла, а также создают существенные препятствия для векторизации кода. Для устранения конфликтов рассмотрен и предложен альтернативный подход, основанный на построении реберной раскраски дуального графа расчетной сетки.

Для построения реберных раскрасок использованы два алгоритма: первый – тривиальный жадный линейный алгоритм раскраски в пять цветов, второй – квадратичный алгоритм построения раскраски Тейта, основанный на последовательном редуцировании и восстановлении графа. В ходе экспериментов выявлено, что при использовании большого количества потоков при распараллеливании использование реберной раскраски является более эффективным методом устранения конфликтов. Также следует отметить, что конкретный алгоритм раскраски графа не оказывает влияния на эффективность расчетов, то есть на практике может быть применен простой алгоритм раскраски в пять цветов.

Список литературы

1. Guzmán F. Finite volume methods. In: Numerical methods for initial value problems in physics, Springer Publ., 2023, pp. 205–258. doi: 10.1007/978-3-031-33556-3_5.
2. Tey W.Y., Asako Y., Wong K.Y. Comparison study between Galerkin finite element method and finite volume method for diffusion problem. J. of Advanced Research in Numerical Heat Transfer, 2024, vol. 15, no. 1, pp. 24–42. doi: 10.37934/arnht.15.1.2442.
3. Sun J.-K., Lin Ch.-D., Su X.-L., Tan Z.-Ch., Chen Y.-L., Ming P.-J. Solution of the discrete Boltzmann equation: Based on the finite volume method. Acta Phys. Sin., 2024, vol. 73, no. 11, art. 110504. doi: 10.7498/aps.73.20231984.
4. Río-Martín L., Busto S., Dumbster M. Massively parallel hybrid finite volume/finite element scheme for computational fluid dynamics. Math., 2021, vol. 9, no. 18, art. 2316. doi: 10.3390/math9182316.
5. Afzal A., Saleel C.A., Prashantha K., Bhattacharyya S., Sadhikh M. Parallel finite volume method-based fluid flow computations using OpenMP and CUDA applying different schemes. J. of Thermal Analysis and Calorimetry, 2021, vol. 145, pp. 1891–1909. doi: 10.1007/s10973-021-10637-1.
6. Tsoutsanis P., Antoniadis A.F., Jenkins K.W. Improvement of the computational performance of a parallel unstructured WENO finite volume CFD code for Implicit Large Eddy Simulation. Computers & Fluids, 2018, vol. 173, pp. 157–170. doi: 10.1016/j.compfluid.2018.03.012.
7. Брыков Н., Волков К., Емельянов В. Использование векторизованных структур данных при реализации вычислительных алгоритмов решения задач механики сплошной среды // Науч.-технич. вестн. информационных технологий, механики и оптики. 2022. Т. 22. № 1. С. 193–205. doi: 10.17586/2226-1494-2022-22-1-193-205.
8. Рыбаков А.А., Мещеряков А.О. Векторизация трехмерного метода погруженных границ для повышения эффективности расчетов на микропроцессорах Intel // Программные продукты и системы. 2023. Т. 36. № 1. С. 130–143. doi: 10.15827/0236-235X.141.130-143.
9. Mandel S. Predicting the formation of ice on aircraft. Scilight, 2019, vol. 2019, no. 52, art. 521108. doi: 10.1063/10.0000503.
10. Martini F., Contreras L.T., Ilinca A. Review of wind turbine icing modelling approaches. Energies, 2021, vol. 14, no. 16, art. 5207. doi: 10.3390/en14165207.

11. Ignatowicz K., Morency F., Beaugendre H. Numerical simulation of ice accretion using Messinger-based approach: Effects of surface roughness. Proc. CASI Aero Conf., 2019, art. hal-02409011.
12. Ignatowicz K., Morency F., Beaugendre H. Sensitivity study of ice accretion simulation to roughness thermal correction model. Aerospace, 2021, vol. 8, no. 3, art. 84. doi: 10.3390/aerospace8030084.
13. Гильфанов Л.Л., Мигранов С.В., Бикбов А.А. Распараллеливание решения задач с использованием раскраски графов // Молодой ученый. 2021. № 5. С. 4–6.
14. Курапов С.В., Давидовский М.В., Толоч А.В. Визуальный алгоритм раскраски плоских графов // Научная визуализация. 2018. Т. 10. № 3. С. 1–33. doi: 10.26583/sv.10.3.01.
15. Курапов С.В., Давидовский М.В., Толоч А.В. Алгебраические методы раскраски кубических графов // Научная визуализация. 2020. Т. 2020. № 2. С. 21–36. doi: 10.26583/sv.12.2.03.
16. Gulicheva A. Parallelization of finite-volume numerical methods of computational fluid dynamics by means of shared memory computing systems. In: CCIS. Proc. HPCST, 2024, vol. 1986, pp. 105–116. doi: 10.1007/978-3-031-51057-1_8.

Software & Systems

doi: 10.15827/0236-235X.142.374-383

2024, 37(3), pp. 374–383

Cubic graph edge coloring in the problem of calculation parallelization on an unstructured surface computational grid

Anastasiya A. Gulicheva¹, Alexey A. Rybakov^{2,3}✉¹ MIREA – Russian Technological University, Moscow, 119454, Russian Federation² Udmurt State University, Izhevsk, 426034, Russian Federation, Russian Federation³ JSCC RAS – branch of SRISA RAS, Moscow, 119334, Russian Federation

For citation

Gulicheva, A.A., Rybakov, A.A. (2024) ‘Cubic graph edge coloring in the problem of calculation parallelization on an unstructured surface computational grid’, *Software & Systems*, 37(3), pp. 374–383 (in Russ.). doi: 10.15827/0236-235X.142.374-383

Article info

Received: 06.05.2024

After revision: 10.06.2024

Accepted: 14.06.2024

Abstract. The paper considers the issue of improving the performance of finite-volume numerical methods on shared-memory computing systems. In these methods, data conflicts are possible at the stage of calculation of conservative value flows across computational cell boundaries. This leads to performance degradation, especially with a large number of simultaneously operating flows. To eliminate data conflicts, the paper proposes a solution based on partitioning the set of processed cell boundaries of the computational grid into subsets without conflicts and processing these sets separately. The authors consider the solution on the example of calculations on surface unstructured computational grids. For them the problem is reduced to the problem of constructing a cubic graph edge coloring. To construct a cubic graph edge coloring, the authors apply two algorithms: the trivial linear algorithm of coloring in five colors and the algorithm of Tait coloring in three colors. The authors of the paper compare two coloring algorithms, as well as measure the influence of the proposed approach on the efficiency of calculation parallelization on a surface computational grid. The use of the data dependency elimination approach using cubic graph edge coloring was tested on the numerical problem of calculating body surface icing. Specialists performed launches on an Intel Xeon Phi microprocessor with a large number of parallel flows. The results showed that when the number of flows increases up to 144 and more, the efficiency of parallelization using edge coloring is twice as high as that of the usual approach of dependency elimination using OpenMP directives.

Keywords: numerical methods, parallelization, vectorization, computational grid, dual graph, edges coloring, Tait coloring

Acknowledgements. The work was performed within the framework of the SIC «Kurchatov Institute» state assignment on the topic FNEF 2024 0016

References

1. Guzmán, F. (2023) ‘Finite volume methods’, in *Numerical Methods for Initial Value Problems in Physics*, Springer Publ., pp. 205–258. doi: 10.1007/978-3-031-33556-3_5.
2. Tey, W.Y., Asako, Y., Wong, K.Y. (2024) ‘Comparison study between Galerkin finite element method and finite volume method for diffusion problem’, *J. of Advanced Research in Numerical Heat Transfer*, 15(1), pp. 24–42. doi: 10.37934/arnht.15.1.2442.
3. Sun, J.-K., Lin, Ch.-D., Su, X.-L., Tan, Z.-Ch., Chen, Y.-L., Ming, P.-J. (2024) ‘Solution of the discrete Boltzmann equation: Based on the finite volume method’, *Acta Phys. Sin.*, 73(11), art. 110504. doi: 10.7498/aps.73.20231984.

4. Río-Martín, L., Busto, S., Dumbster, M. (2021) ‘Massively parallel hybrid finite volume/finite element scheme for computational fluid dynamics’, *Math.*, 9(18), art. 2316. doi: 10.3390/math9182316.
5. Afzal, A., Saleel, C.A., Prashantha, K., Bhattacharyya, S., Sathikh, M. (2021) ‘Parallel finite volume method-based fluid flow computations using OpenMP and CUDA applying different schemes’, *J. of Thermal Analysis and Calorimetry*, 145, pp. 1891–1909. doi: 10.1007/s10973-021-10637-1.
6. Tsoutsanis, P., Antoniadis, A.F., Jenkins, K.W. (2018) ‘Improvement of the computational performance of a parallel unstructured WENO finite volume CFD code for Implicit Large Eddy Simulation’, *Computers & Fluids*, 173, pp. 157–170. doi: 10.1016/j.compfluid.2018.03.012.
7. Brykov, N.A., Volkov, K.N., Emelyanov, V.N. (2022) ‘Vectorized numerical algorithms for the solution of continuum mechanics problems’, *Sci. Tech. J. Inf. Technol. Mech. Opt.*, 22(1), pp. 193–205 (in Russ.). doi: 10.17586/2226-1494-2022-22-1-193-205.
8. Rybakov, A.A., Meshcheryakov, A.O. (2023) ‘Vectorization of the three-dimensional immersed boundary method for improving the efficiency of calculations on Intel microprocessors’, *Software & Systems*, 36(1), pp. 130–143 (in Russ.). doi: 10.15827/0236-235X.141.130-143.
9. Mandel, S. (2019) ‘Predicting the formation of ice on aircraft’, *Scilight*, 2019(52), art. 521108. doi: 10.1063/1.0000503.
10. Martini, F., Contreras, L.T., Ilinca, A. (2021) ‘Review of wind turbine icing modelling approaches’, *Energies*, 14(16), art. 5207. doi: 10.3390/en14165207.
11. Ignatowicz, K., Morency, F., Beaugendre, H. (2019) ‘Numerical simulation of ice accretion using Messinger-based approach: Effects of surface roughness’, *Proc. CASI Aero Conf.*, art. hal-02409011.
12. Ignatowicz, K., Morency, F., Beaugendre, H. (2021) ‘Sensitivity study of ice accretion simulation to roughness thermal correction model’, *Aerospace*, 8(3), art. 84. doi: 10.3390/aerospace8030084.
13. Gilfanov, L.L., Migranov, S.V., Bikbov, A.A. (2021) ‘Parallelization of problem solving using graph coloring’, *Young Sci.*, (5), pp. 4–6 (in Russ.).
14. Kurapov, S.V., Davidovsky, M.V., Tolok, A.V. (2018) ‘Visual algorithm for coloring planar graphs’, *Sci. Visualization*, 10(3), pp. 1–33 (in Russ.). doi: 10.26583/sv.10.3.01.
15. Kurapov, S.V., Davidovsky, M.V., Tolok, A.V. (2020) ‘Algebraic methods for coloring cubic graphs’, *Sci. Visualization*, 2020(2), pp. 21–36 (in Russ.). doi: 10.26583/sv.12.2.03.
16. Gulicheva, A. (2024) ‘Parallelization of finite-volume numerical methods of computational fluid dynamics by means of shared memory computing systems’, in *CCIS. Proc. HPCST*, 1986, pp. 105–116. doi: 10.1007/978-3-031-51057-1_8.

Авторы

Гуличева Анастасия Алексеевна¹,
 ассистент, gulicheva@mirea.ru
Рыбаков Алексей Анатольевич^{2,3},
 к.ф.-м.н., начальник отдела,
 ведущий научный сотрудник,
 rybakov@jscc.ru

¹ МИРЭА – Российский технологический университет, г. Москва, 119454, Россия

² Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

³ МСЦ РАН – филиал ФГУ ФНЦ НИИСИ РАН г. Москва, 119334, Россия

Authors

Anastasiya A. Gulicheva¹,
 Assistant, gulicheva@mirea.ru
Alexey A. Rybakov^{2,3},
 Cand. of Sci. (Physics and Mathematics),
 Head of Department, Leading Researcher,
 rybakov@jscc.ru

¹ MIREA – Russian Technological University, Moscow, 119454, Russian Federation

² National Research Centre “Kurchatov Institute”, Moscow, 123182, Russian Federation

³ JSCC RAS – branch of SRISA RAS, Moscow, 119334, Russian Federation

Ускорение расчета критериальной функции в задаче размещения всенаправленных антенн

Ай Мин Тайк¹, С.А. Лупин¹, П.Н. Телегин^{2,3}✉, Б.М. Шабанов^{2,3}

¹ Национальный исследовательский университет «МИЭТ», г. Москва, 124498, Россия

² Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия

³ Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

Ссылка для цитирования

Ай Мин Тайк, Лупин С.А., Телегин П.Н., Шабанов Б.М. Ускорение расчета критериальной функции в задаче размещения всенаправленных антенн // Программные продукты и системы. 2024. Т. 37. № 3. С. 384–392. doi: 10.15827/0236-235X.142.384-392

Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 14.05.2024

После доработки: 13.06.2024

Принята к публикации: 27.06.2024

Аннотация. В работе представлены исследования эффективности табличного метода расчета одного из критериев при решении задачи поиска оптимального варианта размещения набора всенаправленных антенн. Задача определения пространственного положения активных элементов беспроводных сетей относится к задачам многокритериальной дискретной оптимизации. Одним из основных критериев оптимальности решения выступает площадь зоны покрытия. Этот параметр определяет не только доступность сетевых услуг для абонентов, но и затраты оператора на создание и поддержку сетевой инфраструктуры. Для решения задачи размещения антенн в работе используется алгоритм полного перебора вариантов, обеспечивающий нахождение точного решения задач дискретной оптимизации. Вычислительная сложность этого алгоритма зависит как от размеров пространства поиска решения, так и от сложности расчета критериальных функций. Результаты проведенных экспериментов показывают, что табличный метод позволяет в семнадцать раз ускорить вычисление значения критерия, обеспечивающего отсутствие перекрытия зон приема. Метод эффективен и в параллельной реализации алгоритма. При запуске приложения на двенадцати потоках было получено более чем шестикратное снижение времени вычислений. Замена непосредственных вычислений значения критерия на предварительно рассчитанные табличные значения может быть эффективно использована и в других задачах дискретной оптимизации.

Ключевые слова: алгоритм полного перебора, вычислительная сложность, параллельные вычисления, задача максимального покрытия, всенаправленная антенна

Благодарности. Работа выполнена в МСЦ РАН – филиале ФНЦ НИИСИ РАН и НИЦ «Курчатовский институт» по теме FNEF-2024-0016. В проводимых исследованиях использован вычислительный кластер МВС-10П

Введение. Методы оптимизации лежат в основе подавляющего большинства алгоритмов управления, широко используются при проектировании архитектур различных технических и социальных систем [1]. Если пространство поиска решений непрерывно, для нахождения оптимальной стратегии используют методы дифференциальных и интегральных исчислений, аналоговые вычисления. В случае дискретной модели пространства для решения задач используют методы дискретной оптимизации. К подобным задачам сводятся многие практические приложения. Для них существуют алгоритмы, позволяющие находить приемлемое по точности решение за ограниченное время. Однако если необходимо получить точное решение, выбор алгоритмов для его поиска ограничен и сводится к *методу полного перебора* (Brute Force Algorithm, BFA) и его вариациям в виде *метода ветвей и границ* (Branch and Bound, BB) [2]. Эти алгоритмы

имеют высокую вычислительную сложность, что ограничивает их практическое применение. В работе [3] на нескольких примерах показано, что учет особенностей решаемой задачи позволяет существенно сократить пространство поиска решения и, следовательно, вычислительную сложность алгоритма без потери точности решения. В данной работе рассматривается возможность использования BFA для решения еще одной практической задачи – поиска варианта оптимального размещения антенн, которая может быть сведена к задаче дискретной оптимизации.

Цель проводимых исследований заключается в оценке эффективности методов снижения вычислительной сложности BFA. Рассматриваются возможные подходы к ускорению вычисления критериальной функции применительно к задаче нахождения оптимального варианта размещения всенаправленных антенн.

Постановка задачи

Рассмотрим формальную постановку задачи размещения антенн. Отметим, что интерес будут представлять вычислительные аспекты ее решения, а не особенности технической реализации каналов связи.

Введем следующие обозначения:

Π – пространство поиска решения или область размещения антенн;

$A = \{A_i\}, i = 1, N$ – множество подлежащих размещению всенаправленных антенн;

$S_i = \pi R_i^2$ – площадь зоны покрытия i -й антенны, где R_i – дальность действия антенны;

$S(\Pi \cap S_i)$ – площадь пересечения зоны покрытия i -й антенны и области размещения антенн;

$$S_i^{ov} = \sum_{i=1}^{N, j \neq i} S(S_i \cap S_j) \text{ – площадь пересечения}$$

зоны покрытия i -й антенны со всеми остальными антеннами.

Критериальные функции, определяющие цель оптимизации:

$$F_1 = \max \sum_{i=1}^N S(\Pi \cap S_i) \text{ – максимизация площади}$$

покрытия области поиска решения заданным набором антенн;

$$F_2 = \min \sum_{i=1}^N (S_i - S(\Pi \cap S_i)) \text{ – минимизация}$$

площади зон покрытия антенн, выходящих за границу области поиска решения;

$$F_3 = \min \sum_{i=1}^N (S_i^{ov}) / 2 \text{ – минимизация площади}$$

пересечения зон покрытия антенн;

$F_3 = 0$ – зоны покрытия антенн не должны иметь пересечения, предельный случай критерия F_3 ;

$F_2 = 0$ – области зон покрытия антенн не должны выходить за границу области поиска решения, предельный случай критерия F_2 .

На рисунке 1 показаны несколько примеров, иллюстрирующих выполнение описанных условий.

С алгоритмической точки зрения представленную задачу можно рассматривать как NP-полную, поскольку для нахождения ее точного решения нужно перебрать все варианты. Вычислительная сложность NP-полных задач зависит не только от размерности пространства поиска решения, хотя этот фактор в большинстве случаев считается определяющим. Поскольку в процессе перебора всех возможных вариантов необходимо для каждого из них

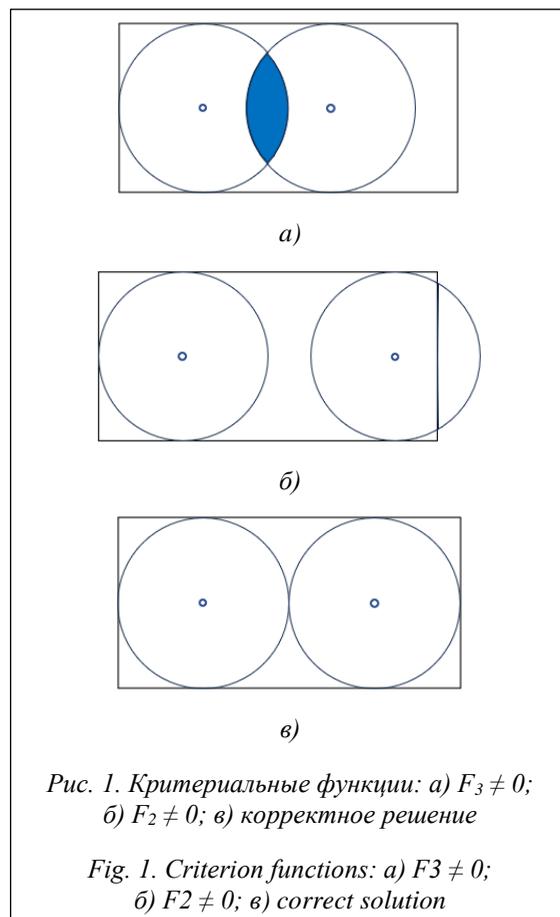


Рис. 1. Критериальные функции: а) $F_3 \neq 0$; б) $F_2 \neq 0$; в) корректное решение

Fig. 1. Criterion functions: а) $F_3 \neq 0$; б) $F_2 \neq 0$; в) correct solution

найти значение критериальной функции, вычислительная сложность этой процедуры влияет на сложность решения задачи, как и размерность пространства поиска. В случае большой размерности задачи единственной возможностью найти ее точное решение является реализация алгоритма на параллельной вычислительной системе. Это значит, что алгоритм поиска должен не только обладать хорошей масштабируемостью для эффективного использования вычислительных ресурсов, но и учитывать ограниченную пропускную способность каналов связи для межузловой коммуникации.

Обзор исследований

Проблема оптимизации покрытия в системах беспроводной связи является сложной комбинаторной задачей. Для ее решения используются методы математического моделирования и дискретной оптимизации. Рассмотрим некоторые наиболее часто используемые подходы.

В статье [4] описан алгоритм нахождения оптимального размещения дискообразных объ-

ектов для максимизации зоны покрытия. Показано, что использование имитационного отжига в контексте максимизации зоны покрытия может быть применимо для решения сложных задач пространственной оптимизации, а параллельный вариант алгоритма обеспечивает и сокращение времени вычислений.

В работе [5] представлен алгоритм, направленный на оптимизацию размещения беспроводных датчиков с различными диапазонами передачи для достижения максимально возможного охвата в пределах региона наблюдения. Авторы реализуют задачу максимизации зоны покрытия в WSNs с использованием генетического алгоритма, который модифицирован для этой проблемы и снижения времени выполнения.

Еще один подход описан в работе [6]. Авторы предлагают решать проблему позиционирования антенн в сотовых сетях с помощью эвристического подхода, основанного на алгоритме табуированного поиска, который, кроме запретов, использует целевые критерии и диверсификацию для снижения сложности проблемы. Показано, что высокая вычислительная сложность и ресурсоемкость алгоритма определяются большим количеством рассматриваемых поисковых комбинаций и требованиями к памяти.

Авторы работы [7] представляют многоцелевой алгоритм размещения передатчиков, который рассматривает зону покрытия и мощность передачи в качестве целей и решает проблему неопределенности при оценке параметров антенны. Предлагаемый алгоритм позволяет оптимизировать размещение всенаправленных передатчиков при одновременном снижении вычислительных затрат. Алгоритм демонстрирует практическое применение многокритериальной оптимизации для максимизации зоны покрытия.

Возможность использования методов математического моделирования для оптимизации покрытия в беспроводных сетях 5G обсуждается в статье [8]. Анализ сложности задачи позволил авторам отнести ее к классу NP-полных, что определяет и выбор алгоритмических подходов к решению.

В работе [9] обсуждается комбинированный эвристический алгоритм для оптимизации зоны покрытия WSNs. Описаны оригинальный алгоритм и его авторская модификация, приводятся экспериментальные результаты, демонстрирующие эффективность предложенного метода для оптимизации параметров WSNs.

Анализ и сравнение различных алгоритмов размещения антенн представлен в статье [10]. Отмечается, что размещение антенн является важнейшим этапом в процессе проектирования сотовых сетей. Эта задача включает в себя принятие решений о нахождении мест установки антенн, а также их типах и количестве для каждой вышки. Классические алгоритмы дискретной оптимизации, обладающие высокой сложностью, применимы только для ограниченного числа антенн, поэтому авторы предлагают использовать генетический алгоритм. Результаты вычислений с применением реальных наборов данных свидетельствуют об эффективности предложенного эвристического подхода.

В работе [11] обсуждается фундаментальная проблема назначения диапазона в WSNs и ставится цель минимизировать стоимость покрытия за счет оптимизации радиусов обнаружения и связи для сенсорных узлов. Такая оптимизация необходима для сокращения затрат времени и энергии на передачу показаний датчиков, что в итоге повышает эффективность и быстродействие WSNs.

Рассмотренные работы показывают, что высокая вычислительная сложность оптимизационной задачи заставляет исследователей отказываться от нахождения точного решения и использовать квазиоптимальные эвристики.

Проводимые авторами данной статьи исследования направлены на поиск методов снижения вычислительной сложности задачи размещения антенн, чтобы для ее решения можно было использовать универсальный оптимизационный алгоритм – метод BFA, гарантирующий нахождение точного решения.

Всенаправленные антенны являются важным компонентом систем радиосвязи для приема и передачи сигналов во всех направлениях. Такая схема приема позволяет устранять мертвые зоны для приложений, требующих равномерного покрытия зоны приема [12]. Их проще устанавливать благодаря 360-градусной диаграмме направленности и отсутствию необходимости в точном позиционировании. В работе [13] выделены следующие области применения всенаправленных антенн: сотовая связь, мобильные устройства, беспроводные телефоны, беспроводные компьютерные сети, антенны радиовещания, точки беспроводного доступа, базовые станции полиции и такси, связь с самолетами, интернет вещей (IoT). Всенаправленные антенны часто используются и в беспроводных сенсорных сетях (WSN).

В работе [14] уже были определены универсальные способы снижения вычислительной сложности BFA. К ним относятся уменьшение пространства поиска решения, сокращение времени расчета критериальной функции, использование эффективных методов генерации вариантов решения, распараллеливание.

Рассмотрим, как можно использовать эти способы для ускорения вычислений в задаче максимизации зоны покрытия всенаправленных антенн.

1. *Уменьшение пространства поиска.* Метод BFA предполагает, что должны быть исследованы все возможные варианты расположения антенн на заданной территории [15]. В случае дискретного пространства число позиций, которые могут занимать антенны, задается количеством шагов сетки M_x и M_y по осям X и Y соответственно. Если $M_x = M_y = M$, то количество положений антенны равно $K = M^2$. Тогда общее количество вариантов решения для N антенн составит K^N . Если используется критерий $F_2 = 0$, то при размере сетки меньше радиуса антенны крайние позиции можно исключить (рис. 2).

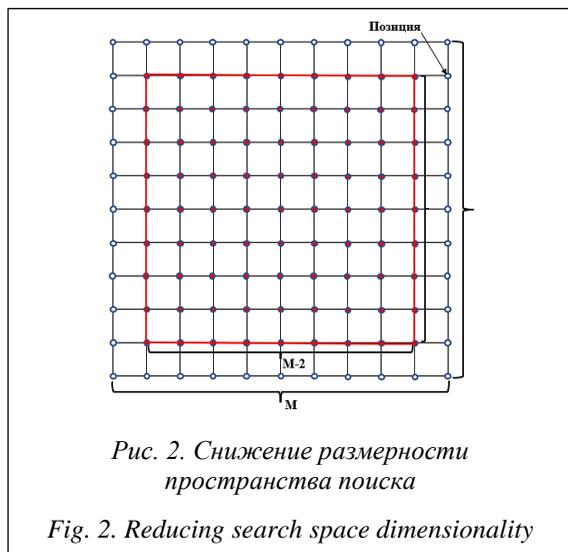


Рис. 2. Снижение размерности пространства поиска

Fig. 2. Reducing search space dimensionality

Тогда количество положений антенны уменьшится и составит $K_1 = (M - 2)^2$. Разница будет $K - K_1 = M^2 - (M - 2)^2 = 4(M - 1)$. При этом ускорение для N элементов определяется как

$$Acc = \left(\frac{M}{M - 2} \right)^{2N}.$$

2. *Сокращение времени расчета критериальной функции.* В рассматриваемой задаче выполнение критерия $F_3 = 0$ предполагает, что расстояние между центрами расположения пары антенн будет больше суммы их радиусов

покрытия. Пусть первая антенна A_1 расположена в позиции с координатами (x_1, y_1) , а вторая A_2 – в позиции (x_2, y_2) . Тогда для проверки выполнимости критерия необходимо проверить условие:

$$(R_1 + R_2) \leq \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (1)$$

Отметим, что координаты расположения антенны необходимо предварительно вычислить, используя номер занимаемой ею позиции. Реализация алгоритма, когда для вычисления критерия используется выражение (1), обозначена в таблицах как P_{calc} .

Вычисление правой части выражения (1) требует многократного выполнения нескольких операций, которые можно однократно произвести заранее. Применительно к рассматриваемой задаче можно сделать предположение, что формирование таблицы расстояний между всеми позициями антенн позволит сократить время проверки критерия. Тогда выражение (1) примет вид

$$(R_1 + R_2) \leq D_{1,2}, \quad (2)$$

$$\text{где } D_{1,2} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

Вычисленные значения расстояний D_{ij} между центрами всех возможных мест расположения антенн формируют матрицу, и для проверки условия (2) достаточно взять нужное значение из памяти. Реализация алгоритма, в котором для вычисления критерия используются табличные значения (2), обозначена как P_{tabl} .

3. *Использование эффективных методов генерации вариантов решения.* В исследовании [16] показано, что генерация вариантов размещения с помощью метода лексикографических перестановок обеспечивает ускорение решения задачи.

4. *Распараллеливание.* BFA хорошо распараллеливается, что позволяет использовать для его реализации и многоядерные процессоры, и графические ускорители. Важно отметить, что метод генерации вариантов также должен позволять параллельную реализацию. Выбранный метод удовлетворяет этому требованию.

Тестовая задача

Условия тестовых задач приведены в таблице 1, характеристики антенн – в таблице 2. Общее число размещаемых антенн $N = 6$, пространство поиска решения или область размещения антенн Π представляет собой квадрат размером 9×9 км, шаг сетки 1×1 км. В первой

задаче размещаются антенны трех разных типов, а во второй все антенны одинаковые.

Целью оптимизации является максимизация покрытия $F_1 \rightarrow \max$ при условии равенства критериев F_2 и F_3 нулю.

Таблица 1

Условия задач

Table 1

Task conditions

Задача_1		Задача_2	
Тип	Количество	Тип	Количество
A ₁	4	A ₁	6
A ₂	1	A ₂	0
A ₃	1	A ₃	0

Таблица 2

Параметры всенаправленных антенн

Table 2

Parameters of omnidirectional antennas

Тип антенны	Количество антенн	R (км)
A ₁	4	1
A ₂	1	2
A ₃	1	3

Решением задачи является N -мерный вектор $P = \{P_i\}$, $i = 1, 6$, элементы которого определяют номер позиции, которую занимает соответствующая антенна.

Проводимые исследования направлены на оценку эффекта, который дает сокращение времени вычисления критериального выражения. Для этого сравним время решения задачи для двух реализаций алгоритма – P_{calc} и P_{tabl} . Теоретический анализ сложности показывает, что реализация P_{tabl} должна работать быстрее.

Программная реализация алгоритма

Основные процедуры алгоритма полного перебора вариантов применительно к задаче нахождения максимального покрытия включают следующие этапы.

1. Задание параметров пространства поиска П.
2. Чтение матриц описания антенн $A = \{A_i\}$.
3. Расчет матрицы расстояний $D = \{D_{i,j}\}$.
4. Генерация варианта решения $P = \{P_i\}$.
5. Проверка выполнимости критериев и запись рекорда.
6. Проверка завершения поиска, если нет, то повтор п. 4.
7. Завершение работы.

В параллельной реализации алгоритма процедуры 4–6 исполняются на разных ядрах. Для этого используется технология многопоточ-

ных вычислений, а распределение вычислений между потоками производится с помощью функции `omp parallel for` из библиотеки OpenMP (<https://hpc-tutorials.llnl.gov/openmp/>). При этом каждый поток формирует свой массив найденных локальных оптимумов, а глобальный оптимум будет найден после объединения всех локальных массивов. Такой подход исключает зависимость по данным между потоками и минимизирует взаимодействие потоков.

Результаты эксперимента

Для оценки эффективности замены вычислений на использование табличных значений при проверке выполнимости критерия проведены вычислительные эксперименты.

В экспериментах использован компьютер с процессором Intel Xeon Processor E5-2630 v2 с 12 физическими ядрами, работающими на частоте 2,6 ГГц. Запуск приложения на одном ядре соответствовал последовательному варианту программы. Для параллельного приложения использованы все 12 физических ядер без режима гипертренинга. Программирование проведено в среде разработки Visual Studio 2019 на языке C++ с использованием библиотеки OpenMP.

Результаты работы последовательного приложения представлены в таблице 3.

Таблица 3

Сокращение времени проверки выполнимости критерия F3 (задача 1)

Table 3

Reducing time for verifying the feasibility of criterion F3 (task 1)

Метод расчета	Время вычисления (сек.)
P_{calc}	19012,4
P_{tabl}	1100,45

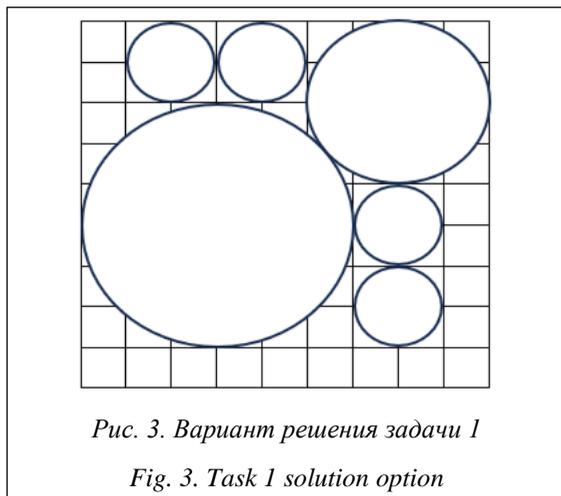
Результаты подтверждают эффективность применения табличных значений для проверки выполнимости критерия. Табличный вариант обеспечивает значительное ускорение вычислений:

$$Acc = T_{calc}/T_{tabl} = 19012,4/1100,45 = 17,28.$$

Оба варианта программы находят все 38 016 возможных вариантов решения задачи. Несколько решений приведены в таблице 4.

Один из представленных вариантов (выделен цветом в таблице) показан на рисунке 3.

На следующем этапе исследований была проверена эффективность табличного варианта проверки выполнимости критерия и при парал-



льной реализации алгоритма. Отметим, что для достижения максимального ускорения вычислений в многопоточном приложении все ядра процессора должны быть загружены одинаково. Это условие выполнимо, если количе-

ство итераций кратно числу иницируемых потоков, а в данном случае – числу физических ядер процессора. В противном случае на последнем этапе не все ядра будут участвовать в вычислениях. Поясним это на небольшом примере. Пусть процессор имеет 12 ядер, а количество итераций внешнего цикла *for* равно 122. Тогда на первом этапе все 12 ядер выполняют по 10 итераций, а на втором будут работать только 2 ядра. В результате максимально ожидаемое ускорение, которое мы сможем получить, уменьшится с 12 до 11,1. По мере увеличения сложности задачи влияние этого фактора на эффективность вычислений будет уменьшаться. В реализованном приложении выравнивание нагрузки ядер производится в генераторе вариантов решения.

Время работы параллельного приложения при решении задач 1 и 2 показано в таблице 5.

Проанализируем полученные в экспериментах результаты.

Некоторые варианты решения (задача 1)

Таблица 4

Several solutions (task 1)

Table 4

Вариант	Расположение антенн											
	A ₁		A ₁		A ₁		A ₁		A ₂		A ₃	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
1	7	1	7	3	1	7	3	7	7	6	3	3
2	8	3	4	7	2	8	7	1	7	7	3	3
3	7	2	7	4	2	8	4	8	7	7	3	4
4	6	1	3	8	1	6	8	1	2	2	6	5
5	2	8	2	6	8	2	6	2	2	3	6	6

Время работы параллельного приложения (сек.)

Таблица 5

Parallel application operation time (sec.)

Table 5

Число потоков	Задача 1		Задача 2	
	Время	Ускорение	Время	Ускорение
1	1100,45	–	7101,92	–
2	570,57	1,93	3861,46	1,84
3	451,31	2,44	2624,64	2,71
4	304,47	3,61	2036,13	3,51
5	284,11	3,87	1767,37	4,02
6	240,43	4,58	1451,81	4,92
7	247,92	4,44	1405,01	5,1
8	174,86	6,29	1306,07	5,44
9	176,99	6,22	1270,83	5,59
10	184,26	5,97	1255,08	5,66
11	183,47	5,99	1114,65	6,37
12	184,59	5,96	1080,03	6,58

1. *Точность решения.* Использование предварительно сформированной таблицы расстояний сохраняет основное достоинство BFA – нахождение точного решения задачи.

2. *Сокращение времени решения задачи.* Табличный метод обеспечивает значимое ускорение вычислений. В рассмотренной задаче удалось более чем в 17 раз сократить время вычислений. Следует отметить, что размерность матрицы расстояний (N^2) позволяет кешировать ее в памяти каждого ядра. Матрица $\{D_{i,j}\}$ симметрична относительно главной диагонали, что дает возможность хранить в памяти только одну ее половину, но это усложняет процесс чтения данных. Следует заметить, что эффективность табличного хранения данных будет зависеть от вида критериальной функции. Полученное значение ускорения справедливо в отношении критерия $F_3 = 0$.

3. *Масштабируемость.* Эксперименты с параллельной реализацией алгоритма позволили сделать вывод о достаточно хорошей масштабируемости приложения. Однако следует отметить, что в данном случае не исследовалась возможность распараллеливания процедуры вычисления критерия, поскольку для табличной реализации это не имеет смысла. Эксперименты подтвердили только функциональность табличной реализации и в параллельной версии алгоритма.

Причина ограниченности полученных ускорений связана с другими компонентами приложения, а не с табличным методом вычисления критерия. В первую очередь, это относится к используемому для генерации вариантов методу лексикографических перестановок. Эффективно устраняя необходимость проверки получаемого вектора, он не позволяет сделать нагрузку ядер одинаковой. Деградация ускорения наблюдается при числе потоков более 8. Метод, используемый для ограничения размерности пространства поиска, не ограничивает масштабируемость. Параллельная реализация алгоритма с табличным методом вычисления критерия обеспечила значительное ускорение вычислений: $Acc = 17,28 \times 6,29 = 108,69$.

Выводы

Использование различных подходов к снижению вычислительной сложности отдельных этапов алгоритма полного перебора вариантов позволяет значительно расширить границы его применимости при решении практических задач. Низкие требования к межузловому взаимодействию дают возможность реализации BFA не только на суперкомпьютерах, но и в распределенной среде.

В дальнейших работах планируется исследование эффективности табличного метода и для других задач дискретной оптимизации.

Список литературы

1. Zou Y., Zhan Q., Xiang K. A comprehensive method for optimizing the design of a regular architectural space to improve building performance. *Energy Reports*, 2021, vol. 7, pp. 981–996. doi: 10.1016/j.egy.2021.01.097.
2. Тайк А.М., Лупин С.А., Кхаинг М.Т. Методы повышения эффективности алгоритма полного перебора на примере решения задачи о неограниченном ранце // INJOIT. 2023. Т. 11. № 5. С. 41–46.
3. Lupin S., Thike A.M., Tun H. Comparison the various criteria in wireless network topology optimization task. *Proc. OPTIMA-2017*, 2017, pp. 370–377.
4. Coll N., Fort M., Saus M. Coverage area maximization with parallel simulated annealing. *Expert Systems with Applications*, 2022, vol. 202, art. 117185. doi: 10.1016/j.eswa.2022.117185.
5. Hanh N.T., Binh T.T.H., Hoai N.X., Palaniswami M.S. An efficient genetic algorithm for maximizing area coverage in wireless sensor networks. *Information Sci.*, 2019, vol. 488, pp. 58–75. doi: 10.1016/j.ins.2019.02.059.
6. Benmezal L., Benhamou B., Boughaci D. Evolutionary Iterated Local Search meta-heuristic for the antenna positioning problem in cellular networks. *Computational Intelligence*, 2022, vol. 38, no. 3, pp. 1183–1214. doi: 10.1111/coin.12454.
7. Parnianifard A., Mumtaz S., Chaudhary S., Imran M.A., Wuttisittikulkij L. A data driven approach in less expensive robust transmitting coverage and power optimization. *Sci. Reports*, 2022, vol. 12, art. 17725. doi: 10.1038/s41598-022-21490-z.
8. Seda P., Seda M., Hosek J. On mathematical modelling of automated coverage optimization in wireless 5G and beyond deployments. *Applied Sci.*, 2020, vol. 10, no. 24, art. 8853. doi: 10.3390/app10248853.
9. Zhu F., Wang W. A coverage optimization method for WSNs based on the improved weed algorithm. *Sensors*, 2021, vol. 21, no. 17, art. 5869. doi: 10.3390/s21175869.
10. Calles-Esteban F., Olmedo A.A., Hellin C.J. et al. Optimizing antenna positioning for enhanced wireless coverage: A genetic algorithm approach. *Sensors*, 2024, vol. 24, no. 7, art. 2165. doi: 10.3390/s24072165.
11. Das S.K., Kapelko R. On the range assignment in wireless sensor networks for minimizing the coverage-connectivity cost. *ACM TOSN*, 2021, vol. 17, no. 4, art. 46, pp. 1–48. doi: 10.1145/3457408.
12. Yang C., Huo X., Wang X., Yang B., Zhang G. Design of an omnidirectional antenna with a compact size. *Proc. IRC*, 2019, vol. 2019, no. 19, pp. 6268–6271. doi: 10.1049/joe.2019.0254.

13. Qing X., Ning C.Z. Omnidirectional antennas. Handbook of Antenna Technologies, 2016, pp. 1415–1478. doi: 10.1007/978-981-4560-44-3_52.

14. Ай Мин Тайк, Лупин С.А., Федяшин Д.А. Использование библиотеки MPI для параллельной реализации алгоритма полного перебора вариантов // Программные продукты и системы. 2023. Т. 36. № 4. С. 607–614. doi: 10.15827/0236-235X.144.607-614.

15. Liu Y., Jian Y., Sivakumar R., Blough D.M. Maximizing line-of-sight coverage for mmWave wireless LANs with multiple access points. IEEE/ACM Transactions on Networking, 2022, vol. 30, no. 2, pp. 698–716. doi: 10.1109/TNET.2021.3122378.

16. Лупин С.А., Ай Мин Тайк. Сравнение методов перестановки с двумя типами структур данных // Международный научно-исследовательский журнал. 2023. № 9. С. 1–5.

Software & Systems

doi: 10.15827/0236-235X.142.384-392

2024, 37(3), pp. 384–392

Accelerating criterion function calculation in an omnidirectional antenna placement problem

Aye Min Thike¹, Sergey A. Lupin¹, Pavel N. Telegin^{2,3}✉, Boris M. Shabanov^{2,3}

¹ National Research University of Electronic Technology, MIET,
Moscow, 124498, Russian Federation

² Joint Supercomputer Center of RAS, Moscow, 119334, Russian Federation

³ National Research Centre “Kurchatov Institute”,
Moscow, 123182, Russian Federation

For citation

Aye Min Thike, Lupin, S.A., Telegin, P.N., Shabanov, B.M. (2024) ‘Accelerating criterion function calculation in an omnidirectional antenna placement problem’, *Software & Systems*, 37(3), pp. 384–392 (in Russ.). doi: 10.15827/0236-235X.142.384-392

Article info

Received: 14.05.2024

After revision: 13.06.2024

Accepted: 27.06.2024

Abstract. The paper presents a study of the tabular effectiveness method for calculating one of the criteria when solving the problem of the optimal solution for placing a set of omnidirectional antennas. The problem of the active network elements placement belongs to multicriteria discrete optimization problems. The coverage area is one of the main criteria of the solution optimality. This parameter determines both the availability of network services for subscribers and operator's costs for creating and maintaining a network infrastructure. In order to solve the antenna placement problem, this paper proposes using a brute-force algorithm (BFA), which can find an exact solution to discrete optimization problems. The computational complexity of BFA depends on the size of the solution search space and on the complexity of calculating criterion functions. The experiments show that the tabular method is 17 times faster than direct calculation of the criterion value that ensures no overlaps of reception areas. The tabular method for the criterion function calculation is also effective for parallel BFA implementation. When running the application on 12 threads, computation time decreased more than sixfold. The substitution of direct computation of a criterion value on pre-calculated tabulated values can be also effective in other discrete optimization problems.

Keywords: brute-force algorithm, computational complexity, parallel computing, maximum coverage problem, omnidirectional antenna

Acknowledgements. The research was carried out at the Joint Supercomputer Center of the Russian Academy of Sciences – the Branch of Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences” and the National Research Centre “Kurchatov Institute” as a part of the government assignment on the topic FNEF-2024-0016

References

1. Zou, Y., Zhan, Q., Xiang, K. (2021) ‘A comprehensive method for optimizing the design of a regular architectural space to improve building performance’, *Energy Reports*, 7, pp. 981–996. doi: 10.1016/j.egy.2021.01.097.
2. Thike, A.M., Lupin, S., Khaing, M.T. (2023) ‘Methods for improving the efficiency of Brute-force algorithm by the example of solving an Unbounded Knapsack Problem’, *INJOIT*, 11(5), pp. 41–46 (in Russ.).
3. Lupin, S., Thike, A.M., Tun, H. (2017) ‘Comparison the various criteria in wireless network topology optimization task’, *Proc. OPTIMA-2017*, pp. 370–377.
4. Coll, N., Fort, M., Saus, M. (2022) ‘Coverage area maximization with parallel simulated annealing’, *Expert Systems with Applications*, 202, art. 117185. doi: 10.1016/j.eswa.2022.117185.

5. Hanh, N.T., Binh, T.T.H., Hoai, N.X., Palaniswami, M.S. (2019) 'An efficient genetic algorithm for maximizing area coverage in wireless sensor networks', *Information Sci.*, 488, pp. 58–75. doi: 10.1016/j.ins.2019.02.059.
6. Benmezal, L., Benhamou, B., Boughaci, D. (2022) 'Evolutionary Iterated Local Search meta-heuristic for the antenna positioning problem in cellular networks', *Computational Intelligence*, 38(3), pp. 1183–1214. doi: 10.1111/coin.12454.
7. Pamianifard, A., Mumtaz, S., Chaudhary, S., Imran, M.A., Wuttisittikulij, L. (2022) 'A data driven approach in less expensive robust transmitting coverage and power optimization', *Sci. Reports*, 12, art. 17725. doi: 10.1038/s41598-022-21490-z.
8. Seda, P., Seda, M., Hosek, J. (2020) 'On mathematical modelling of automated coverage optimization in wireless 5G and beyond deployments', *Applied Sci.*, 10(24), art. 8853. doi: 10.3390/app10248853.
9. Zhu, F., Wang, W. (2021) 'A coverage optimization method for WSNs based on the improved weed algorithm', *Sensors*, 21(17), art. 5869. doi: 10.3390/s21175869.
10. Calles-Esteban, F., Olmedo, A.A., Hellin, C.J. et al. (2024) 'Optimizing antenna positioning for enhanced wireless coverage: A genetic algorithm approach', *Sensors*, 24(7), art. 2165. doi: 10.3390/s24072165.
11. Das, S.K., Kapelko, R. (2021) 'On the range assignment in wireless sensor networks for minimizing the coverage-connectivity cost', *ACM TOSN*, 17(4), art. 46, pp. 1–48. doi: 10.1145/3457408.
12. Yang, C., Huo, X., Wang, X., Yang, B., Zhang, G. (2019) 'Design of an omnidirectional antenna with a compact size', *Proc. IRC*, 2019(19), pp. 6268–6271. doi: 10.1049/joe.2019.0254.
13. Qing, X., Ning, C.Z. (2016) 'Omnidirectional antennas', *Handbook of Antenna Technologies*, pp. 1415–1478. doi: 10.1007/978-981-4560-44-3_52.
14. Aye Min Thike, Lupin, S.A., Fedyashin, D.A. (2023) 'Using MPI library for parallel implementation of a brute-force algorithm', *Software & Systems*, 36(4), pp. 607–614 (in Russ.). doi: 10.15827/0236-235X.144.607-614.
15. Liu, Y., Jian, Y., Sivakumar, R., Blough, D.M. (2022) 'Maximizing line-of-sight coverage for mmWave wireless LANs with multiple access points', *IEEE/ACM Transactions on Networking*, 30(2), pp. 698–716. doi: 10.1109/TNET.2021.3122378.
16. Lupin, S.A., Aye Min Thike (2023) 'Comparison of permutation methods with two types of data structures', *Int. Research J.*, (9), pp. 1–5.

Авторы

Ай Мин Тайк¹, к.т.н., докторант,
ayeminthike52@gmail.com
Лупин Сергей Андреевич¹, к.т.н.,
профессор, lupin@miec.ru
Телегин Павел Николаевич^{2,3}, к.т.н.,
ведущий научный сотрудник, ptelegin@jssc.ru
Шабанов Борис Михайлович^{2,3},
д.т.н., чл.-корр. РАН, директор,
shabanov@jssc.ru

Authors

Aye Min Thike¹, Cand. of Sci. (Engineering),
Doctoral Student, ayeminthike52@gmail.com
Sergey A. Lupin¹, Cand. of Sci. (Engineering),
Professor, lupin@miec.ru
Pavel N. Telegin^{2,3}, Cand. of Sci. (Engineering),
Leading Researcher, ptelegin@jssc.ru
Boris M. Shabanov^{2,3}, Dr.Sci. (Engineering),
Corresponding Member of the RAS,
Director, shabanov@jssc.ru

¹ Национальный исследовательский университет «МИЭТ», г. Москва, 124498, Россия
² Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия
³ Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

¹ National Research University of Electronic Technology, MIET, Moscow, 124498, Russian Federation
² Joint Supercomputer Center of RAS, Moscow, 119334, Russian Federation
³ National Research Centre "Kurchatov Institute", Moscow, 123182, Russian Federation

Визуальная оценка качества работы генератора псевдослучайных чисел для решения криптографических задач

Т.М. Татарникова ¹✉, Д.А. Булгаков ¹

¹ Санкт-Петербургский государственный университет аэрокосмического приборостроения, г. Санкт-Петербург, 190000, Россия

Ссылка для цитирования

Татарникова Т.М., Булгаков Д.А. Визуальная оценка качества работы генератора псевдослучайных чисел для решения криптографических задач // Программные продукты и системы. 2024. Т. 37. № 3. С. 393–401. doi: 10.15827/0236-235X.142.393-401

Информация о статье

Группа специальностей ВАК: 2.3.6

Поступила в редакцию: 02.05.2024

После доработки: 10.06.2024

Принята к публикации: 27.06.2024

Аннотация. Статья посвящена методам и тестам, используемым для оценки различных генераторов случайных чисел (ГСЧ), и разработке приложения для проведения приблизительной визуальной оценки ГСЧ. Описываются разновидности и ключевые характеристики ГСЧ. Рассматриваются математические методы и программные тестовые пакеты, предназначенные для оценки ГСЧ, такие как тесты Национального института стандартов и технологий США (NIST), Dieharder, PractRand, ENT и RaBiGeTe. Предлагается альтернативный способ проведения быстрой приблизительной оценки качества генерации при помощи визуализации. Суть предлагаемого решения в том, чтобы, во-первых, разделять каждое целое число на три части – по 8 бит каждая и, во-вторых, преобразовывать полученные значения в цвета RGB, которые кодируются тремя байтами. Функционал движка рендеринга позволяет сформировать растровую текстуру – матрицу пикселей из полученных чисел с наложением этой текстуры на 2D-плоскость. Для реализации этой процедуры использованы средства игрового движка Unity. Также в приложении предусмотрен инструмент генерации последовательностей из 65 536 целых положительных чисел при помощи четырех разных алгоритмов генерации случайных чисел: System Random от компании Microsoft, Unity Random от разработчиков игрового движка Unity, стандартного алгоритма Rand языка C и алгоритма Лемера. Получаемая визуализация позволяет пользователю легко обнаруживать в исходном наборе чисел различные повторяющиеся структуры или артефакты. Работа визуализатора протестирована как на заведомо неслучайных наборах чисел, так и на истинно случайных числах, полученных от квантового генератора. В перспективе планируется применение визуализации для начальной быстрой оценки результатов генерации случайных чисел в блокчейн сети.

Ключевые слова: генератор случайных чисел, энтропия, тестирование генератора случайных чисел, визуализация работы генератора случайных чисел, приложение для визуализации

Введение. Генераторы случайных и псевдослучайных чисел являются одной из ключевых компонент информационных систем для разнообразных сфер деятельности. Случайные числа используются в криптографических приложениях для создания шифров и ключей безопасности, в финансовых моделях для симуляции рыночных колебаний и прогнозирования инвестиций, в научных исследованиях при моделировании случайных процессов, в статистическом анализе и, конечно, в разнообразных игровых и развлекательных приложениях. Сегодня существует множество алгоритмов и техник для получения случайных и псевдослучайных чисел, поэтому разработчик приложения неизбежно столкнется с задачей оценки качества работы того или иного генератора для принятия решения о пригодности его использования.

Существуют два основных типа генераторов случайных последовательностей: *генера-*

торы случайных чисел (ГСЧ, англ. RNG) и *генераторы псевдослучайных чисел* (ГПСЧ, англ. PRNG). Оба типа создают набор нулей и единиц, которые затем могут быть разделены на блоки чисел требуемых длины и типа.

Генератор истинно случайных нулей и единиц можно представить как абстрактный механизм для подбрасывания идеальной монетки со сторонами, помеченными 0 и 1. Если принять в расчет, что броски не зависят друг от друга, вероятность получения 0 или 1 будет равна 50 %. Главное преимущество такого идеального генератора в том, что значение следующего элемента последовательности невозможно предсказать. На практике ГСЧ для создания случайности использует источник энтропии, обрабатываемый при помощи некоторой функции. Этот процесс, известный как дистилляция энтропии, позволяет нивелировать слабости источника, которые могут приводить к появлению чрезмерно длинных или повторя-

ющихся последовательностей нулей и единиц. В качестве источника энтропии современные ГСЧ могут использовать, например, шумы токов в электрической цепи (https://www.researchgate.net/publication/306370204_Intel's_digital_random_number_generator_DRNG), движения курсора мыши по экрану или белый шум звукового кодека.

ГПСЧ использует одно или несколько входных данных – так называемых зерен (англ. seed), которые в идеальном случае должны быть случайными и непредсказуемыми. Можно сказать, что для получения максимально качественного результата генерации ГПСЧ должен получать начальные значения из выходных данных ГСЧ, то есть от источника энтропии. Выходные данные ГПСЧ обычно являются детерминированными функциями начального значения, из-за чего при описании таких генераторов и применяется термин «псевдослучайный». Важная особенность всех ГПСЧ – повторяемость. Это значит, что каждый элемент псевдослучайной последовательности можно воспроизвести из начального числа. Свойство повторяемости позволяет выполнять проверку генерации или генерировать общий набор данных для отдельных клиентов приложения.

Ключевой характеристикой генераторов, пригодных для использования в криптографических приложениях, является непредсказуемость. ГПСЧ обладают свойствами прямой и обратной непредсказуемости. Прямая непредсказуемость – это невозможность предсказать следующее выходное число данной последовательности (даже если известны все предыдущие числа), если зерно остается неизвестным. Обратная непредсказуемость – это невозможность определить зерно на основе знания о любых сгенерированных значениях. Для качественного ГПСЧ не должна быть очевидной никакая корреляция между зерном и любым числом, сгенерированным с использованием этого зерна. То есть каждый элемент последовательности должен казаться результатом независимого случайного события. Однако если зерно и алгоритм генерации станут известны, тогда все генерируемые значения можно предсказать и случайность исчезнет. Математические алгоритмы большинства современных ГПСЧ общеизвестны (например, вихрь Мерсенна), поэтому для получения непредсказуемых последовательностей крайне важно хранить начальное значение в секрете или изменять его непредсказуемым образом.

Обзор существующих методов оценки ГСЧ

Свойства любой случайной последовательности могут быть охарактеризованы и описаны с точки зрения вероятности. Существует несколько способов оценить качество ГСЧ, включая статистические тесты и оценку аппаратуры. Каждый статистический тест служит для проверки гипотезы о случайности входной последовательности. Для этого задается параметр α , обозначающий уровень значимости (то есть вероятность ложноотрицательного результата). Обычно α берется равным 0,01 или 0,001. На основе входных данных статистический тест вычисляет так называемое P -значение – вероятность того, что идеальный ГСЧ сгенерирует последовательность менее случайную, чем проверяемая этим тестом. Если P -значение больше α , то гипотеза о случайности принимается, иначе отвергается [1]. Аппаратные тесты позволяют удостовериться, что физические приборы, используемые для генерации случайных чисел, работают правильно и не содержат уязвимостей или слабых мест, которые могут поставить под угрозу безопасность ГСЧ.

При тестировании двоичных последовательностей принимаются три предположения [2].

1. Единообразие: в любой момент генерации последовательности появление нуля или единицы одинаково вероятно. Для генераторов, выдающих последовательности натуральных десятичных чисел, предположение о единообразии можно переформулировать так: в любой момент генерации последовательности появление любого числа из заданного диапазона одинаково вероятно.

2. Масштабируемость: любой тест, применимый к последовательности, также может быть применен к подпоследовательностям, случайным образом извлеченным из заданной последовательности.

3. Согласованность: генератор должен продемонстрировать единообразное поведение для любых начальных значений.

Существуют две общепризнанные организации, занимающиеся разработкой стандартов, алгоритмов, протоколов и методов тестирования и оценки криптографических систем и продуктов. Это Национальный институт стандартов и технологий США (NIST) и Национальное агентство по кибербезопасности Германии (BSI). Но, помимо государственных институтов, свои программные решения и методы оценки ГСЧ также предлагают независимые

организации и исследователи. Рассмотрим некоторые из наиболее популярных.

Dieharder – расширенный набор из 15 статистических тестов Джорджа Марсальи [3]. Для проведения серии тестов необходимо сформировать бинарный файл, состоящий из последовательности нулей и единиц. Программа генерирует текстовый файл отчета, в котором можно найти рассчитанные P -значения. Также в пакет входит приложение для быстрой генерации псевдослучайных последовательностей по одному из выбранных алгоритмов. Например, первый тест из пакета Dieharder называется Birthday Spacings. Его суть в следующем: выбираются m дней рождения в году, состоящем из n дней. Указываются промежутки между днями рождения. Пусть j – количество значений, которые встречаются в этом списке более одного раза. Значит, j асимптотически распределено по Пуассону со средним значением $m \frac{3}{4^n}$. В тесте принято $n = 224$ и $m = 29$. Отсюда базовое распределение j принимается пуассоновским с параметром $\lambda = \frac{2^{27}}{2^{26}} = 2$. Берется выборка из 500 значений j , и критерий соответствия χ^2 дает P -значение. Первый тест использует биты 1–24 (считая слева) целых чисел в указанном файле. Затем файл закрывается и снова открывается. Далее биты 2–25 используются для указания дней рождения. Затем берутся биты 3–26 и так далее до битов 9–32. Каждый набор битов предоставляет P -значение, а девять P -значений предоставляют выборку для проведения теста Колмогорова–Смирнова.

PractRand – набор из пяти статистических тестов: BCFN, DC6, Gap16, BRank, FPF [4]. Например, тест Gap16 используется для определения значимости интервала между повторением одной и той же цифры. Результаты Gap-16 не должны иметь существенную корреляцию с другими стандартными тестами. Провалы теста Gap16 при P -значениях, близких к 1, обычно свидетельствуют о том, что выходные данные ГПСЧ были слишком систематизированными, например, когда одноцикловый ГПСЧ приблизился к концу своего цикла. С другой стороны, провалы при P -значениях, близких к 0, указывают на то, что промежутки между некоторыми числами оказались более экстремальными, чем ожидалось. Чаще всего такое поведение характерно для одноцикловых ГПСЧ и ГПСЧ типа RC4.

Тест ENT – инструмент командной строки, рассчитывает пять параметров: энтропию, хи-квадрат, среднее арифметическое, значение Монте Карло для π и коэффициент последовательной корреляции (<https://www.fourmilab.ch/random/>). Тут интересно рассмотреть принцип работы теста Монте Карло для π . Каждая следующая последовательность из шести байтов преобразуется в 24-битные координаты (x, y) точки внутри квадрата. Если расстояние до точки меньше радиуса круга, вписанного в квадрат, шестибайтовая последовательность считается попаданием. Процент попаданий можно использовать для расчета значения π . Если последовательность близка к случайной, то для очень больших потоков чисел результат теста будет приближаться к истинному значению π .

RaBiGeTe MT – инструмент для оценки качества ГПСЧ, содержащий серию из 24 статистических тестов для обнаружения систематических ошибок или закономерностей в выходных данных ГПСЧ (http://cristianopi.altervista.org/RaBiGeTe_MT/). Тест RaBiGeTe MT поддерживает многопоточность, имеет графический интерфейс и обладает широкими возможностями настройки параметров тестов. Входные данные представляют собой библиотеку DLL, содержащую последовательность выходных битов генератора. В качестве примера можно рассмотреть алгоритм теста коротких блоков. Тестируемая последовательность преобразуется в последовательность блоков длины S . Берется некоторое число n , после чего первый блок формируется из битов $0, n, 2n, \dots, (s-1)n$; второй блок формируется из битов $1, n+1, 2n+1, \dots, (s-1)n+1$ и т.д. Каждый блок рассматривается как число в диапазоне от 0 до $2^S - 1$. Распределение этих чисел сравнивается с равномерным распределением.

NIST – тестовый пакет, включающий 15 тестов [5]: частотный (монобитный) тест, проверку частоты внутри блока, тест на прогоны, тест на самую длинную серию единиц в блоке, тест ранга двоичной матрицы, тест дискретного преобразования Фурье (спектральный тест), тест на соответствие непересекающихся шаблонов, тест на соответствие перекрывающихся шаблонов, универсальный статистический тест Маурера, тест линейной сложности, последовательный тест, приближенный тест энтропии, тест кумулятивных сумм, тест на случайные отклонения и тест на вариант случайного отклонения.

Например, в монобитном тесте внимание уделяется соотношению нулей и единиц во всей

последовательности. Этот тест оценивает близость доли единиц к 0,5, то есть количество единиц и нулей в последовательности должно быть примерно одинаковым.

Применить его для тестирования натуральных десятичных чисел тоже возможно – для этого понадобится записать все десятичные числа в двоичном виде в единую строку. Нули полученной последовательности преобразуются в -1 типа `int`, единицы преобразуются в $+1$ типа `int`. После этого все -1 и $+1$ последовательно складываются:

$$S_n = x_1 + x_2 + \dots + x_n \quad (1)$$

и полученная сумма делится на квадратный корень из общего количества слагаемых:

$$S_{abs} = \frac{|S_n|}{\sqrt{n}}. \quad (2)$$

На финальном этапе теста вычисляется хвостовая вероятность P путем применения дополняющей функции ошибок к результату предыдущей операции, деленному на корень из 2:

$$\operatorname{erfc}\left(\frac{S_{abs}}{\sqrt{2}}\right), \quad (3)$$

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^{\infty} e^{-u^2} du. \quad (4)$$

Перечисленные тестовые системы позволяют математически точно оценить качество того или иного ГСЧ, отчего широко используются при их проектировании. Например, в [6] представлено семейство равномерных генераторов псевдослучайных чисел, основанных на обобщенных отображениях Коллатца, полученных на основе гипотезы Коллатца и последовательностей Вейля. Для тестирования генераторов автор использует тестовые пакеты NIST, Dieharder, TestU01 (Big Crush) и PractRand. Генераторы Коллатца–Вейля прошли все тесты межпоточковых корреляций, и это подтверждает, что их можно использовать для генерации множества независимых последовательностей.

Работа [7] посвящена разработке цифрового криптографически безопасного ГПЧ (CSPRNG), который включает в себя детерминированный генератор случайных битов (DRBG), отвечающий требованиям безопасности для криптографических приложений, а также источник энтропии, реализованный на базе кольцевого осциллятора Фибоначчи–Галуа (FiGaRO). Авторы данного источника продемонстрировали хорошую портативность и высокий уровень энтропии. Для тестирования генератора, оценки его энтропии и случайности использованы пакеты NIST и BSI.

Предлагаемое решение

Все рассмотренные методы оценки являются математическими, то есть выдают некую численную оценку для входного набора чисел (битов). При написании собственного ГСЧ или модификации существующего алгоритма наравне с математическими методами хочется также иметь средство быстрой, но приближенной оценки качества перемешивания, чтобы можно было на раннем этапе разработки выявлять ошибки, паттерны и повторяющиеся структуры. Человеческий глаз способен очень быстро фиксировать закономерности и повторения на изображениях, но не в наборах из нулей и единиц. Поэтому визуализация массива чисел в виде цветов палитры (красный R, зеленый G, синий B) позволит быстро фиксировать ошибки и артефакты, появляющиеся в выходных последовательностях ГСЧ (<https://habr.com/ru/companies/vk/articles/574414/>).

Кодирование цвета. Положительный диапазон типа `int` составляет от 0 до 2 147 483 647, соответственно, для описания любого числа из этого диапазона понадобится 31 бит. Цветовая палитра TrueColor, являющаяся на сегодня стандартом представления цвета, позволяет описать 16,7 млн. значений цвета, а также 256 уровней прозрачности (альфа-канала). Каждой цветовой компоненте выделяется по 8 бит. Если же отбросить канал прозрачности, останутся 24 бита цвета. Соответственно, можно разбить любое целое 24-битное число на три байта и преобразовать его в цвета для отображения на экране. Существуют три основных типа записи цветов по каналам: десятичный – когда каждому цвету соответствуют значения от 0 до 255, нормализованный – значения от 0 до 1 и шестнадцатеричный – значения от 0 до FF. При использовании шестнадцатеричного кодирования перед числом, обозначающим итоговый цвет, ставится символ решетки.

Конкретно в C# цвет представляется в виде векторной структуры с тремя параметрами типа `float`. Здесь применяется нормализованный тип кодирования с диапазоном от 0.0f до 1.0f. Для разбиения целого числа на байты был написан метод `GetByte`, у которого первым аргументом указывается число, а вторым – желаемый байт. Представим его код:

```
int GetByte (int rn, int byte)
{
    int number = ((rn >> (8 * byte))
% 256 + 256) % 256;
    return number;
}
```

Чтобы получить нормализованное представление цвета, нужно вызвать метод `GetByte` три раза – для нулевого, первого и второго байтов числа, поделить выходное значение на 255 и записать результат в переменные r , g и b , из которых затем будет сформирован вектор типа `Color` для хранения цвета:

```
float r = GetByte(result, 0) / 255f;
float g = GetByte(result, 1) / 255f;
float b = GetByte(result, 2) / 255f;
colorToPaintPixel = new Color(r, g, b);
```

Полученный цвет можно использовать для закраски пикселя текстуры.

Попиксельное закрашивание. В движке Unity существует метод `SetPixel` класса `Texture2D`, который используется для закрашивания конкретного текселя текстуры в заданный цвет. Позиция текселя задается целыми числами x , y . Для облегчения текстурной развертки и корректной работы мип-мэппинга в компьютерной графике принято использовать квадратные текстуры со стороной, кратной степени двойки. Текстура размером 256×256 текселей позволит визуализировать 65 536 случайных чисел, чего вполне достаточно для оценки возникновения паттернов и повторяющихся структур.

Чтобы получить визуализацию массива чисел в виде закрашенных пикселей, необходимо проделать следующие шаги:

- создать квад (`Quad`) – квадратную плоскую поверхность, состоящую из двух треугольников;
- применить к кваду материал типа `Unlit/Texture`;
- при запуске сгенерировать новую 2D-текстуру и применить ее к материалу квада;
- установить режим фильтрации – точечная выборка, чтобы тексели не размывались по краям;
- отключить тайлинг (повторение), чтобы текстура не размывалась по границам квада.

Для модификации текстуры квада в реальном времени был написан скрипт `PixelPaint.cs`:

```
pixelTexture = new Texture2D(xMax,
yMax);
GetComponent<Renderer>().material.mainTexture = pixelTexture;
pixelTexture.filterMode = FilterMode.Point; // Точечная выборка (без фильтрации)
pixelTexture.wrapMode = TextureWrapMode.Clamp; // Отключение тайлинга
```

Для тестирования визуализатора в программе использованы четыре разных алгоритма

ГПСЧ: метод `System.Random` языка C#, метод `Unity Random` игрового движка Unity, классическая функция `Rand` языков C/C++ и алгоритм Лемера.

C Rand. Стандартная функция C и C++ `rand` представляет собой линейный конгруэнтный генератор. Алгоритм имеет следующую формулу. Значения a , c и m установлены как 1 103 515 245, 12 345 и 2^{31} соответственно (<https://microsin.net/programming/dsp/rand-c-source-code.html>):

$$X_{n+1} = (aX_n + c) \bmod m. \quad (5)$$

Поскольку функция `rand` возвращает целое число в диапазоне от 0 до 32 767, для получения чисел от 0 до 2^{24} самым простым способом будем использовать три вызова метода `Rand()` с последующей конкатенацией (<https://learn.microsoft.com/en-us/cpp/c-runtime-library/reference/rand?view=msvc-170>). Первой операцией выполнится побитовый сдвиг влево на 8 бит результата предыдущей генерации. Затем выполнится логическое «И» между результатом новой генерации и числом 255. Далее между двумя полученными операндами выполнится логическое «ИЛИ». В коде этот алгоритм выглядит следующим образом:

```
int randomCRand = (int)rand.Rand() &
255;
randomCRand = randomCRand << 8 |
(int)rand.Rand() & 255;
randomCRand = randomCRand << 8 |
(int)rand.Rand() & 255;
writer.WriteLine(randomCRand); //
Запись результата в файл
```

Алгоритм Лемера. В дополнение к стандартным алгоритмам Microsoft и Unity для проверки работоспособности приложения был взят довольно простой алгоритм американского математика Д. Лемера, представленный в 1954 году. Формула Лемера выглядит следующим образом:

$$X(i) = a \cdot X(i-1) \bmod m. \quad (6)$$

Значение константы $m = 2\,147\,483\,647$ – это максимальное положительное число типа `int`. Константа a изначально равнялась 16 807, но в 1993 году был предложен более качественный вариант алгоритма, где $a = 48\,271$. Для инициализации генератора Лемера можно использовать любое целочисленное зерно в диапазоне от 1 до 2^{31} (<https://learn.microsoft.com/en-us/archive/msdn-magazine/2016/august/test-run-lightweight-random-number-generation>). На практике зерно обычно высчитывается путем преобразования текущих даты и времени в единственное целое

число. Помимо a и m , в генераторе Лемера используются еще две константы: $q = 127\,773$ и $r = 2\,836$. Приведем фрагмент кода алгоритма Лемера на языке C#:

```
public double Next()
{
    int hi = seed / q;
    int lo = seed % q;
    seed = (a * lo) - (r * hi);
    if (seed <= 0)
        seed = seed + m;
    return (seed * 1.0) / m;
}
```

Переменные hi и lo здесь обозначают верхнюю и нижнюю границы диапазона чисел.

Получение инициализирующего числа. У структуры `DateTime` платформы Microsoft .NET есть свойство `Ticks`, которое возвращает число тактов, представляющее дату и время экземпляра (<https://learn.microsoft.com/en-us/dotnet/api/system.datetime.ticks?view=net-8.0>). Для получения псевдослучайного зерна для алгоритмов `C Rand` и `Lehmer` можно использовать свойство `Ticks` для получения количества тактов, прошедших, например, со Дня Победы (09.05.1945). Один такт равен 100 наносекундам, поэтому полученное значение будет 19-значным числом, что потребует использования типа `long`. Теперь для получения псевдослучайного зерна достаточно посчитать остаток от деления количества тактов на 131 072 и преобразовать результат в тип `uint`. Приведем фрагмент кода класса `GetSeedFromTicks`, реализующий этот функционал:

```
DateTime myBirthday = new
DateTime(1945, 05, 09);
DateTime currentDate = DateTime.Now;
// 1 тик = 100 нс
long elapsedTicks = currentDate.
Ticks - myBirthday.Ticks;
long seed; // Зерно = остаток от де-
ления тиков на 65536
seed = (uint)elapsedTicks % 131072;
Debug.Log("Полученное зерно: " +
seed);
return seed;
```

Полученные результаты и экспериментальная проверка

С использованием игрового движка Unity был разработан 2D-интерфейс приложения для визуализации ГПСЧ (рис. 1).

Центральную часть экрана занимает квадрат, на который будет накладываться текстура с попиксельным закрасиванием. В правой части располагаются четыре кнопки выбора алгоритмов. При нажатии на любую из них будут сге-

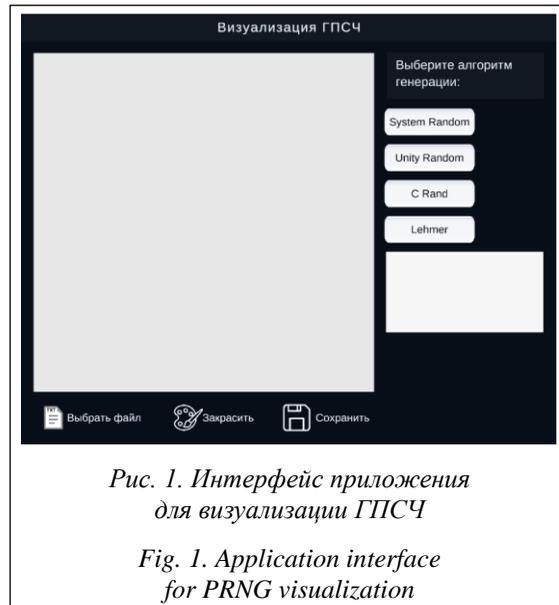


Рис. 1. Интерфейс приложения для визуализации ГПСЧ

Fig. 1. Application interface for PRNG visualization

нерированы 65 536 псевдослучайных чисел с помощью выбранного алгоритма и записаны в текстовый файл. В окне ниже отобразится имя файла, соответствующее названию выбранного алгоритма.

Кнопки внизу экрана позволяют выбрать текстовый файл для визуализации, запустить скрипт, который преобразует числа из файла в цвета текстуры, и сохранить закрасенную текстуру в файл с расширением `.PNG` в папку `Output` в директории программы.

На изображениях рисунка 2 визуально не удается выделить какие-либо повторяющиеся структуры или преобладающие цвета. Это значит, что рассмотренные ГПСЧ в первом приближении дают достаточно качественный результат перемешивания и имеет смысл подвергнуть их дальнейшей углубленной оценке при помощи математических методов NIST или Dieharder.

Если же, например, изменить константу a в алгоритме `C Rand` на 111, то на изображении проступят четко различимые горизонтальные линии, свидетельствующие о наличии повторений (рис. 3а). На рисунке 3б приведена визуализация повторяющихся числовых последовательностей от 1 до 1 023 – в результате получается красно-черный градиент.

Также визуальной оценке был подвергнут ГПСЧ, описанный в [8], где для выбора зерна используются данные с аппаратных датчиков давления и цвета. Его визуализация представлена на рисунке 3с. Зерно меняется через каждые 1 024 числа. Как можно заметить, внешние датчики также обеспечивают высокое качество перемешивания.

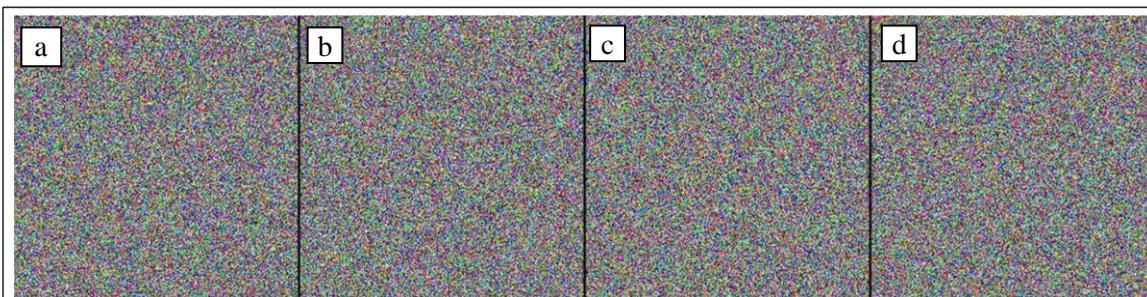


Рис. 2. Визуализации ГПСЧ: a) System Random; b) Unity Random; c) C Rand; d) Lehmer

Fig. 2. PRNG visualization: a) System Random; b) Unity Random; c) C Rand; d) Lehmer

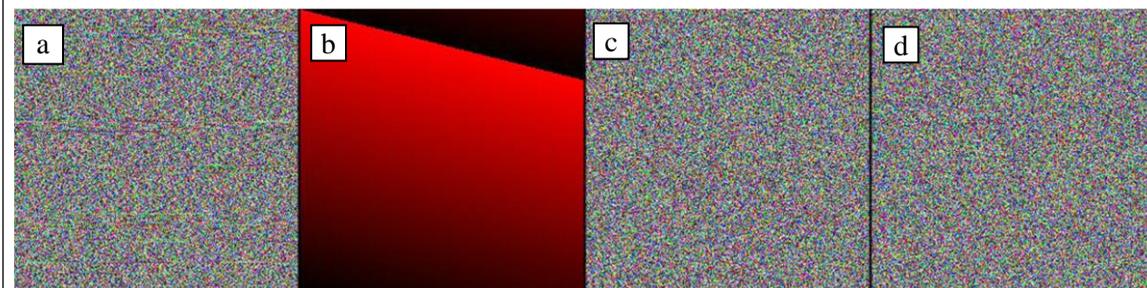


Рис. 3. Визуализация различных числовых последовательностей: a) последовательность, содержащая повторяющиеся числа; b) арифметическая прогрессия; c) последовательность, для определения зерна которой использовались данные с внешних датчиков; d) истинно случайная последовательность

Fig. 3. Visualization of different numerical sequences: a) sequence with repeating numbers; b) arithmetic progression; c) sequence with a grain determined using data from external sensors; d) true random sequence

Для визуального сравнения последовательностей псевдослучайных чисел с истинно случайной последовательностью был взят набор из нулей и единиц, сгенерированный на квантовом генераторе компании «КуРЭйт». Данный генератор основан на интерференции лазерных импульсов со случайной фазой и использует в качестве источника энтропии флуктуации фазы электромагнитного поля в резонаторе полупроводникового лазера [9]. Полученный от него набор данных заранее прошел проверку всеми 15 тестами NIST. Этот случайный набор был разбит на строки по 24 символа (24 бита) с последующим преобразованием строк в 24-битные числа типа int. На рисунке 3d представлена визуализация итоговой случайной последовательности целых десятичных чисел.

Дальнейшие перспективы и выводы

Случайные числа играют ключевую роль в функционировании всей экосистемы блокчейна. Они используются, в частности, для генерации криптографических ключей, обеспе-

чивающих безопасность транзакций и защиту пользовательских криптоактивов, выбора узлов-валидаторов, распределения вознаграждений, генерации случайных идентификаторов транзакций или блоков, определения свойств NFT. При этом подходы к получению случайных чисел в блокчейнах и децентрализованных приложениях делятся на две категории: одни решения работают вне цепочки блоков, другие же для генерации чисел и их проверки используют хэши блоков и смарт-контракты.

Самый простой способ получения случайных чисел вне цепочки – обратиться к третьей стороне, поставщику услуг, который по запросу пользователя будет генерировать число и отправлять его в блокчейн. Минус такого подхода очевиден: все участники сети должны доверять честности поставщика (<https://pyth.network/blog/secure-random-numbers-for-block-chains>).

Для генерации случайных чисел внутри сети можно использовать в качестве источника энтропии хэш будущего блока. Такой подход называется Blockhash. Транзакция запроса сохраняет номер текущего или будущего блока,

после чего узлы-валидаторы сети вычисляют его хэш. Как только хэш станет доступным, транзакция раскрытия вернет его значение. Однако здесь участникам сети тоже приходится доверять честности стороннего участника – валидатора, который может переопределить порядок транзакций или пренебречь определенной транзакцией ради изменения хэша блока, что приведет к гарантированному выпадению желаемого случайного числа.

Верифицируемая случайная функция (VRF) призвана защитить результат от влияния со стороны участников сети. Это криптографический примитив, псевдослучайная функция, которая представляет общедоступное достоверное доказательство своего вывода на основе открытых входных данных и закрытого ключа. Ее можно записать так [10]:

$$F_{SK}(x) = (y, p). \quad (7)$$

В уравнении (7) выходное значение y выглядит случайным, но однозначно вычисляется на основе входных данных x и секретного ключа SK. Функция возвращает доказательство p , которое каждый участник сети может проверить, чтобы убедиться, что y является корректным результатом вычислений. В блокчейне x представляет собой комбинацию из пользовательских данных и хэшей блоков. Поставщик услуг с секретным ключом SK находится вне

сети. Он отслеживает появление в блокчейне запросов на генерацию случайного числа и отправляет ответы в цепочке (y, p) . Транзакция раскрытия проверяет доказательство p , чтобы убедиться, что y является правильным значением. Однако главный недостаток использования VRF опять же лежит в области доверия поставщику, поскольку именно он решает, стоит ли отправлять сгенерированное случайное число в блокчейн или нет. К тому же вычисление функции и доказательства могут потребовать несколько транзакций, за проведение которых участнику придется платить комиссию сети.

Как можно заметить, блокчейны предъявляют особые требования к случайности результатов генерации, поскольку получение неслучайного числа может свидетельствовать о компрометации механизма генерации злоумышленником или попытке поставщика услуг подтасовать результаты. Предлагаемое решение является первым этапом на пути оценки качества и достоверности генерации случайных чисел. Визуализация позволяет проводить быструю оценку последовательностей целых чисел и выявлять в них артефакты, паттерны, повторения и области неравномерного распределения для дальнейшего углубленного анализа популярными математическими методами.

Список литературы

1. Ключарев П.Г. О статистическом тестировании блочных шифров // Математика и математическое моделирование. 2018. № 5. С. 35–56. doi: 10.24108/mathm.0518.0000132.
2. Ryabko B. Time-adaptive statistical test for random number generators. *Entropy*, 2020, vol. 22, no. 6, art. 630. doi: 10.3390/e22060630.
3. Marsaglia G. The marsaglia random number CDROM including the diehard battery of tests of randomness. 1997. URL: <http://www.stat.fsu.edu/pub/diehard/> (дата обращения: 21.05.2024).
4. Sleem L., Couturier R. TestU01 and Pracrtrand: Tools for a randomness evaluation for famous multimedia ciphers. *Multimedia Tools and Applications*, 2020, vol. 79, pp. 24075–24088. doi: 10.1007/s11042-020-09108-w.
5. Bassham L., Rukhin A., Soto J. et al. A statistical test suite for random and pseudorandom number generators for cryptographic applications. NIST SP, 2010, 800-22 Rev. 1. doi: 10.6028/NIST.SP.800-22R1A.
6. Dziala T.R. Collatz-Weyl generators: high quality and high throughput parameterized pseudorandom number generators // Декабрь 2023. ArXiv, doi: 10.48550/arXiv.2312.17043.
7. Crocetti L., Di Matteo S., Nannipieri P., Fanucci L., Saponara S. Design and test of an integrated random number generator with all-digital entropy source. *Entropy*, 2022, vol. 24, no. 2, art. 139. doi: 10.3390/e24020139.
8. Булгаков Д.А. Генерация случайных чисел для интерактивных приложений с использованием внешних датчиков // Изв. вузов. Приборостроение. 2024. Т. 67. № 4. С. 338–344. doi: 10.17586/0021-3454-2024-67-4-338-344.
9. Шаховой Р. QRate Chaos – квантовый генератор случайных чисел // QRate. Pat. № 2721585. URL: <https://www.gograte.com/projects/qrate-chaos-kvantovyy-generator-sluchaynykh-chisel-pat-2721585/> (дата обращения: 21.05.2024).
10. Werapun W., Karode T., Jakapan S., Arpornthip T., Sangiamkul E. NativeVRF: A simplified decentralized random number generator on EVM blockchains. *Systems*, 2024, vol. 11, no. 7, art. 326. doi: 10.3390/systems11070326.

Visual quality assessment of the pseudorandom number generator for solving cryptographic tasksTatiana M. Tatarnikova ¹✉, Dmitry A. Bulgakov ¹¹ Saint Petersburg State University of Aerospace Instrumentation,
Saint Petersburg, 190000, Russian Federation**For citation**Tatarnikova, T.M., Bulgakov, D.A. (2024) 'Visual quality assessment of the pseudorandom number generator for solving cryptographic tasks', *Software & Systems*, 37(3), pp. 393–401 (in Russ.). doi: 10.15827/0236-235X.142.393-401**Article info**

Received: 02.05.2024

After revision: 10.06.2024

Accepted: 27.06.2024

Abstract. The paper focuses on studying the methods and tests for evaluating various random number generators (RNG) and developing an application for a rough visual evaluation of RNG. It describes RNG types and key characteristics. It also considers mathematical methods and software test packages designed for evaluating RNG, such as the US National Institute of Standards and Technology (NIST), Dieharder, PractRand, ENT and RaBiGeTe tests. The paper shows an alternative way to perform a fast approximate evaluation of generation quality using visualization. The essence of the proposed solution is, first, to divide each integer into three parts, 8 bits each. Second, to convert the resulting values into RGB colors, which are also encoded in three bytes. The rendering engine functionality allows generating a bitmap texture - a matrix of pixels from the obtained numbers and overlying this texture on a 2D-plane. Implementing this procedure involves using the Unity game engine. The application also provides a tool for generating sequences of 65,536 positive integers using four different random number generation algorithms: System Random from Microsoft, Unity Random from the developers of the Unity game engine, the standard Rand algorithm of C language, and Lehmer's algorithm. The resulting visualization allows the user to detect the presence of various repetitive structures or artifacts in the initial set of numbers easily. The work of the visualizer tests both on obviously non-random sets of numbers and on truly random numbers obtained from a quantum generator. In the future, the authors propose using visualization for initial quick assessment of random number generation results in a blockchain network.

Keywords: random number generator, entropy, random number generator testing, visualization of random number generator operation, visualization application

References

1. Klyucharev, P.G. (2018) 'On statistical testing of block ciphers', *Math. and Math. Modeling*, (5), pp. 35–56 (in Russ.). doi: 10.24108/mathm.0518.0000132.
2. Ryabko, B. (2020) 'Time-adaptive statistical test for random number generators', *Entropy*, 22(6), art. 630. doi: 10.3390/e22060630.
3. Marsaglia, G. (1997) *The Marsaglia Random Number CDROM Including the Diehard Battery of Tests of Randomness*, available at: <http://www.stat.fsu.edu/pub/diehard/> (accessed May 21, 2024).
4. Sleem, L., Couturier, R. (2020) 'TestU01 and Pracrtrand: Tools for a randomness evaluation for famous multimedia ciphers', *Multimedia Tools and Applications*, 79, pp. 24075–24088. doi: 10.1007/s11042-020-09108-w.
5. Bassham, L., Rukhin, A., Soto, J. et al. (2010) 'A statistical test suite for random and pseudorandom number generators for cryptographic applications', *NIST SP*, (800-22 Rev. 1). doi: 10.6028/NIST.SP.800-22R1A.
6. Dziala, T.R. Collatz-Weyl generators: high quality and high throughput parameterized pseudorandom number generators // December 2023, *ArXiv*. doi: 10.48550/arXiv.2312.17043.
7. Crocetti, L., Di Matteo, S., Nannipieri, P., Fanucci, L., Saponara, S. (2022) 'Design and test of an integrated random number generator with all-digital entropy source', *Entropy*, 24(2), art. 139. doi: 10.3390/e24020139.
8. Bulgakov, D.A. (2024) 'Random number generation for interactive applications using external sensors', *J. of Instrument Eng.*, 67(4), pp. 338–344 (in Russ.). doi: 10.17586/0021-3454-2024-67-4-338-344.
9. Shakhovoy, R. 'QRate Chaos – quantum random number generator', *QRate*, Pat. № 2721585, available at: <https://www.goqr.com/projects/qrate-chaos-kvantovyy-generator-sluchaynykh-chisel-pat-2721585/> (accessed May 21, 2024) (in Russ.).
10. Werapun, W., Karode, T., Jakapan, S., Arpornthip, T., Sangiamkul, E. (2024) 'NativeVRF: A simplified decentralized random number generator on EVM blockchains', *Systems*, 11(7), art. 326. doi: 10.3390/systems11070326.

АвторыТатарникова Татьяна Михайловна ¹, д.т.н.,
профессор, директор института,
tm-tatarn@yandex.ruБулгаков Дмитрий Алексеевич ¹,
аспирант, старший преподаватель,
dmbulg@gmail.com**Authors**Tatiana M. Tatarnikova ¹, Dr.Sci. (Engineering),
Professor, Director University,
tm-tatarn@yandex.ruDmitry A. Bulgakov ¹, Postgraduate Student,
Senior Lecturer,
dmbulg@gmail.com¹ Санкт-Петербургский государственный
университет аэрокосмического приборостроения,
г. Санкт-Петербург, 190000, Россия¹ Saint Petersburg State University
of Aerospace Instrumentation,
Saint Petersburg, 190000, Russian Federation

Компоненты и функциональность программного средства моделирования структуры импортозамещения

Ф.Н. Абу-Абед ¹✉, С.А. Жиронкин ²

¹ Тверской государственный технический университет, г. Тверь, 170026, Россия

² Национальный исследовательский Томский политехнический университет, г. Томск, 634050, Россия

Ссылка для цитирования

Абу-Абед Ф.Н., Жиронкин С.А. Компоненты и функциональность программного средства моделирования структуры импортозамещения // Программные продукты и системы. 2024. Т. 37. № 3. С. 402–411. doi: 10.15827/0236-235X.142.402-411

Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 13.05.2024

После доработки: 10.06.2024

Принята к публикации: 17.06.2024

Аннотация. Предметом данного исследования являются создание основных компонентов средства для моделирования структуры импортозамещения в российской экономике и определение его функциональности. Актуальность заключается в отсутствии на сегодняшний день программных аналогов и разработок на отечественном рынке ИТ для решения поставленной задачи. Для исследования применяются методы системного анализа, имитационного моделирования, визуального проектирования и объектно-ориентированного программирования. Моделирование выпуска продукции для внутреннего рынка требует поиска оптимального сочетания значений объемов факторов производства (капитала и труда) и их эффективности. Для этого необходимо создать программное средство, которое отражает математическую модель импортозамещения, основанную на его целевой функции, и связывает различные классы, БД и пользовательский интерфейс. В настоящей статье приведены основные классы, каждый из которых играет определенную роль в расчетах и моделировании импортозамещения, исследуется их взаимодействие, представленное в виде диаграммы. В разработанном приложении через методы классов выполняются команды и процедуры; представлен алгоритм взаимодействия пользователя с приложением. Предложенная БД приложения включает в себя таблицы переменных, необходимых для определения значений целевой функции (количество отраслей и продуктов, массивы стоимости основного капитала и количества работников, изменения фондоотдачи и производительности труда и прочее) и результатов (искомых отраслевых объемов инвестиций, занятости и выпуска). Интерфейс приложения дает пользователю возможность рассчитать значения целевой функции структуры импортозамещения, используя кнопки, открывающие диалоговые окна для ввода значений и запуска расчетов. Научная новизна заключается в получении результатов, позволяющих оптимизировать решение важной задачи в контексте регулирования национальной экономики – в моделировании структуры импортозамещения в условиях усиления внешних шоков и технологических ограничений.

Ключевые слова: математическая модель, программное средство, структура, классы, пользовательский интерфейс, команды и процедуры, БД, импортозамещение

Благодарности. Исследование выполнено за счет гранта Российского научного фонда № 23-28-01423, <https://rscf.ru/project/23-28-01423/>

Введение. Моделирование структуры импортозамещения в российской экономике представляет собой задачу, решаемую с помощью не только экономическо-математических, но информационно-технологических методов, в контексте поиска оптимальных решений по перераспределению инвестиций, рабочей силы, промежуточного продукта между отраслями.

Разработанные ранее подходы к моделированию импортозамещения основаны на целевой функции выпуска продукции для внутреннего рынка (по критерию оптимизации для достижения его максимума), в качестве аргументов которой предлагаются:

– конкурентоспособность отечественной продукции и производительность отраслевых предприятий по ее выпуску [1];

– зависимость экономики от импорта определенной продукции, объемы ее производства и затраты на выпуск внутри страны [2];

– промежуточное и конечное потребление импорта в экономике, выпуск отечественными предприятиями [3];

– выпуск в доминирующей модели российских рынков – монополий, их фондовооруженность, производительность и капиталотдача [4];

– дополнительные налоги, получаемые государством от предприятий, участвующих в импортозамещении, увеличение заработной платы и социальных выплат в целом по экономике, количество новых рабочих мест [5].

В зарубежной литературе моделирование импортозамещения рассматривается с позиции оптимизации общего равновесия при измене-

нии соотношения экспорта и импорта [6], изменения импорта и внутреннего отраслевого выпуска в ответ на изменения внутренних цен [7], выпуска в ресурсовозобновляемом и невозобновляемом секторах экономики [8], выгод и издержек от протекционизма [9].

В целом в литературе преобладают подходы к моделированию импортозамещения, в основе которых лежат переменные, характеризующие экономическую динамику (затраты, выпуск, налоги, импорт, экспорт и пр.). Такие модели достаточно изучены, и нахождение значений их целевых функций представляется сравнительно простым. В свою очередь, в условиях интенсивного воздействия внешних шоков (таких как санкции и эмбарго) на российскую экономику традиционные динамические модели теряют эффективность, поскольку доступ к ряду ресурсов значительно сокращается, а затраты на выпуск продукции для внутреннего рынка растут, что снижает конкурентоспособность – важный аргумент во многих подходах.

Поэтому, учитывая потери российской экономики от санкций, в данной работе предлагается более соответствующий проблеме подход к моделированию импортозамещения, основанный на оптимизации ее структуры. Следует отметить, что исследований, посвященных разработке программных продуктов в данной области, мало, исключением можно считать разработку алгоритмов [10, 11] и предложение модулей оптимального межотраслевого распределения ресурсов [12].

В связи с этим целесообразно создание программного продукта для модели структуры импортозамещения, целевая функция которой имеет два основных аргумента – объемы и эффективность использования труда и капитала как главных факторов производства (новые технологии учитываются как константа [13], а изменения в уровне образования работников – как экзогенный параметр [14]).

Модели и методы

Целевая функция в модели структуры импортозамещения. С учетом сказанного на уровне производства одного продукта на одном предприятии целевая функция модели структуры импортозамещения будет выглядеть следующим образом:

$$\begin{cases} C_x = \gamma FC \times \Delta CP + \lambda C_r + \varepsilon, \\ L_x = \Delta X \times Q_f \times LP^{-1} + \omega N + \varepsilon, \\ Q_x = \lambda C_x^\lambda + \theta L_x^{\Delta LP}, \end{cases} \quad (1)$$

где C_x – функция инвестиций для импортозамещения с учетом прироста новых технологий и воздействия внешних шоков на отечественный рынок капитала; FC – стоимость основного капитала на предприятии, занятого выпуском одного продукта; L_x – функция трудовых ресурсов для импортозамещения с учетом изменения доли отечественной продукции ΔX , заданной в отраслевых программах; Q_x – функция выпуска продукции с учетом двух факторов – труда и капитала (производственная функция); λ – показатель прироста производимого продукта от внедрения передовых производственных технологий (технологический мультипликатор); γ – показатель, учитывающий воздействие внешних шоков (в т.ч. санкций) на основные средства (потребность в дополнительных инвестициях): потери от ухода иностранных компаний, от ограничения доступа на мировой рынок капитала, от технологических ограничений; ε – ошибки регрессионных вычислений: $\varepsilon \sim N(0, \Sigma)$, где Σ – ковариационная матрица ошибок ε_t [15], в данной модели равная 0; LP – производительность труда; ΔLP – изменение производительности труда за счет внедрения новых передовых производственных технологий в процессе импортозамещения; ω – показатель потерь рынка труда вследствие внешних шоков (дополнительная потребность в рабочей силе); N – число работников, занятых в производстве продукта в целях импортозамещения; Θ – экзогенный параметр, учитывающий потребность в развитии профессионального образования в условиях внешних шоков и технологических ограничений.

Применительно к модели отраслевой структуры импортозамещения в национальной экономике целевая функция модели структуры импортозамещения для выпуска в отрасли i продукта j может быть представлена в следующем виде:

$$\begin{cases} C_x = \sum_{i=1}^n \sum_{j=1}^m \gamma FC_{ij} \times \Delta CP_{ij} + \lambda C_{r_{ij}} + \varepsilon, \\ L_x = \sum_{i=1}^n \sum_{j=1}^m \Delta X_{ij} \times Q_{f_{ij}} \times LP^{-1} + \omega N_{ij} + \varepsilon, \\ Q_{x_{ij}} = \lambda C_{x_{ij}}^\lambda + \theta L_{x_{ij}}^{\Delta LP}, \end{cases} \quad (2)$$

где n – конечное число отраслей i в экономике, занятых в импортозамещении; FC_{ij} – стоимость основного капитала отрасли i , занятой в выпуске продукта j ; ΔCP_{ij} – изменение фондотдачи в отрасли i по выпуску продукта j ; $C_{r_{ij}}$ – объем инвестиций в отрасли i , необходимый

для импортозамещения продукта j ; ΔX_{ij} – изменение доли продукта j отрасли i на внутреннем рынке; Q_{r_j} – текущий (фактический) объем выпуска продукта j в отрасли i ; N_{ij} – число работников отрасли i , занятых в импортозамещении продукта j ; C_{x_j} – функция инвестиций для импортозамещения продукта j в отрасли i ; L_{x_j} – функция трудовых ресурсов для импортозамещения продукта j в отрасли i ; Q_{x_j} – производственная функция для импортозамещения продукта j в отрасли i .

Далее описаны значения нормированного весового показателя факторов, определяющих их значимость, которые были определены в ходе построения матрицы критериев n (конечное значение количества видов структуры импортозамещения) и m (количество видов шоков).

γ – эндогенный показатель, учитывающий воздействие внешних шоков (в том числе санкций) на основные средства – потребность в дополнительных инвестициях: потери от ухода иностранных компаний – эндогенный фактор шокового воздействия на воспроизводственную структуру, $a_{ij} = 0,170$; потери от ограничения доступа на мировой рынок капитала – эндогенный фактор воздействия шоков на структуру факторов производства, $a_{ij} = 0,044$; потери от технологических ограничений – эндогенный фактор шокового воздействия на технологическую структуру, $a_{ij} = 0,325$.

Авторы использовали среднее квадратичное отклонение для усреднения весовых показателей воздействия шоков и санкций на инвестирование основных средств предприятий, занятых в импортозамещении, при котором $\gamma = 1 + a_{ij}^{cp.kb} = 1 + 0,149 = 1,149$.

λ – экзогенный показатель прироста производимого продукта от внедрения передовых производственных технологий (технологический мультипликатор) – экзогенный фактор шокового воздействия на технологическую структуру, $a_{ij} = 0,139$. Тогда $\lambda = 1 + a_{ij} = 1,139$.

ω – экзогенный показатель потерь рынка труда вследствие внешних шоков (дополнительная потребность в рабочей силе вследствие внешних санкций и утечки мозгов и рук), $a_{ij} = 0,68$. Тогда $\omega = 1 + a_{ij} = 1,68$.

Θ – экзогенный параметр, учитывающий потребность в развитии профессионального образования в условиях внешних шоков и технологических ограничений, $a_{ij} = 0,136$. Тогда $\Theta = 1 + a_{ij} = 1,136$.

Описание документации вычислительного обеспечения представленной выше модели

предназначено для пользователей, которые хотят понять структуру, функции и взаимодействие классов и модулей программного средства, разработанного для решения задачи межотраслевого распределения факторов производства (труда и капитала) в условиях внешних шоков с целью импортозамещения.

Разработка программного и алгоритмического обеспечения

Классы и структура их связей. Программное средство состоит из нескольких основных классов, каждый из которых играет определенную роль в расчетах и моделировании.

Класс *Constants* отвечает за хранение и управление константами, используемыми в расчетах. Константы включают в себя такие параметры, как погрешность регрессионных вычислений (`_epsilon`), показатель влияния внешних шоков на основные средства (`_gamma`), показатель прироста производимого продукта от внедрения новых технологий (`_lambda`), показатель потерь рынка труда вследствие внешних шоков (`_omega`) и потребность в развитии профессионального образования (`_teta`). Эти параметры могут быть заданы или изменены пользователем через методы класса.

```
class Constants:
    """
    Класс с константами
    """

    def __init__(self):
        self._epsilon = 0
        self._gamma = 1.149
        self._lambda = 1.139
        self._omega = 1.068
        self._teta = 1.136

    def set_epsilon(self, new_value:
float) -> None:
        """
        Задать эпсилон
        :param new_value: новое значение
        """
        self._epsilon = new_value

    def set_gamma(self, new_value:
float):
        """
        Задать гамма
        :param new_value: новое значение
        """
        self._gamma = new_value

    def set_lambda(self, new_value:
float):
        """
```

```

        Задать лямбда
        :param new_value: новое значение
        """
        self._lambda = new_value

    def set_omega(self, new_value:
float):
        """
        Задать омега
        :param new_value: новое значение
        """
        self._omega = new_value

    def set_teta(self, new_value: float):
        """
        Задать тета
        :param new_value: новое значение
        """
        self._teta = new_value

```

Класс *Variables* предназначен для хранения переменных, необходимых для расчетов. В него входят такие переменные, как количество отраслей (n), количество продуктов (m), массивы стоимости основного капитала (FC_array), изменения фондоотдачи ($delta_CP_array$), изменения фондоотдачи на уровне отрасли (C_r_array), изменения доли отечественного продукта ($delta_X_array$), фактический объем выпуска продукции (Q_f_array), количество работников (N_array), производительность труда (lp) и изменение производительности труда ($delta_lp$). Класс также содержит массивы для хранения результатов расчетов C_x , L_x и Q_x .

```

class Variables:
    """
    Класс переменных для расчета
    """
    def __init__(self, **kwargs):
        self.n = kwargs.get('n')
        self.m = kwargs.get('m')
        self.FC_array =
= kwargs.get('FC_array')
        self.delta_CP_array =
= kwargs.get('delta_CP_array')
        self.C_r_array =
= kwargs.get('C_r_array')
        self.delta_X_array =
= kwargs.get('delta_X_array')
        self.Q_f_array =
= kwargs.get('Q_f_array')
        self.N_array = kwargs.get('N_array')
        self.lp = kwargs.get('lp')
        self.delta_lp =
kwargs.get('delta_lp')

        # Инициализация массивов в этом
        классе
        self.C_x_array = [[0 for _ in
range(self.m)] for _ in range(self.n)]
        self.L_x_array = [[0 for _ in
range(self.m)] for _ in range(self.n)]
        self.Q_x_array = [[0 for _ in
range(self.m)] for _ in range(self.n)]

```

Для вычисления массива C_x используется класс *C_x_calculator*. Этот класс включает метод `_calculate_C_x_array`, который выполняет необходимые расчеты на основе значений констант и переменных. Метод `get_result` возвращает результат вычислений.

```

class C_x_calculator:
    """
    Класс подсчета массива C_x
    """
    def __init__(self, constants_obj:
Constants, variables_obj: Variables):
        self.C_x_result = self._calcula
te_C_x_array(constants_obj, varia
bles_obj)

    def _calculate_C_x_array(self, con
stants_obj, variables_obj):
        """
        Посчитать массив C_x
        """
        for i in range(variables_obj.n):
            for j in range(varia-
bles_obj.m):
                # Вычисляем значение для
                текущего элемента массива
                value = con-
stants_obj._gamma * varia-
bles_obj.FC_array[i][j] * varia-
bles_obj.delta_CP_array[i][
j] + con-
stants_obj._lambda * \
                varia-
bles_obj.C_r_array[i][j] + con-
stants_obj._epsilon
                # Запишем вычисленное
                значение в элемент массива
                variables_obj.C_x_ar-
ray[i][j] = value

            return variables_obj.C_x_array

    def get_result(self) ->
List[List[float]]:
    """
    Вернуть результат
    :return: массив
    """
    return self.C_x_result

```

Аналогично класс *L_x_calculator* выполняет вычисления для массива L_x с помощью метода `_calculate_L_x_array` и предоставляет результаты через метод `get_result`.

```

class L_x_calculator:
    """
    Класс подсчета массива L_x
    """
    def __init__(self, constants_obj:
Constants, variables_obj: Variables):
        self.L_x_result = self._calcu-

```

```

late_L_x_array(constants_obj, variables_obj)

def _calculate_L_x_array(self, constants_obj: Constants, variables_obj: Variables):
    """
    Посчитать массив L_x
    """
    for i in range(variables_obj.n):
        for j in range(variables_obj.m):
            # Вычисляем значение для
            # текущего элемента массива
            value = variables_obj.delta_X_array[i][j] * variables_obj.Q_f_array[i][j] * variables_obj.lp ** (
                -1) + constants_obj._omega * \
                variables_obj.N_array[i][j] + constants_obj._epsilon
            # Запишем вычисленное
            # значение в элемент массива
            variables_obj.L_x_array[i][j] = value

    return variables_obj.L_x_array

def get_result(self) -> List[List[float]]:
    """
    Вернуть результат
    :return: массив
    """
    return self.L_x_result

```

Класс *Q_x_calculator* отвечает за вычисление массива Q_x, используя метод *_calculate_Q_x_array* и предоставляя результаты через метод *get_result*.

```

class Q_x_calculator:
    """
    Класс подсчёта массива Q_x
    """
    def __init__(self, constants_obj: Constants, variables_obj: Variables):
        self.Q_x_result = self._calculate_Q_x_array(constants_obj, variables_obj)

    def _calculate_Q_x_array(self, constants_obj: Constants, variables_obj: Variables):
        """
        Посчитать массив Q_x
        """
        for i in range(variables_obj.n):
            for j in range(variables_obj.m):
                # Вычисляем значение для
                # текущего элемента массива
                value = con-

```

```

stants_obj._lambda * variables_obj.C_x_array[i][j] ** constants_obj._lambda + constants_obj._teta * variables_obj.L_x_array[i][j] ** variables_obj.delta_lp
        # Запишем вычисленное
        # значение в элемент массива
        variables_obj.Q_x_array[i][j] = value

    return variables_obj.L_x_array
def get_result(self) -> List[List[float]]:
    """
    Вернуть результат
    :return: массив
    """
    return self.Q_x_result

```

Для координации выполнения всех расчетов служит класс *FormulaSolve*. Он управляет процессом решения целевой функции модели структуры импортозамещения. В его методы входят *_solve*, который выполняет необходимые вычисления, и методы *get_C_x*, *get_L_x*, *get_Q_x* для получения результатов соответствующих массивов.

```

class FormulaSolve:
    """
    Класс решения целевой функции модели
    структуры импортозамещения
    """
    def __init__(self, constants_obj: Constants, variables_obj: Variables):
        self.C_x_obj = None
        self.L_x_obj = None
        self.Q_x_obj = None

        self._solve(constants_obj, variables_obj)

    def _solve(self, constants_obj: Constants, variables_obj: Variables):
        """
        Посчитать целевую функцию
        """
        self.C_x_obj = C_x_calculator(constants_obj, variables_obj)
        self.L_x_obj = L_x_calculator(constants_obj, variables_obj)
        self.Q_x_obj = Q_x_calculator(constants_obj, variables_obj)

    def get_C_x(self) -> List[List[float]]:
        """
        Получить массив C_x
        :return: двумерный массив
        """

```

```

return self.C_x_obj.get_result()

def get_L_x(self) ->
List[List[float]]:
    """
    Получить массив L_x
    :return: двумерный массив
    """
    return self.L_x_obj.get_result()

def get_Q_x(self) ->
List[List[float]]:
    """
    Получить массив Q_x
    :return: двумерный массив
    """
    return self.Q_x_obj.get_result()
    
```

Взаимодействие классов. Пользователь задает через интерфейс значения констант и переменных, которые затем передаются в соответствующие классы Constants и Variables. После этого классы C_x_calculator, L_x_calculator и Q_x_calculator выполняют необходимые расчеты. Результаты расчетов координируются и сохраняются классом FormulaSolve. Взаимодействие классов отражено на рисунке 1.

Формат команд и процедур. Команды и процедуры в приложении выполняются через методы классов. Пользователь взаимодействует с приложением через графический интерфейс, который вызывает соответствующие методы для выполнения расчетов и отображения результатов. Работа алгоритма в виде блок-схемы представлена на рисунке 2.

Работа СУБД. Используемая в приложении БД состоит из двух основных таблиц – variables и results. Таблица variables хранит переменные, используемые в расчетах, такие как количество отраслей (n), количество продуктов (m), массивы стоимости основного капитала (FC_array), изменения фондоотдачи (delta_CP_array), изменения фондоотдачи на уровне отрасли (C_r_array), изменения доли отечественного продукта (delta_X_array), фактический объем выпуска продукции (Q_f_array), количество работников (N_array), производительность труда (lp) и изменение производительности труда (delta_lp). Каждая запись в таблице variables идентифицируется уникальным идентификатором.

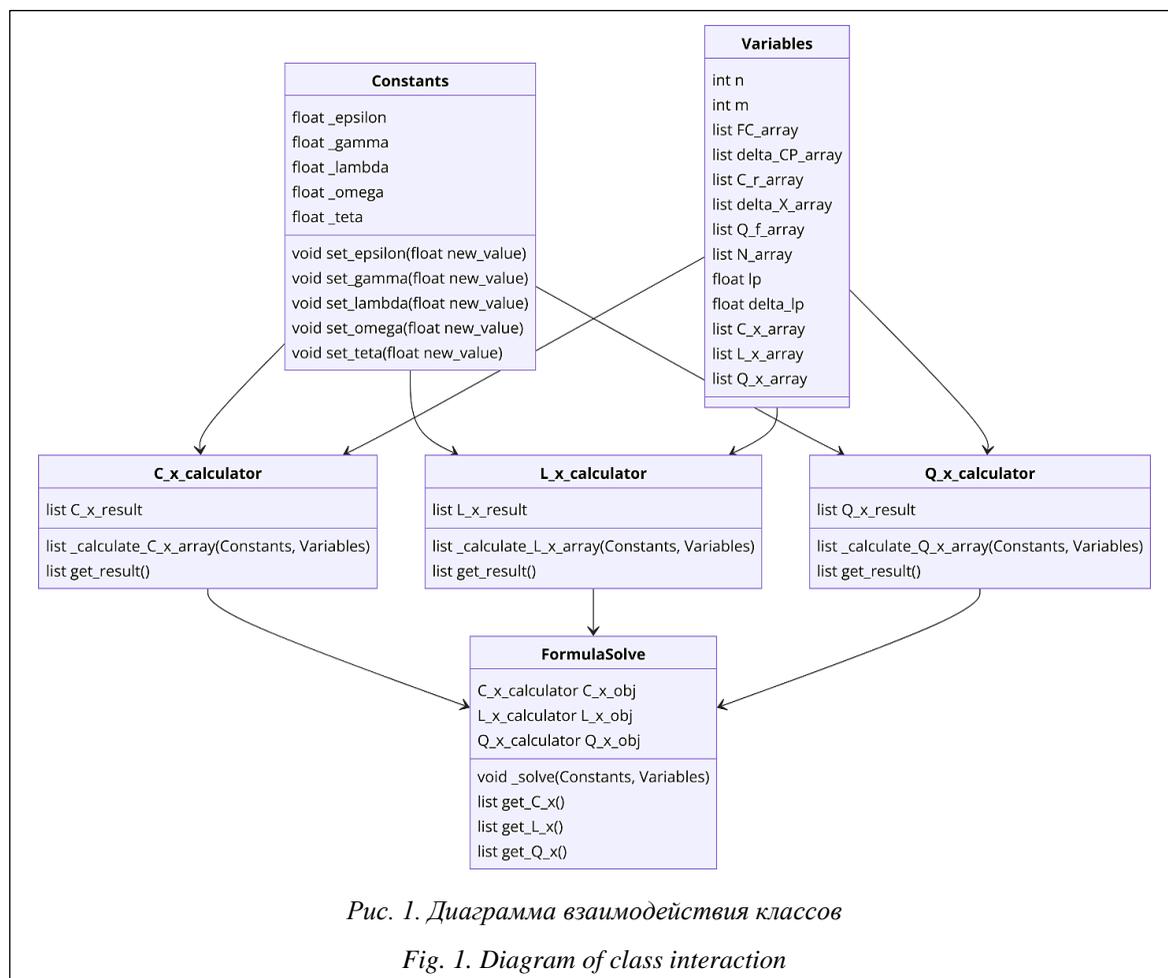
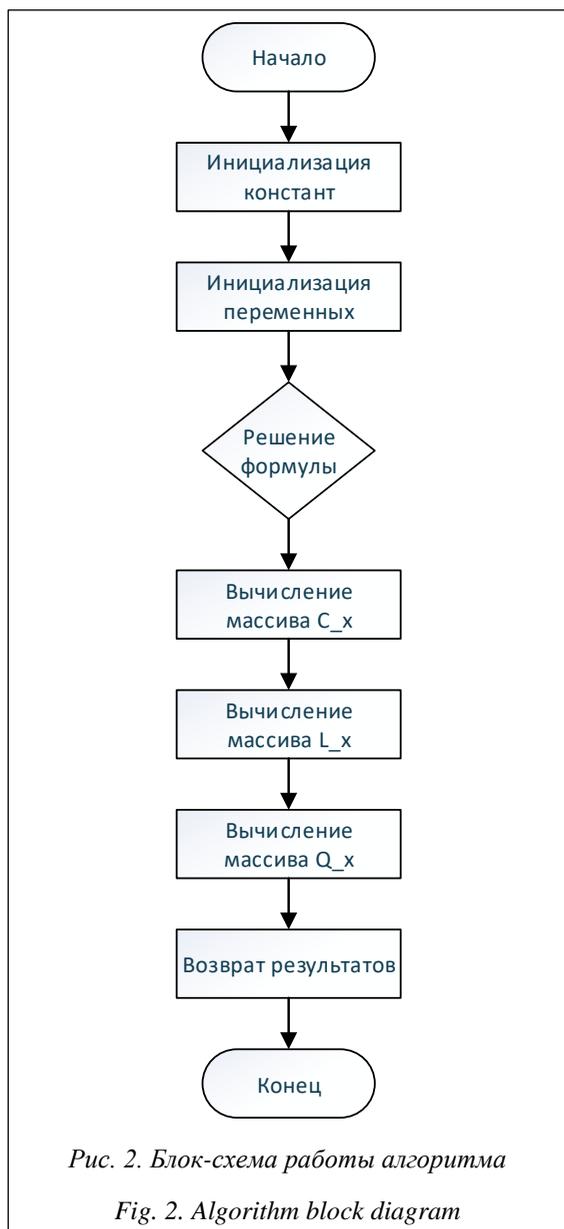


Рис. 1. Диаграмма взаимодействия классов

Fig. 1. Diagram of class interaction



ром (uid) (<http://www.swsys.ru/uploaded/image/2024-3/15.jpg>).

Таблица results хранит результаты расчетов, такие как массивы C_x , L_x и Q_x . Эти результаты связываются с соответствующими переменными через uid (<http://www.swsys.ru/uploaded/image/2024-3/16.jpg>).

Связь между таблицами обеспечивается использованием uid, что позволяет отслеживать все вычисления, связанные с конкретным набором переменных.

Результаты разработки

Основная функция приложения. Основная функция приложения заключается в мо-

делировании структуры импортозамещения в условиях внешних шоков. Это позволяет пользователям оценивать воздействие различных факторов на производственные процессы и планировать необходимые изменения для повышения эффективности импортозамещения.

Алгоритм работы приложения. Работа приложения начинается с ввода данных пользователем через графический интерфейс. Пользователь вводит значения переменных и констант, таких как количество отраслей, количество продуктов, значения массивов стоимости основного капитала, изменения фондоотдачи, фактический объем выпуска продукции и количество работников. Введенные данные сохраняются в таблице variables БД.

После ввода всех необходимых данных пользователь нажимает кнопку для запуска расчетов. Программа выполняет расчеты, используя методы классов $C_x_calculator$, $L_x_calculator$ и $Q_x_calculator$. Результаты расчетов сохраняются в таблице results БД и связываются с соответствующими переменными через uid.

Результаты расчетов отображаются на экране в виде таблиц и графиков. Пользователь может просмотреть массивы C_x , L_x и Q_x , а также другие производные данные. Интерфейс предоставляет возможность визуализации данных для удобства анализа.

Графический интерфейс. Графический интерфейс приложения состоит из нескольких основных компонентов, обеспечивающих удобный ввод данных, запуск расчетов и просмотр результатов (<http://www.swsys.ru/uploaded/image/2024-3/17.jpg>).

Пользователь вводит значения переменных и констант через соответствующие поля ввода. Например, он может ввести количество отраслей (n) и количество продуктов (m), а также значения массивов стоимости основного капитала, изменения фондоотдачи, изменения доли отечественного продукта, фактического объема выпуска продукции и количества работников. Для каждого массива предусмотрены кнопки, открывающие диалоговые окна для ввода значений (<http://www.swsys.ru/uploaded/image/2024-3/21.jpg>).

После ввода всех данных пользователь подтверждает ввод и нажимает кнопку для запуска расчетов. Результаты расчетов отображаются на экране в виде таблиц и графиков, что позволяет пользователю анализировать данные и принимать решения на основе полученной информации.

Таблицы отображают результаты расчетов массивов C_x , L_x и Q_x , а графики визуально представляют эти данные для удобства анализа.

Заключение

Математическое моделирование структуры импортозамещения на основе определения оптимальных пропорций двух факторов производства – труда и капитала, а также их эффективности (производительности и фондоотдачи) дает возможность преодолеть те ограничения моделей импортозамещения, учитывающих динамические характеристики производства (выпуск, затраты, субсидии и пр.), которые усиливаются в периоды действия внешних шоков, таких как санкционные ограничения. Для этого в статье предложен подход к созданию программного средства, предназначенного для моделирования инвестиций и занятости, выпуска в импортозамещающих отраслях, включающего ряд основных классов (в том числе

класс решения целевой функции модели структуры импортозамещения), диаграмму, отражающую порядок их взаимодействия, алгоритм вызова соответствующих методов для выполнения расчетов и отображения результатов, а также систему управления БД.

Разработанный алгоритм работы приложения предусматривает ввод данных – переменных и констант – через графический интерфейс, запуск расчетов, сохранение результатов и вывод их на экран в виде графиков и таблиц. С учетом этого предложенное программное средство может быть использовано в процессе создания отраслевых программ импортозамещения и их корректировки в условиях усиления внешних шоков, а также для гармонизации федеральных и региональных направлений промышленной политики.

Дальнейшее направление исследования связано с развитием СУБД, отражающих результаты воздействия внешних шоков на структуру и масштаб импортозамещения.

Список литературы

1. Уразаева Л.Ю., Галимов И.А. Математическое моделирование импортозамещения // Вестн. евразийской науки. 2022. Т. 14. № 3. С. 1–9.
2. Стрижкова Л.А. Использование таблиц «затраты–выпуск» при оценке зависимости российской экономики от импорта и процессов импортозамещения // Вопросы статистики. 2016. № 5. С. 3–22.
3. Моисеев Н.А., Внуков И.А., Сокерин П.О. Оценка эффектов различных вариантов импортозамещения методом «затраты–выпуск» на примере Российской Федерации // Экономика и математические методы. 2023. Т. 59. № 1. С. 30–47. doi: 10.31857/S042473880024869-4.
4. Кутернин М.И. Моделирование комплексного процесса импортозамещения и модернизации экономики страны при управлении через ее естественные монополии // Вестн. университета. 2017. № 9. С. 39–44.
5. Палаш С.В. Моделирование экономических эффектов импортозамещения в обрабатывающей промышленности Российской Федерации // Науч.-технич. ведомости СПбГПУ. Экономические науки. 2019. Т. 12. № 1. С. 59–69.
6. Clements K.W., Mariano M.J.M., Verikios G. Foreign-domestic substitution, import penetration and CGE modeling. *Applied Economics*, 2021, vol. 53, no. 35, pp. 4080–4099. doi: 10.1080/00036846.2021.1897072.
7. Енина Е.С. Модель вовлечения в процесс импортозамещения: маркетинговый подход // Экономика и управление. 2022. Т. 28. № 9. С. 905–911. doi: 10.35854/1998-1627-2022-9-905-911.
8. Nurkomariyah S., Vierke I.M. Import substitution in circular economy perspective: A study on textile and apparel industry. *IJSSHR*, 2023, vol. 6, no. 12, pp. 12–33. doi: 10.47191/ijsshr/v6-i12-33.
9. Cherif R., Hasanov F. The pitfalls of protectionism: Import substitution vs. export-oriented industrial policy. *ICT*, 2024, vol. 24, art. 14. doi: 10.1007/s10842-024-00414-9.
10. Кордюков Р.Ю., Допира Р.В., Иванова А.В., Абу-Абед Ф.Н., Мартынов Д.В. Модель и алгоритмизация оптимизационной задачи о назначениях в условиях дополнительных ограничений // Программные продукты и системы. 2016. Т. 29. № 2. С. 16–22. doi: 10.15827/0236-235X.114.016-022.
11. Моисеев Н.А., Ахмадеев Б.А. Алгоритм оценки импортозамещения на основе таблиц затрат–выпуска // Вестн. РЭА им. Г.В. Плеханова. 2021. № 3. С. 117–129. doi: 10.21686/2413-2829-2021-3-117-129.
12. Абу-Абед Ф.Н., Жиронкин С.А. Моделирование структуры импортозамещения на базе модели системы оптимального распределения // Программные продукты и системы. 2023. Т. 36. № 4. С. 644–653. doi: 10.15827/0236-235X.144.644-653.
13. Саидзода И.М., Ризоев С.С. Компьютерное исследование производственной функции Кобба–Дугласа с использованием электронной таблицы MS Excel // Вестн. ТГПУ. 2022. № 2. С. 80–85.
14. Сазонов А.А., Сазонова М.В. Выявление свойств экзогенности в моделях прогнозирования экономического роста // Фундаментальные и прикладные исследования: проблемы и результаты. 2016. № 24. С. 176–180.
15. Коротких О. Межстрановая BVAR-модель внешнего сектора // Деньги и кредит. 2020. Т. 79. № 4. С. 98–112. doi: 10.31477/tjmf.202004.98.

Components and functionality of a software tool for modelling an import substitution structure

Fares N. Abu-Abed ¹✉, Sergey A. Zhironkin ²

¹Tver State Technical University, Tver, 170026, Russian Federation

²National Research Tomsk Polytechnic University, Tomsk, 634050, Russian Federation

For citation

Abu-Abed, F.N., Zhironkin, S.A. (2024) 'Components and functionality of a software tool for modelling an import substitution structure', *Software & Systems*, 37(3), pp. 402–411 (in Russ.). doi: 10.15827/0236-235X.142.402-411

Article info

Received: 13.05.2024

After revision: 10.06.2024

Accepted: 17.06.2024

Abstract. The subject of this research is creating the main components of a tool for modelling an import substitution structure in the Russian economy and determining its functionality. The research is relevant due to the lack of software analogues and developments in the domestic IT market to solve the task. The authors apply the methods of system analysis, simulation modelling, visual design and object-oriented programming as research methods. Modelling the output of products for the domestic market requires finding the optimal combination of the volume values of production factors (capital and labour) and their efficiency. For this purpose, it is necessary to create a software tool that reflects a mathematical model of import substitution based on its target function and links various classes, databases and user interface. This paper summarizes the main classes, each of which plays a certain role in calculating and modelling of import substitution, explores their interaction presented as a diagram. The developed application executes commands and procedures through class methods; there is an algorithm of user interaction with the application. The proposed application database includes tables of variables necessary for determining the target function values (the number of industries and products, arrays of the cost of fixed capital and the number of employees, changes in stock returns and labour productivity, etc.) and results (the desired sectoral volumes of investment, employment and output). The application interface allows the user to calculate the values of the target function of an import substitution structure by using buttons that open dialogue windows for entering values and starting calculations. The scientific novelty consists in obtaining results that allow optimizing the solution of an important problem in the context of national economy regulation – in modelling an import substitution structure under the conditions of increasing external shocks and technological constraints.

Keywords: structure, mathematical model, software tool, classes, user interface, commands and procedures, database, import substitution

Acknowledgements. The study was supported by the Russian Science Foundation grant no. 23-28-01423, <https://rscf.ru/project/23-28-01423/>

References

1. Urazaeva, L.Yu., Galimov, I.A. (2022) 'Mathematical modeling of import substitution', *The Eurasian Scientific J.*, 14(3), pp. 1–9 (in Russ.).
2. Strizhkova, L.A. (2016) 'Using "input–output" tables in estimating the dependence of Russian economy on import and import substitution processes', *Bull. of Statistics*, (5), pp. 3–22 (in Russ.).
3. Moiseev, N.A., Vnukov, I.A., Sokerin, P.O. (2023) 'Evaluation of the effects from various ways of import substitution "input–output" method: Evidence from the Russian Federation', *Economics and Mathematical Methods*, 59(1), pp. 30–47 (in Russ.). doi: 10.31857/S042473880024869-4.
4. Kuternin, M.I. (2017) 'Modeling of the import substitution and national economy modernization complex process with a management through the natural monopolies of the country', *Vestn. Univ.*, (9), pp. 39–44 (in Russ.).
5. Palash, S.V. (2019) 'Modeling the economic effects of import substitution in Russian manufacturing industry', *SPbSTU J. Economics*, 12(1), pp. 59–69 (in Russ.).
6. Clements, K.W., Mariano, M.J.M., Verikios, G. (2021) 'Foreign-domestic substitution, import penetration and CGE modelling', *Applied Economics*, 53(35), pp. 4080–4099. doi: 10.1080/00036846.2021.1897072.
7. Enina, E. (2022) 'Model of involvement in import substitution: A marketing approach', *Economics and Management*, 28(9), pp. 905–911 (in Russ.). doi: 10.35854/1998-1627-2022-9-905-911.
8. Nurkomariyah, S., Vierke, I.M. (2023) 'Import substitution in circular economy perspective: A study on textile and apparel industry', *IJSSHR*, 6(12), pp. 12–33. doi: 10.47191/ijsshr/v6-i12-33.
9. Cherif, R., Hasanov, F. (2024) 'The pitfalls of protectionism: Import substitution vs. export-oriented industrial policy', *JICT*, 24, art. 14. doi: 10.1007/s10842-024-00414-9.
10. Kordyukov, R.Yu., Dopira, R.V., Ivanova, A.V., Abu-Abed, F.N., Martynov, D.V. (2016) 'A model and algorithmization of the assignment problem under additional constraints', *Software & Systems*, 29(2), pp. 16–22 (in Russ.). doi: 10.15827/0236-235X.114.016-022.
11. Moiseev, N.A., Akhmadeev, B.A. (2021) 'Algorithm of estimating import-substitution based on costs – output tables', *Bull. of the Russian Plekhanov University of Economics*, 3(117), pp. 117–129 (in Russ.). doi: 10.21686/2413-2829-2021-3-117-129.

12. Abu-Abed, F.N., Zhironkin, S.A. (2023) 'Russian import-substitution structure based on an optimal distribution system model', *Software & Systems*, 36(4), pp. 644–653 (in Russ.). doi: 10.15827/0236-235X.144.644-653.
13. Saidzoda, I.M., Rizoiev, S.S. (2022) 'Computer study of Cobb–Douglas production function using MS excel spreadsheet', *Herald TSPU*, (2), pp. 80–85 (in Russ.).
14. Sazonov, A.A., Sazonova, M.V. (2016) 'Identification of exogeneity properties in economic growth forecasting models', *Fundamental and Applied Research: Problems and Results*, (24), pp. 176–180 (in Russ.).
15. Korotkikh, O. (2020) 'A multi-country BVAR model for the external sector', *Russ. J. of Money and Finance*, 79(4), pp. 98–112 (in Russ.). doi: 10.31477/rjmf.202004.98.

Авторы

Абу-Абед Фарес Надимович¹, к.т.н.,
доцент, декан, aafares@mail.ru

Жиронкин Сергей Александрович², д.э.н.,
профессор, zhironkin@tpu.ru

¹ Тверской государственный технический
университет, г. Тверь, 170026, Россия

² Национальный исследовательский Томский
политехнический университет,
г. Томск, 634050, Россия

Authors

Fares N. Abu-Abed¹, Cand. of Sci. (Engineering),
Associate Professor, Dean, aafares@mail.ru

Sergey A. Zhironkin², Dr.Sci. (Economics),
Professor, zhironkin@tpu.ru

¹ Tver State Technical University,
Tver, 170026, Russian Federation

² National Research Tomsk Polytechnic University,
Tomsk, 634050,
Russian Federation

Компьютерное моделирование сложных взаимодействий отдельных элементов гранулированных материалов с применением функционально заданных объектов

С.И. Вяткин ¹✉, Б.С. Долговесов ¹

¹ Институт автоматизации и электрометрии СО РАН, г. Новосибирск, 630090, Россия

Ссылка для цитирования

Вяткин С.И., Долговесов Б.С. Компьютерное моделирование сложных взаимодействий отдельных элементов гранулированных материалов с применением функционально заданных объектов // Программные продукты и системы. 2024. Т. 37. № 3. С. 412–420. doi: 10.15827/0236-235X.142.412-420

Информация о статье

Группа специальностей ВАК: 1.2.2

Поступила в редакцию: 19.01.2024

После доработки: 11.03.2024

Принята к публикации: 19.03.2024

Аннотация. В статье предложен метод контактной динамики для гранулированных материалов с применением функционально заданных объектов. Рассмотрена задача моделирования сложных взаимодействий большого количества отдельных элементов этих материалов с использованием информации о массе, местоположении и контактах соприкасающихся поверхностей для получения линейной сложности по общему числу контактов, обнаруженных на каждой итерации. Предложенный метод контактной динамики основан на численном моделировании с предоставлением доступа ко всем переменным, на негладкости в пространстве времени и силы, то есть в пространстве конфигурационных переменных допустимым множеством является область, ограниченная набором поверхностей. В случае столкновения функция времени, представляющая скорость, будет прерывистой. Трение определяется соотношением между силой реакции и локальной относительной скоростью. Предложен метод вычисления скорости для моделирования больших наборов твердых тел. В качестве отдельных элементов гранулированных материалов используются функционально заданные объекты, в частности, эллипсоиды. В методе отсутствуют повторные парные сравнения между объектами, что обуславливает линейную сложность по общему числу контактов, обнаруженных на каждой итерации. Разработана модель трения в конфигурационном пространстве твердых тел, основанная на принципе максимального рассеивания. В результате она объединяет трение качения и скольжения, обеспечивая ожидаемые характеристики, с вычислением допустимой скорости и реакции на трение для всех объектов. Работоспособность метода подтверждена компьютерной симуляцией. Эксперименты показали, что в микроскопическом масштабе времени столкновение может разделиться на несколько отдельных эпизодов контакта. Это приводит к тому, что исход столкновения значительно зависит от формы соответствующих тел. В случае, когда отдельные элементы гранулированных материалов имеют вид эллипсоидов, результат оказывается более приемлемым. Предложенный метод контактной динамики может быть полезен для вычисления квазистатических эволюций гранулированных материалов, например, в машиностроении, строительстве, при численной обработке с помощью конечных элементов динамических и квазистатических задач, связанных с деформируемыми телами.

Ключевые слова: функционально заданные объекты, компьютерное моделирование, динамическое взаимодействие, контакт, трение

Благодарности. Работа выполнена в ИАиЭ СО РАН при поддержке Минобрнауки в рамках работ по Госзаданию № 124041700102-4

Введение. Модели твердого тела широко используются, в частности, в компьютерных играх, художественных фильмах, в технике, промышленности, военном деле. Динамическое взаимодействие твердых тел применяется в наноструктурированных объектах, где даже небольшое изменение дисперсности или состава компонентов нанокompозита приводит к существенному изменению макроскопических физико-механических характеристик.

Моделирование динамики множества тел требует точного определения контакта, и из-за небольших временных интервалов необходимы большие вычислительные затраты. Это является фундаментальной задачей во многих областях. Методы моделирования динамиче-

ских систем с событиями контактного воздействия являются важным инструментом и популярной стратегией для моделирования твердых тел. В то же время с помощью таких методов приходится решать ряд хорошо известных проблем, включая пересечения, нестабильность, неточности и низкие эксплуатационные характеристики, которые возрастают по мере усложнения контактной задачи. Контактная динамика – это численный метод для вычисления динамического движения совокупности тел, подверженных односторонним ограничениям невзаимопроникновения, с учетом трения в случае контакта, а также скачков скорости, которые могут возникнуть в результате столкновений.

Одной из значительных проблем при моделировании твердых тел является присущая динамике негладкость [1]. Это происходит, когда твердые тела мгновенно меняют свою скорость при столкновениях с другими поверхностями. Жесткие ограничения на твердое тело, связанные с отсутствием взаимопроникновения и незначительной деформацией, затрудняют определение контакта, который сильно зависит от изменений конфигурации и положения [2, 3].

Основная цель данного исследования – разработка алгоритмов, регулирующих поведение функционально заданных объектов, которые всегда дают реалистичные и выполнимые результаты с максимально возможной скоростью.

Задачи – оценка влияния трения, необходимого для реалистичности физического моделирования, а также разработка быстрого и эффективного симулятора функционально заданных объектов. Геометрический анализ осуществимости импульсов трения важен при построении модели трения.

Новизна работы заключается в способе определения трения, основанного на объединении трения качения и скольжения, а также в вычислении допустимой скорости и реакции на трение для всех объектов. Предложен простой и эффективный метод реализации ограничений. При моделировании основной упор сделан на физическую точность.

В работе рассмотрен симулятор функционально заданных объектов, который может быть интегрирован с различными системами визуализации, анимации и обнаружения столкновений. Симулятор включает множество сложных режимов, таких как перекачивание, скольжение, укладка, переворачивание и распространение ударов, а также соответствующие переходы между ними. Это дает реалистичное поведение для твердого тела, находящегося в контакте с множеством фиксированных ограничений.

Для настройки с несколькими объектами каждое тело рассматривается как независимая система, в которой все остальные движущиеся тела накладывают аффинные ограничения на уровне скорости. Это эффективно, поскольку включается контактная информация первого порядка для каждого тела. Таким образом, получаются скорости, обеспечивающие непроникающее поведение, по крайней мере, локально на следующем этапе. При этом вводится приближительное сохранение импульса для многих одновременных контактов, для чего вычисля-

ется аффинное ограничение для каждого контакта, использующее скорость контакта в момент максимального сжатия. В совокупности это реализуется алгоритмом моделирования множества тел, который является линейным по общему количеству контактов, обрабатываемых для всех объектов на каждом шаге. Внедрено несколько практических оптимизаций для дальнейшего повышения производительности – так сокращается количество обрабатываемых контактов и для увеличения скорости, и для упрощения внедрения. Также отфильтровываются многие первоначально обнаруженные контакты, игнорируются удовлетворяющие ограничения, то есть контакты, от которых объект удаляется с достаточной скоростью.

Динамика твердого тела находит применение, например, в машиностроении, робототехнике, нанотехнологии, аэрокосмической промышленности, молекулярной динамике. Она позволяет предсказывать и объяснять движение и поведение твердых тел в различных условиях и с разными внешними воздействиями.

В статье [4] рассматривается алгоритм идентификации инерционных параметров и контактных сил, связанных с объектом. Используется известная формулировка взаимодополняемости для разрешения контакта, чтобы установить взаимосвязь замкнутой формы между инерционными параметрами, контактными силами и наблюдаемыми движениями. Взаимодополняемость является полезным средством анализа контактов состояния.

Современные популярные подходы к решению проблемы множественных контактов с трением зависят от порядка столкновений и применяют решение линейной задачи о дополнительности. В [5] описана аппаратная структура параллельной обработки для решения линейной задачи о дополнительности (LCP).

В статье [6] предлагается новый метод для решения линейной задачи о дополнительности при интерактивном моделировании твердого тела, основанный на методе Гаусса–Зайделя.

Возникающие контактные проблемы можно также решить, используя множители Лагранжа для представления неизвестной величины сил реагирования. В отличие от решения линейной задачи о дополнительности в работе [7] применяется прямое решение задачи смешанной линейной дополнительности, не требующее построения матрицы LCP.

Большая часть трудностей при разрешении контактов возникает из-за неопределенности, связанной с установлением того, какие точки

контакта действительно активны, то есть способствуют реагированию в данный момент времени. Для проверки, является ли ограничение активным или пассивным, требуется функция расстояния между частями тел. В статье [8] дается строгое определение функции расстояния и выводятся ее некоторые свойства.

В [9] представлен метод обработки фрикционного контакта для жестких и деформируемых объектов. Метод позволяет решить основные трудности, присущие негладкому характеру фрикционного контакта.

В статье [10] описывается конечно-элементная модель при моделировании твердого тела. После определения столкновения между объектами генерируется ответный импульс для распространения на другие объекты, уже находящиеся в контакте со сталкивающимися телами.

В [11] представлен неявный пошаговый алгоритм для динамики твердого тела с контактом и трением, который гарантирует отсутствие пересечений конфигураций на каждом временном шаге. Хотя метод и менее эффективен, чем альтернативные подходы, с его помощью можно надежно обрабатывать широкий спектр сложных сцен.

В обзорной работе [12] рассматриваются основные методы моделирования динамических систем с событиями контактного воздействия. Представлено описание контактной динамики, а также сделан обзор современного состояния основных аспектов, связанных с дисциплиной контактной динамики. Отмечены недостатки существующих методов и перспективы будущих исследований в области контактной механики в динамике множества тел.

К сожалению, нелинейное поведение трения усложняет проблему контакта твердого тела. Традиционно контакт разрешался с использованием сложных схем и методов обнаружения столкновений с учетом того, что он происходит между полигонами. Хотя полигональные поверхности являются эффективным средством представления твердых тел, им не хватает некоторых внутренних свойств, важных для разрешения контакта. Как правило, поверхности базы полигонов не снабжены внутренним и внешним разделением или плавным полем расстояний вблизи поверхности.

Авторы данного исследования предлагают новый метод решения проблемы фрикционных контактов с использованием функций возмущения. В работе представлен алгоритм определения скорости для моделирования больших

наборов функциональных объектов. При этом в методе отсутствуют повторные попарные сравнения между объектами, что дает линейную сложность по общему числу контактов, вычисленных на каждой итерации.

Описание метода

Функционально заданный объект. Объект описывается с помощью функций возмущения и базовой квадратики [13]:

$$F'(x, y, z) = F(x, y, z) + \sum_{i=1}^N R_i(x, y, z), \quad (1)$$

где $F'(x, y, z)$ – функционально заданный объект; $F(x, y, z)$ – базовая квадратика; $i = 1, \dots, N$ – количество функций возмущения; $R_i(x, y, z)$ – возмущение, при этом

$$R_i(x, y, z) = \begin{cases} Q_i^3(x, y, z), & \text{if } Q_i(x, y, z) \geq 0, \\ 0, & \text{if } Q_i(x, y, z) < 0, \end{cases} \quad (2)$$

где $Q_i(x, y, z)$ – возмущающая квадратика.

Динамическое взаимодействие функциональных объектов. При отсутствии контактов динамика функционального объекта i в его собственной системе координат определяется как

$$M^i \dot{v} = {}^i f + {}^i c, \quad (3)$$

где ${}^i v$ – поворот функционального объекта i ; ${}^i f$ – суммарная сила, состоящая из неконтактных видов, таких как гравитация; M – тензор инерции объекта; ${}^i c$ – слагаемое Кориолиса (${}^i c = ({}^i v)^T M^i \dot{v}$).

В случае наличия контакта скорость функционального объекта до обнаружения столкновения составит

$${}^i v = {}^i v' + s({}^i f + {}^i c), \quad (4)$$

а скорость после обработки всех контактов –

$${}^i v = {}^i v'^{+1} + {}^i \xi, \quad (5)$$

где s – размер шага; ${}^i \xi$ – изменение пространственной скорости, вызванное активными контактными ограничениями.

В статье [14] описан алгоритм обнаружения столкновений функционально заданных объектов, где применен менее универсальный, но более быстрый способ определения контакта из [15].

Анализируется характеристический многочлен, связанный с эллипсоидами и квадратиками, для обнаружения контакта. Условие определения столкновения проверяется непосредственно по их параметрам. Контакт определяется с помощью дискриминантов характеристического многочлена. Относительные положения классифицируются по знаку коэффициентов этого многочлена.

Чтобы определить контакты, каждый функциональный объект обрабатывается один раз за шаг, при этом каждый шаг является линейным по общему количеству контактов. Эти границы поддерживаются путем рассмотрения каждого объекта как независимой системы, используя движение других объектов для создания ограничений уровня скорости.

Поскольку другие объекты двигаются, следовательно, накладываются ограничения на перемещение, поэтому добавляется смещение к каждому ограничению контакта на уровне скорости, которое учитывает скорость контактирующего объекта.

Чтобы компенсировать изменение ограничения в течение каждого интервала, каждое ограничение контакта компенсируется относительной скоростью. Каждое ограничение уровня скорости будет определять аффинное полупространство.

Для функционального объекта i обозначим пространственную скорость, с которой он имеет общий контакт c , через ${}^i v_c$. Эта скорость преобразуется в систему координат объекта i . Величина скорости контакта вдоль нормали ограничения будет

$$m_c = {}^i n_c^T M {}^i v_c. \quad (6)$$

Затем генерируем аффинное ограничение скорости для объекта i , требуя, чтобы его скорость не только была в допустимом направлении, но и имела достаточно большую составляющую вдоль нормали контакта (рис. 1):

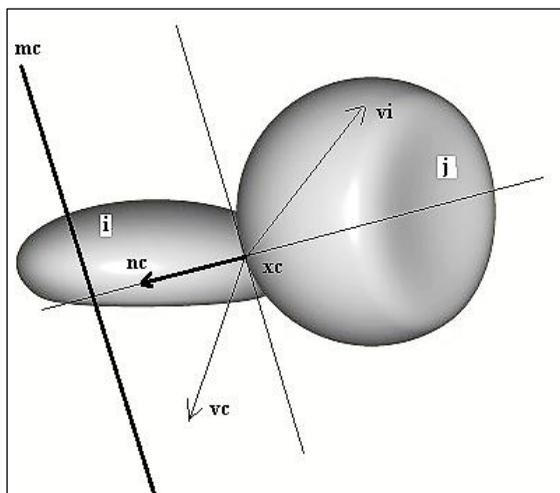


Рис. 1. Функционально заданный объект j , накладывающий ограничение на перемещение объекта i

Fig. 1. Functionally defined object j that imposes a restriction on the movement of object i

$${}^i n_c^T M {}^i v_c - m_c \geq 0. \quad (7)$$

Каждое такое ограничение, вычисленное по контакту c , определяет допустимое аффинное полупространство.

В общей точке контакта \bar{x}_c с нормалью \bar{n}_c задается скорость v_c . Скорость объекта i после столкновения должна быть больше, чем m_c при проецировании на \bar{n}_c (рис. 1).

Сохранение импульса позволяет получить скорость точки контакта вдоль нормали к контакту в момент максимального сжатия. Сначала вычисляется эффективная масса объекта i с инерцией, определяемая силой, действующей вдоль нормали контакта \bar{n}_c , в точке контакта \bar{x}_c .

Пусть v_{nc} – скорость вдоль контактной нормали \bar{n}_c , mv_{nc} – соответствующий импульс. Во время максимального сжатия между объектами i и j сохранение импульса подразумевает, что существует скорость вдоль нормали, которая вычисляется следующим образом:

$$\bar{v}_{nca} = \frac{(mv_{nci} + mv_{ncj})}{(m_i + m_j)}, \quad (8)$$

где v_{nci} и v_{ncj} – скорости объектов i и j в точке соприкосновения, спроецированные вдоль \bar{n}_c . Данная скорость является средним значением этих двух скоростей, взвешенным по их соответствующим массам.

Модель трения, применяемая в настоящей работе, основана на импульсном методе, в частности, на алгоритме отслеживания импульсов энергии [16].

Чтобы контакты могли генерировать как нормальные, так и тангенциальные импульсы, запишем следующее уравнение:

$${}^i v^{t+1} - {}^i v^t = s({}^i f + {}^i c) + {}^i \xi + {}^i \lambda, \quad (9)$$

где ${}^i \xi$ – изменение пространственной скорости, вызванное активными контактными ограничениями; ${}^i \lambda$ – встроенный импульс трения, противодействующий скольжению вдоль касательных плоскостей ограничения в \mathcal{R}^3 .

Для окончательных результатов необходимо дополнительно ограничить скользящие повороты, которые будут генерироваться из единичных касательных векторов в \mathcal{R}^3 .

Каждому сгенерированному скользящему повороту определяем обобщенный коэффициент трения ψ_c . Это позволяет создавать характеристики трения, которые могут и быть анизотропными, и изменяться в зависимости от местоположения контакта. Также вычисляем крутящий момент ${}^i v^t$. Нормальный импульс ${}^i \xi$

и совокупность фрикционных коэффициентов ψ_c используются для определения импульса трения $i\lambda$.

Алгоритм вычислений. Пошагово выполняем алгоритм вычислений, используя фиксированный размер шага s :

- для каждого функционального объекта предварительно вычисляем базовую поверхность;

- генерируем матрицу инерции с фиксированной диагональю;

- вычисляем положение объекта в мировой системе координат;

- на каждом шаге вычисляем скорости $i v_i$ ((4), (5));

- определяем столкновения;

- вычисляем нормали и аффинные расстояния; поскольку вычисляется каждое контактное ограничение, можно немедленно отбросить те, которые строго выполняются (эти ограничения генерируются на ранней стадии конвейера, что позволяет эффективно снизить стоимость каждого шага);

- вычисляем активные контакты для дальнейших вычислений; эти детали отражают встроенную в подход взаимодополняемость на уровне скорости (контактные ограничения, когда удаляется объект, выполняются и поэтому не будут создавать нормальные и тангенциальные импульсы, в то время как все остальные ограничения потенциально активны и поэтому будут создавать импульсы пропорционально степени нарушения их ограничений);

- если контакт обнаружен, для каждого используем нормаль для равномерной выборки конечного набора касательной плоскости в точке контакта; эти выборки используются для аппроксимации набора всех возможных тангенциальных импульсов при контакте.

Алгоритм вычислений показан на рисунке 2.

Результаты

Производительность метода протестирована на процессоре 12 core 2.5 GHz Intel Xeon E5-2678 v3. Все численные методы реализованы с использованием C++ на CPU, выбран Eigen (https://eigen.tuxfamily.org/index.php?title=Main_Page) в качестве основной библиотеки линейной алгебры, включающей все решения для разреженных линейных систем. Для распараллеливания процессора используется Intel TBB, который позволяет упростить создание и развертывание многопоточных приложений (<https://github.com/oneapi-src/oneTBB>). С помощью Intel TBB не надо решать, на какой архи-



Рис. 2. Алгоритм динамического взаимодействия функционально заданных объектов

Fig. 2. Algorithm of dynamic interaction of functionally specified objects

тектуре и на какой платформе будет использоваться программа. Она берет на себя ответственность за работу с потоками.

Результаты тестов сведены в таблицу (для 1 000 эллипсоидов) и проиллюстрированы рисунками 3 и 4.

Каждый шаг реализации разбит на четыре этапа: предварительные вычисления, обнару-



Рис. 3. Пример моделирования 27 функционально заданных объектов с трением с использованием предлагаемого подхода

Fig. 3. Example of modeling 27 functionally defined objects with friction using the proposed approach

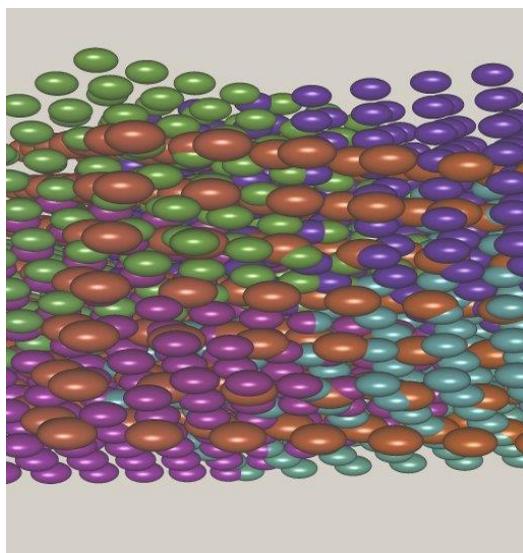


Рис. 4. Пример моделирования 1 000 эллипсоидов с трением с использованием предлагаемого подхода

Fig. 4. Example of modeling 1 000 ellipsoids with friction using the proposed approach

жение контакта, разрешение контакта и заключительные вычисления.

Как на предварительном, так и на заключительном этапах каждый объект посещается один раз и выполняется фиксированный объем работы, поэтому сложность для обоих этапов равна $O(n^2)$, где n – количество объектов, обрабатываемых симулятором.

Фаза обнаружения контакта отвечает за нахождение всех точек соприкосновения, возникающих между любыми парами объектов при моделировании.

Сложность представленного детектора столкновений с широкой фазой зависит от количества объектов, которые сталкиваются или близки к столкновению: $O(n)$ – наилучший случай, $O(n^2)$ – наихудший случай (в физически неправдоподобном случае, когда все n объектов находятся достаточно близко друг к другу).

Из тестов видно, что предлагаемая реализация линейна по количеству моделируемых объектов и по общему количеству точек соприкосновения, обнаруживаемых на каждом шаге.

Была запущена симуляция свободно падающих эллипсоидов в наклонную прозрачную трубу, которые высыпаются из нее на плоскую поверхность.

В таблице показаны статистические данные в ходе моделирования.

Следует отметить продолжающееся увеличение числа контактов после насыщения числа сталкивающихся эллипсоидов в течение времени, когда оседающие объекты собираются вместе.

Также построены графики времени, затраченного на каждый из этапов алгоритма (<http://www.swsys.ru/uploaded/image/2024-3/6.jpg>).

Жесткие ограничения на твердое тело, связанные с отсутствием взаимопроникновения и незначительной деформацией, затрудняют разрешение контакта и делают его высокочувствительным даже к незначительным изменениям конфигурации и положения.

Заключение

В работе представлен основанный на контакте подход к динамике функциональных объектов, который приводит к непротиворечивой теории и надежному симулятору линейного времени. Подход сформулирован с учетом геометрических параметров, что дает возможность локально обрабатывать множественные контакты с нелинейным трением. Это, в свою

Статистические данные и время выполнения тестовых примеров

Test case statistics and execution time

Параметр	1К шагов	2К шагов	3К шагов	4К шагов	5К шагов	6К шагов
Количество объектов	800	1 000	1 000	1 000	1 000	1 000
Количество контактных точек	10 104	51 247	201 467	252 378	312 147	354 245
Время вычислений столкновений, сек.	0,037	0,049	0,061	0,069	0,078	0,089

очередь, позволяет получить крупномасштабное правдоподобное моделирование функцио-

нально заданных объектов с использованием реализованного симулятора.

Список литературы

1. Wachs A., Uhlmann M., Derksen J., Huet D.P. Modeling of short-range interactions between both spherical and non-spherical rigid particles. In: Modeling approaches and computational methods for particle-laden turbulent flows, 2023, pp. 217–264. doi: 10.1016/B978-0-32-390133-8.00019-0.
2. Strecke M., Stueckler J. DiffSDFSim: Differentiable rigid-body dynamics with implicit shapes. Proc. Int. Conf. 3DV, 2021, pp. 96–105. doi: 10.1109/3DV53792.2021.00020.
3. Elandt R., Drumwright E., Sherman M.A., Ruina A. A pressure field model for fast, robust approximation of net contact force and moment between nominally rigid objects. Proc. IEEE/RSJ Int. Conf. IROS, 2019, pp. 8238–8245. doi: 10.1109/IROS40897.2019.8968548.
4. Fazeli N., Tedrake R., Rodriguez A. Identifiability analysis of planar rigid-body frictional contact. In: Robotics Research. SPAR, 2018, vol. 3, pp. 665–682. doi: 10.1007/978-3-319-60916-4_38.
5. Kim Y. The Hardware accelerated physics engine with operating parameters controlling the numerical error tolerance. IJAST, 2018, vol. 119, pp. 145–152. doi: 10.14257/ijast.2018.119.13.
6. Miyamoto S., Yamashita M. An improved convergence based on accelerated modulus-based Gauss–Seidel method for interactive rigid body simulations. SN Appl. Sci., 2021, vol. 3, art. 266. doi: 10.1007/s42452-021-04238-8.
7. Verschoor M., Jalba A. Efficient and accurate collision response for elastically deformable models. ACM Transactions on Graphics, 2019, vol. 38, no. 2, art. 17, pp. 1–20. doi: 10.1145/3209887.
8. Lidstrom P. Kinematics for unilateral constraints in multibody dynamics. MMS, 2017, vol. 22, no. 8, pp. 1654–1687. doi: 10.1177/1081286516642270.
9. Geilinger M., Hahn D., Zehnder J., Bacher M. ADD: Analytically differentiable dynamics for multi-body systems with frictional contact. ACM Transactions on Graphics, 2020, vol. 39, no. 6, art. 190, pp. 1–15. doi: 10.1145/3414685.3417766.
10. Coevoet E., Andrews S., Relles D., Kry P.G. Distant collision response in rigid body simulations. Comput. Graphics Forum, 2020, vol. 39, no. 8, pp. 113–122. doi: 10.1111/cgf.14106.
11. Ferguson Z., Li M., Schneider T., Gil-Ureta F. Intersection-free rigid body dynamics. ACM Transactions on Graphics, 2021, vol. 40, no. 4, art. 183, pp. 1–16. doi: 10.1145/3476576.3476773.
12. Flores P. Contact mechanics for dynamical systems: A comprehensive review. Multibody Syst. Dyn., 2022, vol. 54, pp. 127–177. doi: 10.1007/s11044-021-09803-y.
13. Вяткин С.И., Долговесов Б.С. Методы интерактивного моделирования и визуализации функционально заданных объектов для 3D-Web приложений // Автотметрия. 2022. Т. 58. № 1. С. 111–118. doi: 10.15372/AUT20220112.
14. Вяткин С.И., Долговесов Б.С., Корсун А.С. Определение столкновений функционально заданных объектов с рекурсивным делением объектного пространства // Автотметрия. 2003. Т. 39. № 6. С. 119–126.
15. Vazquez M.B., Pereira-Saez M.J., Rodriguez-Raposo A.B., Souto-Salorio M.J. Contact detection between a small ellipsoid and another quadric. Comput. Aided Geometric Design, 2022, vol. 98, art. 102136. doi: 10.1016/j.cagd.2022.102136.
16. Li Y., Asai M., Chandra B., Isshiki M. Energy-tracking impulse method for particle-discretized rigid-body simulations with frictional contact. Computational Particle Mechanics, 2020, vol. 8, pp. 237–258. doi: 10.1007/s40571-020-00326-5.

Computer modeling of complex interactions of granular material distinct elements using functionally defined objects

Sergey I. Vyatkin ¹✉, Boris S. Dolgovesov ¹

¹Institute of Automation and Electrometry SB RAS,
Novosibirsk, 630090, Russian Federation

For citation

Vyatkin, S.I., Dolgovesov, B.S. (2024) 'Computer modeling of complex interactions of distinct elements of granular materials using functionally specified objects', *Software & Systems*, 37(3), pp. 412–420 (in Russ.). doi: 10.15827/0236-235X.142.412-420

Article info

Received: 19.01.2024

After revision: 11.03.2024

Accepted: 19.03.2024

Abstract. The paper proposes a method of contact dynamics for granular materials using functionally specified objects. It also considers the problem of modeling complex interactions of a large number of individual elements of granular materials using information about the mass, location and contacts of contacting surfaces to obtain linear complexity based on the total number of contacts detected at each iteration. The proposed method of contact dynamics is based on numerical modeling with access to all variables, on non-smoothness in space, time and force. That is, an area bounded by a set of surfaces is a valid set in the space of configuration variables. In case of colliding, the time function representing velocity will be intermittent. Friction is determined by the ratio between a reaction force and a local relative velocity. The paper proposed a velocity calculation method for modeling large sets of solids. Functionally defined objects, in particular ellipsoids, are used as separate elements of granular materials. The method has no repeated pairwise comparisons between objects. Therefore, there is linear complexity in terms of the total number of contacts detected at each iteration. The developed model of friction in a configuration space of solids is based on the maximum dispersion principle. As a result, the model combines rolling and sliding friction (while providing the expected characteristics) with calculating the permissible speed and reaction to friction for all objects. The efficiency of the method is confirmed by computer simulation. Experiments have shown that a collision can split into several separate episodes of contact on a microscopic time scale. This leads to the fact that a collision outcome depends heavily on a shape of corresponding bodies. When individual elements of granular materials have an ellipsoid shape, the result is more acceptable. The proposed method of contact dynamics might be useful for calculating quasi-static evolutions of granular materials in mechanical engineering, construction, numerical processing using finite elements of dynamic and quasi-static problems associated with deformable bodies, etc.

Keywords: functionally defined objects, computer modeling, dynamic interaction, contact, friction

Acknowledgements. The work was carried out at the IA&E SB RAS with the support of the Ministry of Education and Science of the Russian Federation within the State Assignment no. 124041700102-4

References

1. Wachs, A., Uhlmann, M., Derksen, J., Huet, D.P. (2023) 'Modeling of short-range interactions between both spherical and non-spherical rigid particles', in *Modeling Approaches and Computational Methods for Particle-Laden Turbulent Flows*, pp. 217–264. doi: 10.1016/B978-0-32-390133-8.00019-0.
2. Strecke, M., Stueckler, J. (2021) 'DiffSDFSim: Differentiable rigid-body dynamics with implicit shapes', *Proc. Int. Conf. 3DV*, pp. 96–105. doi: 10.1109/3DV53792.2021.00020.
3. Elandt, R., Drumwright, E., Sherman, M.A., Ruina, A. (2019) 'A pressure field model for fast, robust approximation of net contact force and moment between nominally rigid objects', *Proc. IEEE/RSJ Int. Conf. IROS*, pp. 8238–8245. doi: 10.1109/IROS40897.2019.8968548.
4. Fazeli, N., Tedrake, R., Rodriguez, A. (2018) 'Identifiability analysis of planar rigid-body frictional contact', in *Robotics Research. SPAR*, 3, pp. 665–682. doi: 10.1007/978-3-319-60916-4_38.
5. Kim, Y. (2018) 'The Hardware accelerated physics engine with operating parameters controlling the numerical error tolerance', *IJAST*, 119, pp. 145–152. doi: 10.14257/ijast.2018.119.13.
6. Miyamoto, S., Yamashita, M. (2021) 'An improved convergence based on accelerated modulus-based Gauss–Seidel method for interactive rigid body simulations', *SN Appl. Sci.*, 3, art. 266. doi: 10.1007/s42452-021-04238-8.
7. Verschoor, M., Jalba, A. (2019) 'Efficient and accurate collision response for elastically deformable models', *ACM Transactions on Graphics*, 38(2), art. 17, pp. 1–20. doi: 10.1145/3209887.
8. Lidstrom, P. (2017) 'Kinematics for unilateral constraints in multibody dynamics', *MMS*, 22(8), pp. 1654–1687. doi: 10.1177/1081286516642270.
9. Geilinger, M., Hahn, D., Zehnder, J., Bacher, M. (2020) 'ADD: Analytically differentiable dynamics for multi-body systems with frictional contact', *ACM Transactions on Graphics*, 39(6), art. 190, pp. 1–15. doi: 10.1145/3414685.3417766.
10. Coevoet, E., Andrews, S., Relles, D., Kry, P.G. (2020) 'Distant collision response in rigid body simulations', *Comput. Graphics Forum*, 39(8), pp. 113–122. doi: 10.1111/cgf.14106.
11. Ferguson, Z., Li, M., Schneider, T., Gil-Ureta, F. (2021) 'Intersection-free rigid body dynamics', *ACM Transactions on Graphics*, 40(4), art. 183, pp. 1–16. doi: 10.1145/3476576.3476773.

12. Flores, P. (2022) 'Contact mechanics for dynamical systems: A comprehensive review', *Multibody Syst. Dyn.*, 54, pp. 127–177. doi: 10.1007/s11044-021-09803-y.
13. Vyatkin, S.I., Dolgovesov, B.S. (2022) 'Methods of interactive modeling and visualization of functionally defined objects for 3D WEB applications', *Avtometriya*, 58(1), pp. 111–118 (in Russ.). doi: 10.15372/AUT20220112.
14. Vyatkin, S.I., Dolgovesov, B.S., Korsun, A.S. (2003) 'Detecting the collision of functionally defined objects with recursive division of object space', *Avtometriya*, 39(6), pp. 119–126 (in Russ.).
15. Vazquez, M.B., Pereira-Saez, M.J., Rodriguez-Raposo, A.B., Souto-Salorio, M.J. (2022) 'Contact detection between a small ellipsoid and another quadric', *Comput. Aided Geometric Design*, 98, art. 102136. doi: 10.1016/j.cagd.2022.102136.
16. Li, Y., Asai, M., Chandra, B., Isshiki, M. (2020) 'Energy-tracking impulse method for particle-discretized rigid-body simulations with frictional contact', *Computational Particle Mechanics*, 8, pp. 237–258. doi: 10.1007/s40571-020-00326-5.

Авторы

Вяткин Сергей Иванович¹, к.т.н.,
старший научный сотрудник, sivser@mail.ru
Долгovesов Борис Степанович¹,
к.т.н., зав. лабораторией, bsd@iae.nsk.su

¹ Институт автоматизации и электрометрии СО РАН,
г. Новосибирск, 630090, Россия

Authors

Sergey I. Vyatkin¹, Cand. of Sci. (Engineering),
Senior Researcher, sivser@mail.ru
Boris S. Dolgovesov¹, Cand. of Sci. (Engineering),
Head of Laboratory, bsd@iae.nsk.su

¹ Institute of Automation and Electrometry SB RAS,
Novosibirsk, 630090, Russian Federation

Распознавание и отслеживание дефектов дорожного полотна в реальном времени на основе комплексного использования стандартных вычислительных процедур и глубоких нейронных сетей

М.О. Антонов ¹✉, И.О. Темкин ¹

¹ Университет науки и технологий МИСиС, г. Москва, 119049, Россия

Ссылка для цитирования

Антонов М.О., Темкин И.О. Распознавание и отслеживание дефектов дорожного полотна в реальном времени на основе комплексного использования стандартных вычислительных процедур и глубоких нейронных сетей // Программные продукты и системы. 2024. Т. 37. № 3. С. 421–430. doi: 10.15827/0236-235X.142.421-430

Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 16.05.2024

После доработки: 17.06.2024

Принята к публикации: 27.06.2024

Аннотация. В данной работе рассматриваются вопросы, связанные с распознаванием и отслеживанием дефектов дорожного покрытия в режиме реального времени. Это является важной задачей для современных систем мониторинга, беспилотного транспорта и управления дорожной инфраструктурой. Предложенная система способна выявлять дефекты дорожного покрытия с помощью глубоких нейронных сетей и отслеживать их путем присвоения уникальных идентификаторов. Используемая для распознавания дефектов модель YOLOv8 продемонстрировала высокую точность и скорость в задачах обнаружения объектов. Модель была обучена на составных наборах данных, что позволило добиться надежного обобщения. Примечательной особенностью предлагаемого подхода является отсутствие жестких требований к техническим характеристикам камер и других компонентов системы, это делает решение гибким и адаптируемым к различным условиям эксплуатации. Отслеживание дефектов реализовано с помощью венгерского алгоритма и фильтра Калмана, что обеспечивает высокую точность и надежность процесса наблюдения. Венгерский алгоритм эффективно устанавливает соответствия между обнаружениями в последовательных кадрах, а фильтр Калмана сглаживает траектории и прогнозирует будущие положения дефектов, что имеет решающее значение для точного присвоения уникальных идентификаторов и анализа данных. Представленные параметры обучения моделей, оценки потребления вычислительных ресурсов и производительности системы показывают, что предлагаемое решение имеет приемлемые характеристики для применения в режиме реального времени. Эти измерения помогают оценить затраты на внедрение и определить потенциал интеграции в существующие системы мониторинга. Таким образом, представленное решение обеспечивает высокоэффективное и точное распознавание и отслеживание дефектов дорожного покрытия в режиме реального времени, повышая безопасность и качество дорожной инфраструктуры, а также упрощая управление и планирование работ по техническому обслуживанию.

Ключевые слова: распознавание в реальном времени, глубокие нейронные сети, отслеживание объектов, YOLOv8, венгерский алгоритм, фильтр Калмана, дефекты дорожного полотна

Введение. Национальный проект «Безопасные и качественные автомобильные дороги» ставит своей целью приведение региональных дорог и дорожной сети городских агломераций в соответствие нормам. Реализация этого проекта обусловила увеличение объема контрактов по комплексной диагностике дорог и оценке их технического состояния. Организации, участвующие в проекте, обязаны предоставлять заказчику следующую информацию на основе проведенной диагностики: видеоданные обследуемого участка, ведомость дефектов дорожного покрытия и отчет о комплексной диагностике с оценкой транспортно-эксплуатационного состояния.

За последнее десятилетие было проведено множество исследований, посвященных применению методов искусственного интеллекта для анализа изображений дорожной обстанов-

ки. Учитывая наличие обширного объема качественных видеоданных в дорожной отрасли и стремительное развитие компьютерного зрения, актуальным и перспективным направлением исследований становится выявление дефектов дорожного покрытия не только по статическим изображениям автомобильных дорог, но и в реальном времени. Данная тематика так же актуальна и в области автономных транспортных средств, где точное и оперативное обнаружение дефектов на дороге является ключевым для обеспечения безопасности и эффективности самоуправляемых автомобилей. Достигновения в этой области не только способствуют общему улучшению состояния дорог, но и играют решающую роль в расширении возможностей и надежности систем навигации автономных транспортных средств. По мере развития индустрии искусственного интеллект-

та и производительности вычислительных систем в целом обнаружение нарушений как отдельная функция становится недостаточным критерием оценки качества, поскольку сформированные данные требуют дальнейшей обработки путем их классификации по степени критичности того или иного дефекта и выполнения статистического анализа для оперативного принятия управленческих решений.

Научная новизна исследования заключается в разработке универсальной методики для обнаружения и отслеживания дефектов дорожного покрытия, в частности, выбоин и проломов. В отличие от традиционных подходов, которые часто требуют адаптации к различным типам дефектов, в рамках настоящей работы предложено обобщенное решение на базе венгерского алгоритма и фильтра Калмана. Представленный механизм обеспечивает плавную интеграцию в другие сценарии обнаружения и отслеживания, например, деформации, износ поверхности и дорожной разметки, предлагая масштабируемое решение для комплексной диагностики состояния дорог.

Краткий анализ существующих подходов

Традиционно основной формой обнаружения дефектов, определенным образом классифицируемых согласно ГОСТ Р 50597-2017, являлся ручной визуальный осмотр профессиональными инспекторами, однако подобный процесс контроля малоэффективен, затратен материально и по времени. Более того, результаты проверки являются субъективными, поскольку полностью зависят от опыта отдельного специалиста. С другой стороны, технологии мониторинга при помощи специальных датчиков (стационарных и мобильных) имеют ограничения, связанные как с точностью и оперативностью обработки информации, так и с их высокой стоимостью.

Разберем основные известные подходы к их идентификации и анализу.

Подход на основе вибрации. Под основной идеей подхода подразумевается использование GPS, акселерометра и гироскопа для картографирования дорожного покрытия [1, 2]. К достоинствам использования метода можно отнести возможность анализа состояния дорожной сети в реальном времени с достаточно высокой точностью [3]. Стоит отметить, что данные точность и полнота охвата зависят в том числе и от того, был или нет наезд на исследуемый дефект, что является недостатком. К дополнительным не-

достаткам следует отнести ограничения в использовании специфичного устройства ПО [4].

3D-лазер. В данном типе исследования применяется 3D-лазерное сканирование, используя сеточный подход для определения характеристик дефектов [5]. В рамках данного метода формируется облако точек в трехмерном пространстве, что позволяет проводить анализ и классифицировать обнаружения по степени критичности [6]. К недостаткам относятся дороговизна используемого оборудования, а также небольшая дальность обнаружения, что, в свою очередь, накладывает ограничения на применение этого метода в реальном времени на борту автономных транспортных средств [7].

3D-стереовизуализация. Развитием и дополнением предыдущего метода является 3D-стереовизуализация. Метод подразумевает формирование трехмерной поверхности дорожного покрытия на основе набора входных изображений. Например, в [8] авторы представили подход по восстановлению полной трехмерной поверхности дорожного полотна. Методика включала в себя исправление любых наблюдаемых искажений, извлечение особенностей дефектов и 3D-реконструкцию. Метод, создающий набор точек в трехмерном пространстве и позволяющий с высокой точностью отображать дорожное покрытие, рассмотрен в [9]. Используя стереоизображения и технологии обработки изображений, система может идентифицировать различные дорожные дефекты, такие как выбоины, неровности, трещины и т.д. Стоит отметить очень высокую точность анализа дорожного покрытия, обеспечиваемую данным методом. Необходимо упомянуть и требования к производительности оборудования, поскольку трехмерной реконструкции поверхности необходима большая вычислительная мощность.

Искусственные нейронные сети. Данный метод наблюдения использует обработку полученных с камер изображений с последующим обнаружением дефектов средствами сверточных сетей и сетей глубокого обучения, причем в качестве камеры необязательно выступает профессиональное оборудование – в некоторых случаях достаточно мобильного телефона с камерой высокого разрешения [10]. Ряд исследователей предложили использовать модифицированную модель ResNet50-RetinaNet при локализации выбоин по тепловизионным изображениям вместо стандартных средств видеofиксации [11]. В работе [12] исследована модель YOLOv3 на основе собранного для обучения

набора данных, состоящего из 1 500 изображений дорог Индии. К достоинствам подхода можно отнести экономичность, позволяющую при этом определить форму и площадь дефекта с достаточно высокой точностью. Стоит отметить присущий данному методу компромисс между разнообразием набора использованных при обучении данных и точностью результата, временем обработки, а также размером модели в целом, который может являться одновременно с этим существенным недостатком – неполный набор данных и его некачественная разметка сводят возможности методологии на нет.

Принимая во внимание достоинства метода, в рамках настоящей работы представлено решение для обнаружения дефектов на базе более современной архитектуры YOLOv8 в составе с разработанным алгоритмом отслеживания, а приведенные недостатки нивелированы за счет ручного сбора и разметки данных.

Отслеживание дефектов покрытия

Основная идея метода заключается в выделении и сохранении набора идентификаторов, каждый из которых связан с конкретным объектом, и в динамическом обновлении их на основе новых данных по каждому из кадров. Предлагается использовать модель движения в комплексе с алгоритмом целераспределения, где в качестве модели используется фильтр Калмана, а целераспределения – венгерский алгоритм. Представленный метод основан на оценке ряда параметров предсказания в каждом кадре, а именно: текущей скорости движения (по координатам X и Y по отдельности) и размера (высоты и ширины обнаруженного объекта), а значит, может быть применим для различных типов дефектов, если таковые будут обнаружены моделью глубокого обучения на этапе первичной идентификации. Соотношения сторон ограничительной рамки вокруг объекта всегда берутся от самого первого обнаружения. Скорости движения предсказываются при помощи фильтров Калмана, а матрица пересечений ограничительных рамок по IoU-коэффициенту, который отражает степень перекрытия ограничительной рамки детектируемого объекта с предыдущим кадром, а сами детекции назначаются при помощи венгерского алгоритма.

Опишем основные принципы работы метода, представленные в виде последовательности этапов.

Инициализация. Начиная с первого кадра видеоряда каждому обнаруженному объекту

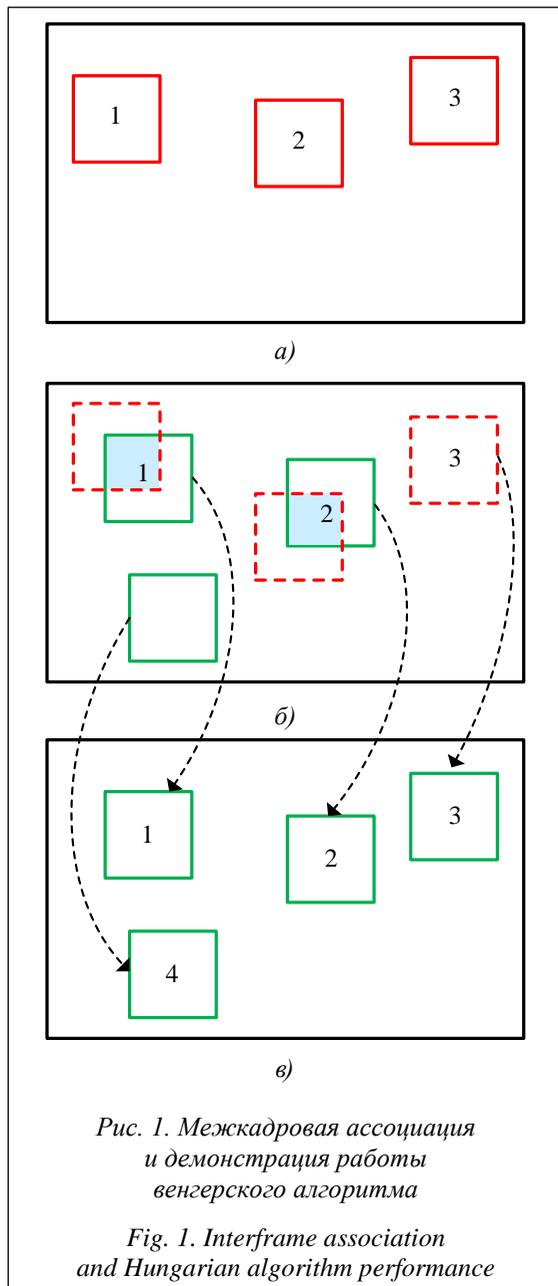
присваивается уникальный идентификатор. В дальнейшем для каждого нового обнаружения вводится новый идентификатор, а отношения между ними при переходе от кадра к кадру реализуются при помощи ассоциации.

Ассоциация. Венгерский алгоритм, также известный как алгоритм Куна–Манкреса, позволяет определить, совпадает ли дефект дорожного покрытия из одного кадра с дефектом из последующего кадра. Сам процесс ассоциации представлен на рисунке 1 на примере трех исходных обнаружений.

Рассмотрим межкадровую ассоциацию и принцип работы на примере двух последовательных кадров живого видеопотока. На первом кадре при инициализации (рис. 1а) обученная модель предсказывает три дефекта дорожного покрытия, в результате чего за счет ассоциации происходит присваивание трех идентификаторов для первого, второго и третьего дефектов соответственно. На втором кадре (рис. 1б) модель идентичным образом предсказывает три дефекта, за исключением третьего, ранее обнаруженного, в результате чего новый объект приобретает четвертое значение идентификатора, а третий подлежит удалению при его отсутствии на последующих кадрах (подробнее в секции «удаление» настоящего раздела), что приводит к состоянию (рис. 1в), позволяющему отслеживать каждое из обнаружений. Стоит отметить, что изменение координат и скорости первого и второго дефектов осуществляется средствами фильтра Калмана.

Предсказание и корректировка. Для формирования оптимальной оценки состояния фильтр Калмана использует динамическую модель системы, известные управляющие воздействия и дальнейшее множество последовательных измерений. Алгоритм на данной стадии состоит из двух повторяющихся фаз – предсказание и корректировка, где в первом случае происходит предсказание состояния в следующий момент времени (с учетом погрешности измерений), а во втором – корректировка за счет поступившей новой информации уже имеющегося предсказанного значения (идентичным образом с учетом погрешности и зашумленности данной информации).

Уравнения для предсказаний и корректировок представлены в матричной форме. Первоначальная оценка параметров осуществляется с помощью вектора средних значений и матрицы ковариаций. Вектор состояния состоит из абсциссы u и ординаты центра области обнаружения объекта v , а также площади области s ,



соотношения сторон r этой области (считается постоянным) и изменения первых трех параметров – скоростей, принимающих нулевое состояние при инициализации. Ковариация – матрица неопределенности в оценке, где большее число означает большую неопределенность. Выбор меньшего значения неопределенности для области обнаружения объекта, а также его площади и соотношения сторон обусловлен наличием погрешности в 10 пикселей, полученной эмпирически на основе среднего размера используемого изображения и размеров детектируемых объектов:

$$x = [u \ v \ s \ r \ \dot{u} \ \dot{v} \ \dot{s}], \quad (1)$$

$$P = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1000 \end{bmatrix} \quad (2)$$

Дополняя ассоциацию, осуществляется измерение областей детекции для каждого объекта соответственно (рис. 1а). Одновременно с этим происходит инициализация среднего значения и ковариации в соответствии с выражениями (1) и (2) для каждой области обнаружения. В состоянии б) в фазе предсказания состояния системы определяются области обнаружения дефектов из состояния а) в момент времени б), после чего полученные значения сопоставляются с новыми областями при помощи венгерского алгоритма, а переход в фазу коррективы завершает цикл.

Рассмотрим подробнее последовательность матричных преобразований.

1. Предсказание состояния системы:

$$\hat{x}_k^- = F\hat{x}_{k-1} + Bu_{k-1}, \quad (3)$$

где \hat{x}_k^- – предсказание состояния системы в текущий момент времени; F – матрица перехода между состояниями (динамическая модель системы приведена далее); \hat{x}_{k-1} – состояние системы в прошлый момент времени; B – матрица применения управляющего воздействия (при его отсутствии значение принимается как нулевое); u_{k-1} – управляющее воздействие в прошлый момент времени, которое в рассматриваемой задаче принимается равным нулю.

Матрица перехода F имеет ненулевые и составленные таким образом коэффициенты, чтобы в случае умножения слева на вектор состояния x значения результирующего вектора представляли собой сумму позиционных значений с их производными:

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (4)$$

2. Предсказание ошибки ковариации:

$$P_k^- = FP_{k-1}F^T + Q, \quad (5)$$

где P_k^- – предсказание ошибки; P_{k-1} – ошибка в прошлый момент времени; Q – ковариация

шума процесса. На основе размерности вектора состояния x , размерности матриц ковариации и шума процесса составят 7×7 для P и Q соответственно.

3. Вычисление усиления Калмана:

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1} \quad (6)$$

где K_k – усиление Калмана; H – матрица измерений, отображающая отношения измерений и состояний; R – ковариация шума измерения, которая в случае использования каких-либо радаров задается конструктивно, а в случае использования модели YOLOv8 – в пикселях (по 1 пикселю в центре и по 10 для ширины и высоты соответственно). Размерность матриц измерений H и ковариации шума измерения R , а также выбор коэффициентов обусловлены набором входящих от модели предсказания значений u , v , s и r для центра обнаруженного объекта и его сторон соответственно:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}, \quad (7)$$

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix}. \quad (8)$$

4. Обновление оценки с учетом измерения z_k :

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - H \hat{x}_k^-), \quad (9)$$

где z_k – измерение в текущий момент времени.

5. Обновление ошибки ковариации:

$$P_k = (I - K_k H) P_k^-, \quad (10)$$

где I – матрица идентичности.

Удаление. Завершаются все преобразования стадией удаления с заданной глубиной хранения объекта, составляющей 10 кадров для настоящего исследования. Принимая значение IoU равным 0.3, достаточно совпадения на 30 % при пересечении областей обнаружения. В случае отсутствия области обнаружения в детекторе (например, вероятность нахождения в кадре или значение IoU ниже пороговых значений) на протяжении 10 кадров подряд равносильно удалению из памяти программы, а значит, на 11-м кадре объект уже получит новый уникальный идентификатор.

Анализ и подготовка наборов данных

В соответствии с поставленными для настоящего изыскания целями были проведены ана-

лиз и подготовка наборов данных для обучения модели, включая наиболее популярные и находящиеся в открытом доступе на момент проведения вычислительных экспериментов изображения с размеченными дефектами дорожного полотна.

В частности, рассмотрев самый популярный из них, авторы [13] представили набор данных из 47 420 изображений дорог разных стран (Японии, Индии, Чешской Республики, Норвегии, США и Китая).

Предварительный анализ показал, что только 26 520 изображений из всего набора аннотированы, иными словами, фактически около половины изображений не содержат заявленных на них дефектов или для них отсутствует разметка. Более глубокое исследование выявило, что из оставшихся 26 520 изображений только 6 544 могут быть классифицированы как содержащие рассматриваемые в настоящей работе выбоины и проломы, а качество их разметки неудовлетворительное (обнаружено множество примеров, где контуры детекции не находятся по границам объекта, вместо дефекта аннотирован, например, светофор и т.п.).

Принимая во внимание этот факт, было принято решение о самостоятельной подготовке набора данных, учитывая опыт и рекомендации других исследователей, но не опираясь на их результаты во избежание попадания некачественных данных в финальную выборку, а разметку осуществить вручную. Результатом стала подборка 30 615 изображений с разрешением не менее $1\,280 \times 1\,280$ только выбоин и проломов, обучение на которых производилось в рамках данного исследования.

При подготовке был сделан акцент на разнообразии сред, в которых запечатлен тот или иной дефект. Например, учитывая потребность применения результатов исследования в среде с большим количеством грунтовых покрытий вместо асфальтовых, а также необходимость размещения камеры на лобовом стекле движущегося транспортного средства вместо создания специализированных видеолaborаторий, выборка была расширена соответствующими, в том числе запечатленными самостоятельно, примерами. Дополнительно набор данных содержит изображения заполненных водой выбоин и проломов (<http://www.swsys.ru/uploaded/image/2024-3/18.jpg>), гарантируя возможность распознавания различных сценариев и повышая надежность результата.

Обучение моделей

В каждой категории YOLOv8 имеются пять моделей для обнаружения, сегментации и классификации, где YOLOv8n самая быстрая и малая по размеру, в то время как YOLOv8x самая точная, но одновременно с этим самая медленная из представленных. Обучение моделей проводилось на графических модулях NVIDIA GeForce RTX 3090 TI в количестве четырех единиц с идентичными параметрами обучения для всех версий модели, представленными в таблице 1.

Таблица 1

Параметры обучения моделей YOLOv8

Table 1

YOLOv8 model training parameters

Параметр	Значение
Эпоха	55
Размер партии	32
Оптимизатор	Adam
Темп обучения	0.000714
Размер изображений	1 280

С целью ускорения обучения был использован модуль DistributedDataParallel (DDP), входящий в состав пакета PyTorch. DDP реализует параллелизм данных на уровне модуля, который может работать на нескольких машинах, а использующие DDP приложения должны порождать несколько процессов и создавать по одному экземпляру DDP для каждого из них. Рекомендуемый способ применения DDP – со-

здать один процесс для каждой реплики модели, которая может охватывать несколько физических устройств. Каждый графический модуль в рамках текущего исследования представляет собой отдельный процесс, и связь между ними осуществляется с помощью стандартного для операционных систем межпроцессного взаимодействия.

На основе прогресса обучения для модели YOLOv8x (рис. 2) можно сделать вывод, что ограничение в виде обозначенных выше 55 эпох необязательно и для достижения лучших результатов обучение может быть продолжено. Дополнительно графики с префиксами train и val (тренировки и валидации соответственно) имеют схожие формы, что свидетельствует о качестве обучения и хорошей работе на данных, которые ранее не демонстрировались. Учитывая, что train loss не приняла нулевое значение, а val loss не искривляются под конец обучения вверх, можно сделать вывод, что модель не была переобучена. Результаты прогресса обучения моделей других размеров носят схожий характер и идентичные для YOLOv8x линии тренда.

После обучения в результате валидации были получены представленные в таблице 2 характеристики моделей. Точность и производительность моделей оценивались на основе параметра mAP50-95 – метрики, представляющей среднюю точность (Average Precision, AP) по нескольким порогам пересечения областей IoU от 0.5 до 0.95. Метрика AP измеряет точ-

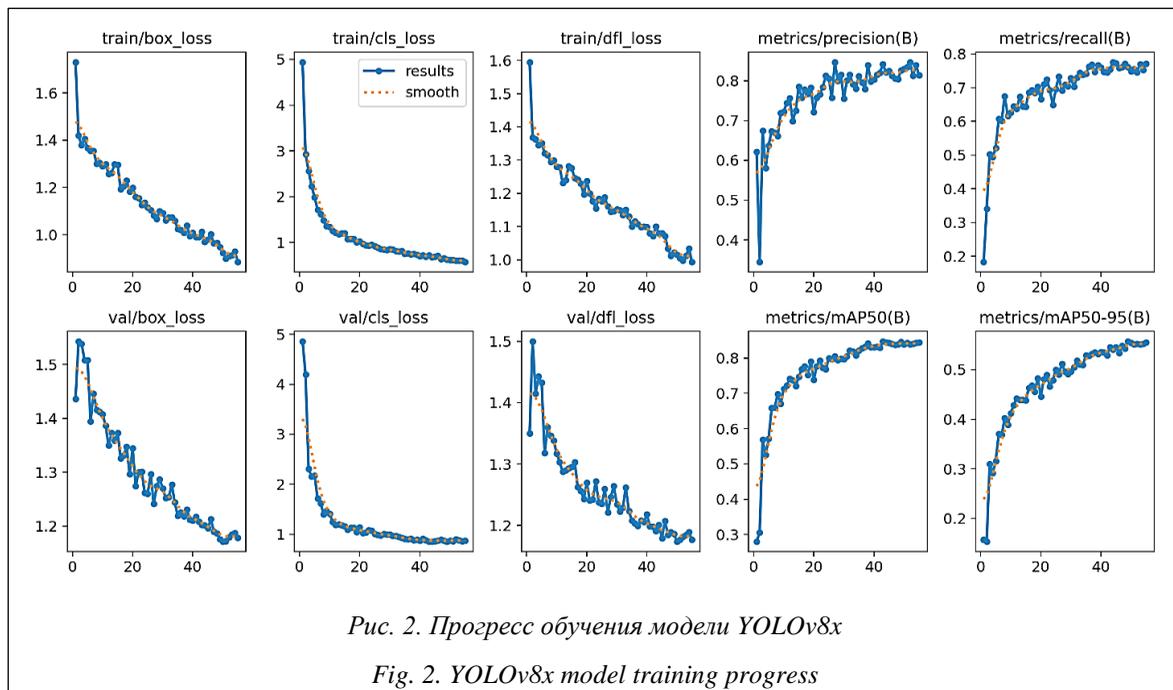


Рис. 2. Прогресс обучения модели YOLOv8x

Fig. 2. YOLOv8x model training progress

Таблица 2

Характеристики обученных моделей YOLOv8

Table 2

Characteristics of trained YOLOv8 models

Модель	Размер, МБ	mAP50-95, ед.	Параметры, млн	FLOPs, млрд
YOLOv8n	6.4	55.8	3.2	8.1
YOLOv8s	22.6	57.6	11.2	28.4
YOLOv8m	52.1	59.9	25.9	78.7
YOLOv8l	87.8	60.3	43.7	164.8
YOLOv8x	136.8	60.6	68.2	257.4

ность модели, суммируя кривую точности-полноты. Усредняя AP по диапазону порогов IoU, mAP50-95 дает всестороннюю оценку производительности модели обнаружения при различных уровнях строгости перекрытия.

Можно наглядно проследить, что наиболее точная, но наименее быстрая модель YOLOv8x достигает значения в 60.6 mAP50-95, в то время как наихудший, но вместе с тем максимально производительный результат достигается моделью YOLOv8n – 55.8 mAP50-95. Одновременно с этим более крупные модели требуют большего количества вычислений, что, в свою очередь, приводит к снижению энергоэффективности. Дополнительно стоит подчеркнуть, что наилучшие характеристики с точки зрения размера, параметров и качества обнаружения у модели YOLOv8m. Здесь и далее будет наглядно прослеживаться тенденция в пользу выбора именно этой версии, а не самой большой YOLOv8x.

Обнаружение дефектов покрытия

По аналогии с приведенными характеристиками для обучения графического модуля тестирование быстродействия моделей осуществляется на NVIDIA GeForce RTX 3090 TI, а в качестве CPU выступают Intel Xeon E5-2696 v4 с частотой 2.20 ГГц в количестве двух единиц и 256 гигабайт RAM.

В таблице 3 представлены результаты потребления вычислительных ресурсов и производительности для каждого типа модели YOLOv8, где под потреблением подразумевается количество занятой видеопамати, а под производительностью – время обработки одного изображения на графическом модуле. Быстрая обработка результатов имеет важное значение для беспилотных транспортных средств, чтобы избежать выбоин. Это делает обязательным минимизацию времени обработки. Учитывая предыдущие результаты точ-

ности, также можно сделать вывод, что модель YOLOv8m обладает наилучшим результатом в том числе по отношению к потреблению и быстродействию, а дальнейшее увеличение скорости работы за счет применения YOLOv8s и YOLOv8n не приносит значительных результатов, а только снижает качество обнаружения в целом. Одновременно с этим модель YOLOv8n обеспечивает исключительное среднее время обработки изображения 8 мс, сохраняя размер модели всего 6.4 МБ, что упрощает ее размещение во встраиваемых системах.

Таблица 3

Потребление и производительность моделей YOLOv8

Table 3

Consumption and performance of YOLOv8 models

Модель	Потребление VRAM, МБ	Быстродействие, мс
YOLOv8n	688	8
YOLOv8s	716	9
YOLOv8m	804	11
YOLOv8l	918	15
YOLOv8x	1086	19

Учитывая, что целью настоящей работы является в том числе исследование возможности применения обычных камер, а не дорогостоящих передвижных видеолaborаторий, на рисунке 3 приведены результаты обнаружения дефектов дорожного полотна из движущегося транспортного средства.

Дополнительно стоит отметить, что результаты тестирования производительности не включают в себя какие-либо методы увеличения быстродействия, например, квантизацию, объединение входящих изображений в группы, конвертацию для поддержки такими сторонними фреймворками, как TensorRT, раскадровку на графическом модуле и т.п. Подобного рода оптимизации находятся вне рамок настоящей работы, но в случае их применения (а ни-



Рис. 3. Работа модели в реальном времени в движущемся транспортном средстве

Fig. 3. Real-time model operation in a moving vehicle

каких ограничений для этого не существует) производительность процесса обнаружения и системы в целом потенциально возрастет.

Рассматривая результаты производительности в составе с упомянутым ранее алгоритмом отслеживания, стоит обратить внимание на представленное в таблице 4 время обработки полного цикла для каждой из моделей. Данное значение является важным показателем для оценки быстродействия и включает в себя суммарно затраченное время на получение кадра из видеопотока, на его предварительную обработку, предсказание моделью, на постобработку, отслеживание и на вывод с формированием соответствующих результатов.

Таблица 4

Затрачиваемое на полный цикл обработки кадра время

Table 4

Time spent on a full frame processing cycle

Модель	Быстродействие, FPS
YOLOv8n	32
YOLOv8s	31
YOLOv8m	29
YOLOv8l	26
YOLOv8x	23

На основании представленных значений можно сделать вывод, что модель YOLOv8m демонстрирует наилучшие по отношению к другим моделям результаты быстродействия. Стоит отметить, что программа выполнялась

на одном ядре центрального процессора Intel Xeon E5-2696 v4 с частотой 2.20 ГГц, а значит, выполнение одноименного кода на более современном оборудовании или применение методов многопоточного программирования вместо текущего однопоточного увеличит производительность системы в целом.

Заключение

Применение обученной в рамках настоящей работы модели YOLOv8 – отличное решение для обнаружения дефектов дорожного полотна, которое обеспечивает баланс между точностью, скоростью и эффективностью использования ресурсов, что позволяет добиться значительных успехов в повышении безопасности дорожного движения и в обслуживании инфраструктуры. Изучение существующих средств диагностики и анализ имеющегося опыта обнаружения дефектов позволяют сделать вывод, что имеющееся техническое обеспечение требует усовершенствования, модернизации и разработки средств диагностики нового поколения. Прослеживается огромный потенциал для использования методов глубокого обучения, что особенно актуально для интеллектуального контроля дорожного движения, где в настоящее время имеется ограниченный объем исследований и данных. Кроме того, исследование подчеркивает важность разнообразного и надежного набора обучающих данных.

Версия модели YOLOv8m продемонстрировала наилучшие результаты, достигнув значе-

ния точности 59.9 mAP при быстродействии в 29 FPS для полного цикла анализа видеопотока в реальном времени на одном ядре центрального процессора. С дальнейшим развитием темы данной работы допустимо применение ее результатов не только для интеллектуального обслуживания городской инфраструктуры, но и для обнаружения и отслеживания иных классов объектов при помощи, например, беспилотных летательных аппаратов.

Таким образом, в рамках настоящего изыскания исследовано обнаружение дефектов дорожного полотна средствами предварительно обученной модели YOLOv8 с последующим отслеживанием объектов при помощи фильтра Калмана и венгерского алгоритма. Критерии оценки качества полученных результатов включали в себя время, размер модели, потребление ресурсов, время обработки полного цикла и надежность в различных условиях.

Список литературы

1. Темкин И.О., Клебанов Д.А., Дерябин С.А., Конов И.С. Метод определения состояния технологических дорог карьера при управлении взаимодействием роботизированных элементов горнотранспортного комплекса // Горный журнал. 2018. № 1. С. 78–82. doi: 10.17580/gzh.2018.01.14.
2. Rana S. Vibration based pavement roughness monitoring system using vehicle dynamics and smartphone with estimated vehicle parameters. Results in Engineering, 2021, vol. 12, art. 100294. doi: 10.1016/j.rineng.2021.100294.
3. Kim Y.-M., Kim Y.-G., Son S.-Y. et. al. Review of recent automated pothole-detection methods. Appl. Sci., 2022, vol. 12, no. 11, art. 5320. doi: 10.3390/app12115320.
4. Kandoi A., Agarwal H., Revadekar A., Chaudhari B.-N. Pothole detection using accelerometer and computer vision with automated complaint redressal. Proc. ICPSC, 2021, pp. 257–264. doi: 10.1109/ICSPC51351.2021.9451762.
5. Issaoui A.E., Feng Z., Lehtomäki M. et al. Feasibility of mobile laser scanning towards operational accurate road rut depth measurements. Sensors, 2021, vol. 21, no. 4, art. 1180. doi: 10.3390/s21041180.
6. Feng Z., Issaoui A.E., Lehtomäki M. et al. Pavement distress detection using terrestrial laser scanning point clouds – Accuracy evaluation and algorithm comparison. ISPRS Open J. of Photogrammetry and Remote Sensing, 2022, vol. 3, art. 100010. doi: 10.1016/j.ophoto.2021.100010.
7. De Blasii M.R., Di Benedetto A., Fiani M. Mobile laser scanning data for the evaluation of pavement surface distress. Remote Sens., 2020, vol. 12, no. 6, art. 942. doi: 10.3390/rs12060942.
8. Fan R., Ozgunalp U., Wang Y., Liu M., Pitas I. Rethinking road surface 3-D reconstruction and pothole detection: From perspective transformation to disparity map segmentation. IEEE Transactions on Cybernetics, 2021, vol. 52, no. 7, pp. 5799–5808. doi: 10.1109/TCYB.2021.3060461.
9. Staniek M. Stereo vision method application to road inspection. The Baltic J. of Road and Bridge Eng., 2017, vol. 12, no. 1, pp. 38–47. doi: 10.3846/bjrbe.2017.05.
10. Maeda H., Sekimoto Y., Seto T., Kashiyama T., Omata H. Road damage detection and classification using deep neural networks with images captured through a smartphone. Computer-Aided Civil and Infrastructure Eng., 2018, vol. 33, no. 12, pp. 1127–1141. doi: 10.1111/mice.12387.
11. Gupta S., Sharma P., Sharma D., Gupta V., Sambual N. Detection and localization of potholes in thermal images using deep neural networks. Multimed Tools Appl., 2020, vol. 79, pp. 26265–26284. doi: 10.1007/s11042-020-09293-8.
12. Dharneeshkar J., Aniruthan S., Karthika R., Parameswaran L. Deep learning-based detection of potholes in Indian roads using yolo. Proc. ICICT, 2020, pp. 381–385. doi: 10.1109/ICICT48043.2020.9112424.
13. Arya D., Maeda H., Ghosh S.K., Toshniwal D., Omata H., Kashiyama T., Sekimoto Y. Crowdsensing-based Road Damage Detection Challenge (CRDDC'2022). Proc. Int. Conf. Big Data, 2022, pp. 6378–6386. doi: 10.1109/Big-Data55660.2022.10021040.

Real-time recognition and tracking of road surface defects based on complex using of standard computational procedures and deep neural networks

Mikhail O. Antonov ¹✉, Igor O. Temkin ¹

¹ University of Science and Technology MISIS, Moscow, 119049, Russian Federation

For citation

Antonov, M.O., Temkin, I.O. (2024) 'Real-time recognition and tracking of road surface defects based on complex using of standard computational procedures and deep neural networks', *Software & Systems*, 37(3), pp. 421–430 (in Russ.). doi: 10.15827/0236-235X.142.421-430

Article info

Received: 16.05.2024

After revision: 17.06.2024

Accepted: 27.06.2024

Abstract. The paper considers several issues related to real-time recognition and tracking of road surface defects. This is an important task for modern monitoring systems, unmanned transportation and road infrastructure management. The pro-

posed system identifies pavement defects using deep neural networks and tracks them by assigning unique identifiers. The YOLOv8 model has demonstrated high accuracy and speed in object detection tasks during defect recognition. Due to training the model on composite datasets, generalization was reliable. A notable feature of this approach is the absence of strict requirements to the technical characteristics of cameras and other system components. This makes the solution flexible and adaptable to various operating conditions. Defect tracking is based on place using the Hungarian algorithm and Kalman filter. This ensures high accuracy and reliability of the observation process. The Hungarian algorithm efficiently establishes correspondences between detections in consecutive frames. The Kalman filter smooths trajectories and predicts future defect positions. This is crucial for accurate assignment of unique identifiers and data analysis. Model training parameters, estimates of computational resource consumption and system performance show that this solution has acceptable performance for real-time applications. These measurements help to estimate implementation costs and determine the potential for integration into existing monitoring systems. Thus, the presented solution provides highly efficient and accurate real-time recognition and tracking of pavement defects. It improves the safety and quality of road infrastructure and simplifies management and maintenance planning.

Keywords: real-time recognition, deep neural networks, object tracking, YOLOv8, Hungarian algorithm, Kalman filter, road surface defects

References

1. Temkin, I.O., Klebanov, D.A., Deryabin, S.A., Konov, I.S. (2018) 'Haul road condition determination under controlled interaction of robotic elements in open pit mining and transport system', *Mining J.*, (1), pp. 78–82 (in Russ.). doi: 10.17580/gzh.2018.01.14.
2. Rana, S. (2021) 'Vibration based pavement roughness monitoring system using vehicle dynamics and smartphone with estimated vehicle parameters', *Results in Engineering*, 12, art. 100294. doi: 10.1016/j.rineng.2021.100294.
3. Kim, Y.-M., Kim, Y.-G., Son, S.-Y. et al. (2022) 'Review of recent automated pothole-detection methods', *Appl. Sci.*, 12(11), art. 5320. doi: 10.3390/app12115320.
4. Kandoi, A., Agarwal, H., Revadekar, A., Chaudhari, B.-N. (2021) 'Pothole detection using accelerometer and computer vision with automated complaint redressal', *Proc. ICPSC*, pp. 257–264. doi: 10.1109/ICSPC51351.2021.9451762.
5. Issaoui, A.E., Feng, Z., Lehtomäki, M. et al. (2021) 'Feasibility of mobile laser scanning towards operational accurate road rut depth measurements', *Sensors*, 21(4), art. 1180. doi: 10.3390/s21041180.
6. Feng, Z., Issaoui, A.E., Lehtomäki, M. et al. (2022) 'Pavement distress detection using terrestrial laser scanning point clouds – Accuracy evaluation and algorithm comparison', *ISPRS Open J. of Photogrammetry and Remote Sensing*, 3, art. 100010. doi: 10.1016/j.ophoto.2021.100010.
7. De Blasiis, M.R., Di Benedetto, A., Fiani, M. (2020) 'Mobile laser scanning data for the evaluation of pavement surface distress', *Remote Sens.*, 12(6), art. 942. doi: 10.3390/rs12060942.
8. Fan, R., Ozgunalp, U., Wang, Y., Liu, M., Pitas, I. (2021) 'Rethinking road surface 3-D reconstruction and pothole detection: from perspective transformation to disparity map segmentation', *IEEE Transactions on Cybernetics*, 52(7), pp. 5799–5808. doi: 10.1109/TCYB.2021.3060461.
9. Staniek, M. (2017) 'Stereo vision method application to road inspection', *The Baltic J. of Road and Bridge Eng.*, 12(1), pp. 38–47. doi: 10.3846/bjrbe.2017.05.
10. Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., Omata, H. (2018) 'Road damage detection and classification using deep neural networks with images captured through a smartphone', *Computer-Aided Civil and Infrastructure Eng.*, 33(12), pp. 1127–1141. doi: 10.1111/mice.12387.
11. Gupta, S., Sharma, P., Sharma, D., Gupta, V., Sambual, N. (2020) 'Detection and localization of potholes in thermal images using deep neural networks', *Multimed Tools Appl.*, 79, pp. 26265–26284. doi: 10.1007/s11042-020-09293-8.
12. Dhameeshkar, J., Aniruthan, S., Karthika, R., Parameswaran, L. (2020) 'Deep learning-based detection of potholes in Indian roads using yolo', *Proc. ICICT*, pp. 381–385. doi: 10.1109/ICICT48043.2020.9112424.
13. Arya, D., Maeda, H., Ghosh, S.K., Toshniwal, D., Omata, H., Kashiyama, T., Sekimoto, Y. (2022) 'Crowdsensing-based Road Damage Detection Challenge (CRDDC'2022)'. *Proc. Int. Conf. Big Data*, pp. 6378–6386. doi: 10.1109/Big-Data55660.2022.10021040.

Авторы

Антонов Михаил Олегович¹, аспирант,
ассистент, amo.mikhail@gmail.com
Темкин Игорь Олегович¹, д.т.н.,
профессор, igortemkin@yandex.ru

Authors

Mikhail O. Antonov¹, Postgraduate Student,
Assistant, amo.mikhail@gmail.com
Igor O. Temkin¹, Dr.Sci. (Engineering),
Professor, igortemkin@yandex.ru

¹ Университет науки и технологий МИСиС,
г. Москва, 119049, Россия

¹ University of Science and Technology MISIS,
Moscow, 119049, Russian Federation

Динамическая генерация пользовательского интерфейса мобильного медицинского приложения на основе онтологического подхода

В.В. Грибова ¹✉, А.Я. Лифшиц ¹, Ф.М. Москаленко ¹,
Е.А. Шалфеева ¹, Н.Е. Шевченко ¹

¹ Институт автоматизации и процессов управления ДВО РАН, г. Владивосток, 690041, Россия

² Владивостокский государственный университет, г. Владивосток, 690014, Россия

Ссылка для цитирования

Грибова В.В., Лифшиц А.Я., Москаленко Ф.М., Шалфеева Е.А., Шевченко Н.Е. Динамическая генерация пользовательского интерфейса мобильного медицинского приложения на основе онтологического подхода // Программные продукты и системы. 2024. Т. 37. № 3. С. 431–439. doi: 10.15827/0236-235X.142.431-439

Информация о статье

Группа специальностей ВАК: 2.3.5

Поступила в редакцию: 19.12.2023

После доработки: 21.02.2024

Принята к публикации: 24.02.2024

Аннотация. Системы поддержки принятия врачебных решений призваны помогать врачам своевременно принимать правильные решения. Особое значение имеет использование таких систем у постели больного удаленно от медицинских центров и поликлиник. Структура электронной медицинской карты имеет много иерархически представляемых разделов, а разнообразие вносимой в нее информации оценивается десятками тысяч возможных терминов и их значений. Учитывая это, актуальной является разработка новых методов генерации мобильного пользовательского интерфейса заполнения электронной медицинской карты. Разработан метод генерации адаптивного интерфейса заполнения электронной медицинской карты пациента для мобильных приложений на основе онтологического подхода. Визуализация осуществляется посредством вертикальных слоев, выбираемых по уровням вершин в иерархической семантической структуре документа электронной медицинской карты. Описана технология генерации, продемонстрирована апробация на мобильной оболочке по диагностике заболеваний. Результатом также является новая технология генерации пользовательского интерфейса и реализованная по ней программная оболочка для компоновки сервисов диагностики по произвольным группам заболеваний для мобильных устройств. Практическая значимость технологии состоит в том, что представленный метод генерации интерфейса универсален: может быть использован для создания адаптивных пользовательских интерфейсов в других предметных областях и задачах, для которых характерны большие объемы вводимой сложноструктурированной информации.

Ключевые слова: мобильное приложение, база знаний, электронная медицинская карта пациента, пользовательский интерфейс, динамическая генерация интерфейса, онтология, интеллектуальные системы

Благодарности. Работа выполнена при поддержке Минобрнауки РФ (задание № FWFV-2021-0004)

Введение. Проблемы в области здравоохранения являются одними из самых острых в современном обществе. Количество различных проявлений и осложнений заболеваний непрерывно растет, и врачи вынуждены запоминать огромные объемы информации по разным областям медицинских знаний. Приходится учитывать различные факторы: особенности пациента и течения заболевания, показания и противопоказания методики обследования или способа лечения, усиление влияния или совместимость лекарственных препаратов, индивидуальную лекарственную непереносимость у пациента и многое другое. Держать это в памяти и своевременно принимать правильные решения становится все сложнее [1, 2]. Как следствие, возникают врачебные ошибки, приводящие к негативным последствиям.

В настоящее время активно внедряются всевозможные *медицинские информационные системы* (МИС). Однако такие системы не имеют функциональных возможностей поддержки

принятия врачебных решений и формализации медицинских клинических документов и процессов [3]. Во многих МИС отсутствует или лишь частично реализован формализованный ввод *электронной медицинской карты* (ЭМК) пациента. Большинство из них являются системами поддержки организационных процессов больницы [4]. Для сокращения же врачебных ошибок нужны *системы поддержки принятия врачебных решений* (СППВР) [5, 6]. Однако системы поддержки таких процессов практической деятельности врача, как, например, диагностика или мониторинг лечения, могут обрабатывать истории болезни только в формализованном виде.

Ожидается, что интеграция СППВР в МИС позволит значительно сократить количество врачебных ошибок в принятии решений. Это является сложной задачей, поскольку в СППВР обрабатываются полностью формализованные истории болезни или ЭМК пациентов.

Наиболее популярны СППВР, основанные на использовании баз знаний, формально пред-

ставляющих интеллектуальную медицинскую информацию, необходимую для принятия решений.

Большинство клинических данных вводится в МИС в неформализованном виде, что является серьезным ограничением для их интеграции с СППВР. При этом формализация ЭМК представляет собой отдельную большую задачу.

Одно из решений указанной проблемы – использование онтологического подхода. Медицинская онтология охватывает множество терминов этой предметной области и задает семантическую структуру для формирования медицинской информации. ЭМК или истории болезни пациентов формируются по этой фиксированной структуре, каждый элемент которой имеет указанный смысл. Базы знаний заполняются конкретными знаниями, например, о диагностике заболеваний, и ограничены в объемах.

Активное развитие мобильных технологий располагает к тому, чтобы СППВР были адаптированы и к мобильным устройствам. В ряде проектов мобильные приложения интегрированы в процесс наблюдения за пациентом, предоставляя возможность упростить самоконтроль, прием препаратов и давая предварительную аналитику внесенных данных [7, 8].

Достижения в цифровизации медицинских процессов не уменьшили недостаток полезных мобильных сервисов для врача (особенно важно, когда врач находится непосредственно у постели больного, где использование стандартной МИС невозможно). Несмотря на наличие различных МИС, имеется необходимость в разработке технологии как создания прикладных систем диагностики заболеваний для мобильных телефонов, так и удаленного взаимодействия между сервисами диагностики, осуществляющими анализ на серверах разработчика, и мобильными телефонами.

Основные проблемы внедрения СППВР на мобильные устройства связаны, во-первых, с большим объемом обрабатываемых данных (не все мобильные устройства имеют достаточные вычислительные ресурсы), во-вторых, с необходимостью такого способа формирования пользовательского интерфейса, как его автоматическая генерация.

Сложность генерации медицинского интерфейса определяется необходимостью именно формализованного ввода данных, общее количество которых (жалобы, результаты объективного, лабораторного и инструментального ис-

следований) велико (десятки тысяч различных наблюдений), а конкретный набор заранее не известен.

Так, актуальна задача по реализации пользовательского интерфейса для ввода данных в ЭМК пациента. Интерфейс должен позволить вносить в систему формализованные данные, не зависящие ни от конкретного наполнения, ни от количества записей.

В данной работе представлен подход к генерации динамического интерфейса мобильного приложения для заполнения формализованных данных истории болезни пациента на основе онтологического подхода. Под динамическим интерфейсом будем понимать интерфейс, который не зависит от конкретного количества возможных данных для ввода информации в ЭМК, а сами данные (жалобы, инструментальные и лабораторные исследования, объективное состояние и др.) могут изменяться в зависимости от заболевания без изменения программного кода интерфейса.

Облачная платформа разработки интеллектуальных сервисов

В Институте автоматизации и процессов управления ДВО РАН реализована облачная платформа IACPaas, которая удаленно доступна (<https://iacpaas.dvo.ru>) пользователям для разработки, управления и использования интеллектуальных ресурсов. Платформа поддерживает модели облачных вычислений PaaS и SaaS, предлагает средства поддержки разработки прикладных интеллектуальных облачных сервисов с реализацией решателей задач как совокупности программных единиц на Java. Платформа IACPaas отличается от остальных облачных платформ тем, что поддерживает свойство объяснения полученных результатов.

Одной из возможностей платформы является формирование знаний и данных на основе онтологического подхода. Данный подход предполагает сначала создание онтологии данных и знаний, а затем формализацию информационных ресурсов (баз знаний, данных), следуя структуре, заданной в онтологии. Такой подход позволяет хранить данные в формализованном виде для их последующей обработки. Онтология единожды задает структуру, а базы знаний постоянно заполняются различными знаниями в соответствии с этой структурой.

Так, на платформе уже сформированы различные базы знаний, в том числе создан спра-

вочник медицинской терминологии и наблюдений, который используется в сервисах ввода данных о пациенте, а также формирования баз знаний для интеллектуальных объясняющих сервисов.

СППВР на платформе основаны на использовании баз знаний, они строятся по онтологии, их базы знаний хранятся в иерархических структурах, задаваемых онтологией, и наполняются терминами из справочника медицинских наблюдений.

Формализация медицинских клинических документов

Для наполнения ЭМК создана семантическая структура (одна из частей медицинской онтологии), в которой заданы семантически значимые отношения между идентификатором пациента и основными характеристиками его состояния и истории развития, включая множество номеров задокументированных историй болезни. Для отдельной истории болезни заданы семантически значимые отношения с ее клинически важными элементами. Все это описывает состав и формат документа, ограничения и правила формирования (<http://www.swsys.ru/uploaded/image/2024-3/2.jpg>), что позволяет иметь формализованные данные пациентов (на платформе IASaaS), пригодные для компьютерной обработки. По данному формату и онтологии на платформе реализовано множество ресурсов – коллекций и архивов ЭМК, отдельных историй болезни по целому спектру нозологий, взятых из публикаций и реальной практики.

Сами ЭМК содержат конкретные данные о пациенте, которые вносятся пользователями приложения – врачами. Фрагмент ЭМК пациента на платформе IASaaS:

- ▼ Боль в грудной клетке – [Признак]
 - ▶ Характер – [Характеристика]
- ▼ Общая слабость – [Признак]
 - ▶ Выраженность – [Характеристика]
- ▼ Объективное состояние
 - ▼ Общий осмотр
 - ▶ Индекс массы тела – [Признак]
 - ▶ Артериальное давление – [Признак]
 - ▶ Положение больного – [Признак]
 - ▼ Осмотр по системам
 - ▼ Система кровообращения
 - ▼ Фракция выброса левого желудочка – [Признак]
 - ▼ Числовые значения
 - 57.0%

▶ Частота сердечных сокращений – [Признак]

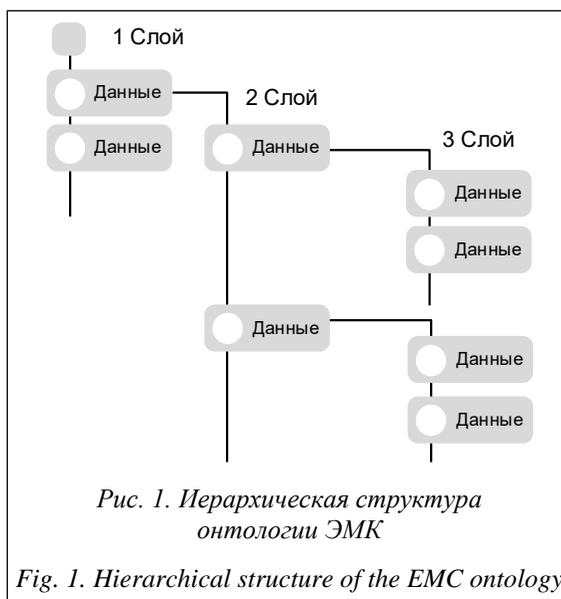
Для удаленной диагностики пациентов с мобильных телефонов разумно ориентироваться на работу с ЭМК. Через мобильный интерфейс можно заполнить ее конкретными данными, основываясь на структуре, заданной в онтологии карты. Формализованное заполнение данных в ЭМК необходимо для последующей диагностики на платформе IASaaS.

Подход к решению проблемы генерации пользовательского интерфейса ЭМК пациента

Структура ЭМК (в рамках онтологии медицины), а также содержащихся в ней историй болезни состоит из множества различных элементов, таких как «Анамнез жизни», «Паспортная часть», «Результаты исследований» и др. Из-за ограничений в размере экрана смартфона невозможно создать удобный интерфейс заполнения ЭМК и текущей истории болезни сразу для всей содержательной структуры, описанной в онтологии. Поэтому был предложен иной подход к генерации интерфейса.

Имея ограничения мобильной платформы (небольшой размер экранов и большая структура онтологии ЭМК), единственно верным способом отображения интерфейса является визуализация посредством вертикальных срезов (слоев), выбираемых по уровням вершин онтологии (рис. 1).

При заполнении конкретной ЭМК пациентов на основе рассматриваемой онтологии можно выделить три основных типа элементов:



- данные (элементы, не имеющие дочерних элементов в структуре онтологии);
- секции (элементы, имеющие некоторое множество дочерних элементов в структуре онтологии);
- элементы онтологии (незаполненные элементы ЭМК).

Каждый слой, отображаемый в приложении, визуализирует эти основные типы элементов:

▼ ИБ Тест 2

19.04.2023-11:31:00 000 – [дата обращения]	Данные
▶ Паспортная часть	Секции
▶ Жалобы	
▶ Объективное состояние	
▶ Результаты компьютерного назначения лечения	
▶ [Анамнез жизни]	Элементы онтологии
▶ [История настоящего заболевания]	
▶ [Исследования лабораторные]	
▶ [Исследования инструментальные]	

Перед визуализацией данных в виде сгенерированного интерфейса их необходимо получить с платформы IASPaas. Взаимодействие внешних систем с платформой IASPaas происходит по реализованной на ней архитектуре удаленного взаимодействия (API).

Так, любая внешняя система, в том числе мобильное приложение, может удаленно получать ресурс с платформы, изменять ресурс, запускать сервис и т.д.

Мобильное приложение формирует запрос на получение ресурса к API платформы, в котором указываются путь до экспортируемого ресурса и тип ресурса.

В ответ на запрос с платформы придет JSON-структура запрашиваемого ресурса. Это ресурс «Архив ИЭМК v.4», созданный по онтологии, в нем содержатся конкретные данные о медицинских картах и историях болезни пациентов, о состоянии их здоровья, заболеваниях и результатах лечения:

```
{
  "title": "Архив ИЭМК v.4",
  "path": "democardio@mail.ru/Мой
  Фонд/Загрузки/ Архив ИЭМК v.4$;",
  "name": "Архив ИЭМК v.4",
  "successors": [
    {
      "name": "ЭМК-q0e11",
```

```
"type": "НЕТЕРМИНАЛ",
"meta": "Онтология электронной
медицинской карты V.4",
"successors": [ ]
},
{
  "name": "ЭМК-med-books.by",
  "type": "НЕТЕРМИНАЛ",
  "meta": "Онтология электронной
медицинской карты V.4",
  "successors": [ ]
},
].
```

Поскольку глубина ресурса ограничена лишь структурой онтологии и может достигать больших размеров, передавать такой ресурс целиком по API долго, поэтому был предложен подход к получению ресурса по отдельным слоям.

При получении данных создаются и визуализируются компоненты пользовательского интерфейса данного слоя.

После отображения интерфейса корневого слоя ЭМК в приложении можно передвигаться к следующему слою при нажатии на интересующий слой (<http://www.swsys.ru/uploaded/image/2024-3/3.jpg>).

Любой переход от слоя к слою данных через приложение инициирует отправку запроса на получение данных слоя, после получения данных с запроса по ним генерируется и визуализируется интерфейс (рис. 2).

Генерация динамического интерфейса на основе слоя данных ЭМК происходит в два

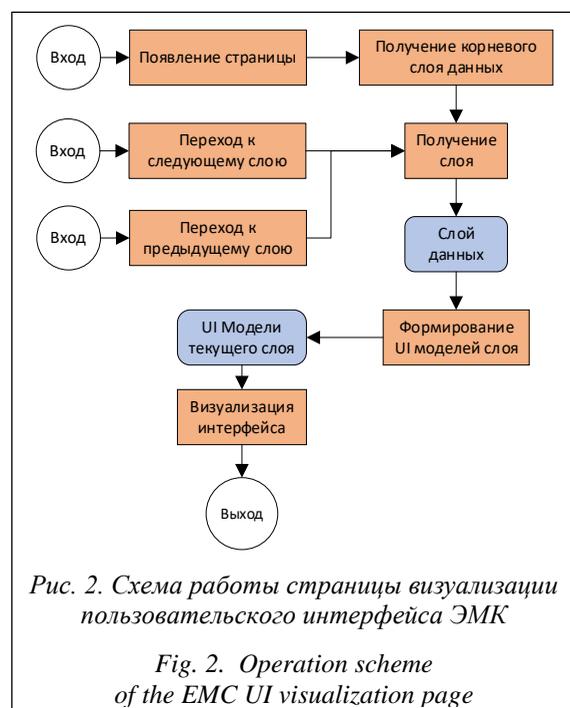


Рис. 2. Схема работы страницы визуализации пользовательского интерфейса ЭМК

Fig. 2. Operation scheme of the EMC UI visualization page

этапа: на первом этапе по данным ЭМК формируются модели интерфейса, а на втором по моделям интерфейса формируется и визуализируется сам пользовательский интерфейс.

Технология динамической генерации пользовательского интерфейса ЭМК

Интерфейс должен позволять заполнять данные в ЭМК пациента в формализованном виде. Распространено мнение, что не все данные ЭМК можно представить формализованно, в частности, во внедряемых на рабочие места врачей МИС анамнез жизни и даже анамнез заболевания принципиально представляются в виде текстов на естественном языке. Однако на медицинском портале платформы IASPaas эксплуатируется обширный справочник медицинской терминологии и наблюдений, с применением которого все факты о пациенте структурируются (<http://www.swsys.ru/uploaded/image/2024-3/4.jpg>).

Сам интерфейс не должен зависеть от конкретных данных и их количества.

Необходимо также обеспечить генерацию интерфейсов под разные структуры из онтологии ЭМК. Способ генерации для каждого элемента конкретной ЭМК зависит от структуры в онтологии. Таким образом, интерфейс можно генерировать на основе структур онтологии.

Алгоритм генератора интерфейса состоит из трех основных шагов:

- определение принадлежности элемента из ЭМК к элементу пользовательского интерфейса;
- формирование модели пользовательского интерфейса из модели элемента ЭМК;
- формирование самого интерфейса в зависимости от модели пользовательского интерфейса.

Схема работы алгоритма представлена на рисунке 3.

Элементы отображаемого слоя ЭМК обрабатываются конструктором. Каждый элемент проходит проверку на соответствие элементу пользовательского интерфейса, при совпадении по элементу ЭМК строится модель пользовательского интерфейса. Пройдя проверку, все модели слоя ЭМК становятся моделями интерфейса, по которым впоследствии будет отображен сам интерфейс для наполнения конкретными жалобами и другими признаками пациента («Головная боль», «Беспокойство», «Боль в грудной клетке», «Температура» и т.п.) и фактами («Пол», «Возраст»). Конкретные названия всех признаков и фактов и их составных частей, а также все возможные значения хранятся в справочнике медицинской терминологии и наблюдений IASPaas для описания с их помощью конкретных состояний пациента.

Структура этих терминов предписывает возможность быть описанным как качественный, числовой или составной. Качественным считается простой признак, конечными элементами которого является список строковых значений. Числовой – это простой признак, имеющий в качестве конечных элементов один вещественный элемент значения и один строковый элемент единицы измерения. Составной признак – комплекс элементов или характеристик, аналогичных простому признаку.

Содержимое справочника медицинской терминологии и сохраняемую ЭМК мобильное приложение получает с платформы IASPaas по протоколу HTTP посредством отправки соответствующих запросов. Такой способ получения информации доступен благодаря реализации удаленного взаимодействия на платформе [9].

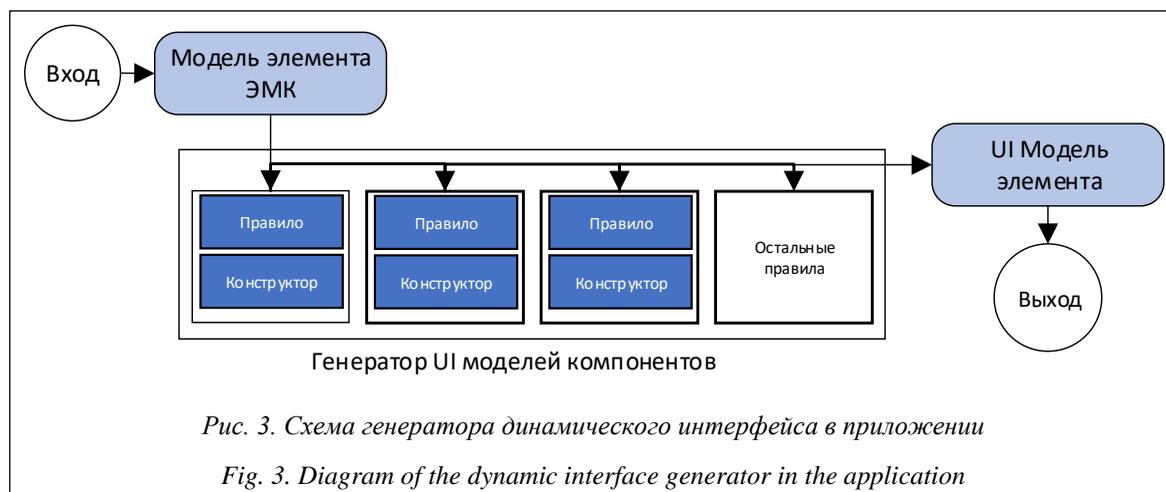


Рис. 3. Схема генератора динамического интерфейса в приложении

Fig. 3. Diagram of the dynamic interface generator in the application

Все признаки и факты проходят через конструктор интерфейсов, где в первую очередь определяется тип элемента – факт или признак, а далее – простой это элемент или составной. Эти два свойства создают четыре варианта элемента: простой факт, составной факт, простой признак, составной признак.

Каждый вариант имеет проверку на соответствие и собственный конструктор, который в случае успешной проверки сможет создать необходимую модель интерфейса из модели данных платформы IASaaS (рис. 4).

Сформированные модели пользовательского интерфейса состоят только из тех данных, которые необходимы для отображения конкретных элементов интерфейса, то есть они отделяют данные базы от данных, необходимых для отображения:

```

{
  "Name": "дата обращения",
  "Value": "10.20.2022-12:01:33.000",
  "ComponentType": "Date",
}

{
  "Name": "Боль в глазу",
  "ComponentType": "AttributeFact",
  "Characteristics": [{
    "Name": "Присутствие",
    "Value": ["Имеется"]
  }], {
    "Name": "Локализация",
    "Value": ["Правый глаз", "Левый глаз"]
  } ]
}.
{
  "Name": "Жалобы",
  "ComponentType": "Section",

```

```

}

{
  "Name": "Беспокойство",
  "ComponentType": "AttributeFact",
  "Value": ["Имеется"]
}

```

Сформированные модели пользовательского интерфейса передаются фабрике формирования интерфейса. Список моделей циклом обрабатывается в фабрике, где в зависимости от типа модели интерфейса строится его соответствующий элемент (рис. 5).

По структуре признаков и фактов этой терминологии формируются модели интерфейса, которые далее наполняются конкретными данными по конкретным заболеваниям или фактам.

Интерфейс, сгенерированный по своим моделям, визуализируется в мобильном приложении (<http://www.swsys.ru/uploaded/image/2024-3/5.jpg>). Реализованное мобильное приложение, используя инструменты взаимодействия с внешними системами, обменивается данными с интеллектуальной платформой IASaaS [10].

Таким образом, медицинская онтология позволяет реализовать конструктор интерфейса для мобильного приложения, не зависящий от наполнения, благодаря чему приложение легко адаптируется к любому количеству признаков или фактов и других элементов ЭМК.

С помощью создаваемых приложений врачи могут заполнять ЭМК пациентов в формализованном виде и проводить их диагностику как поддержку в принятии более точных и обоснованных решений, что на сегодня очень актуально [2, 5, 6]. Это отличает разрабатываемые в рамках предложенной технологии приложения с возможностью ввода любых текущих данных, готовых для машинной интерпретации, от распространенных разработок с фиксированным шаблоном для ввода данных пациента [7, 8]. Их интерфейсы позволяют легко вводить формализованные данные в условиях развиваемой предметной области: база терминов может расширяться, в структуру ЭМК могут добавляться разделы.

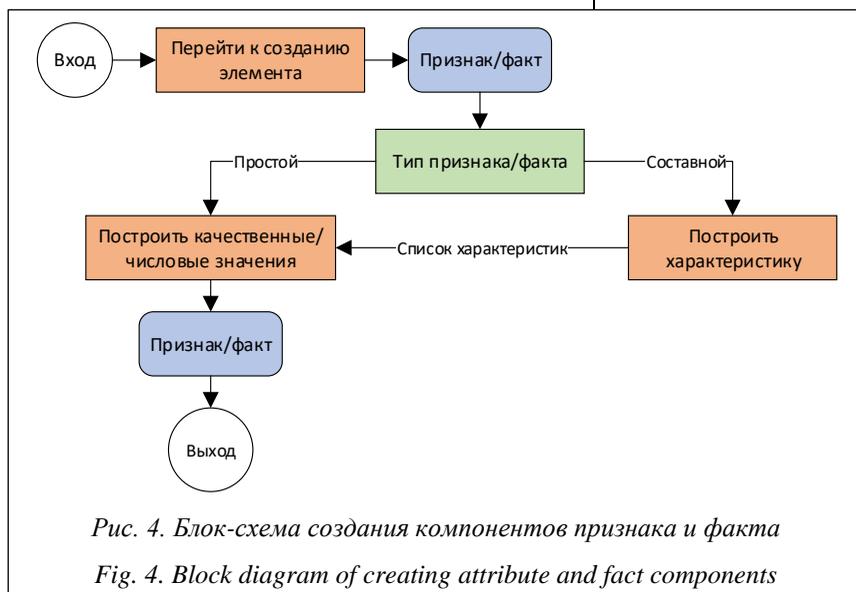


Рис. 4. Блок-схема создания компонентов признака и факта

Fig. 4. Block diagram of creating attribute and fact components

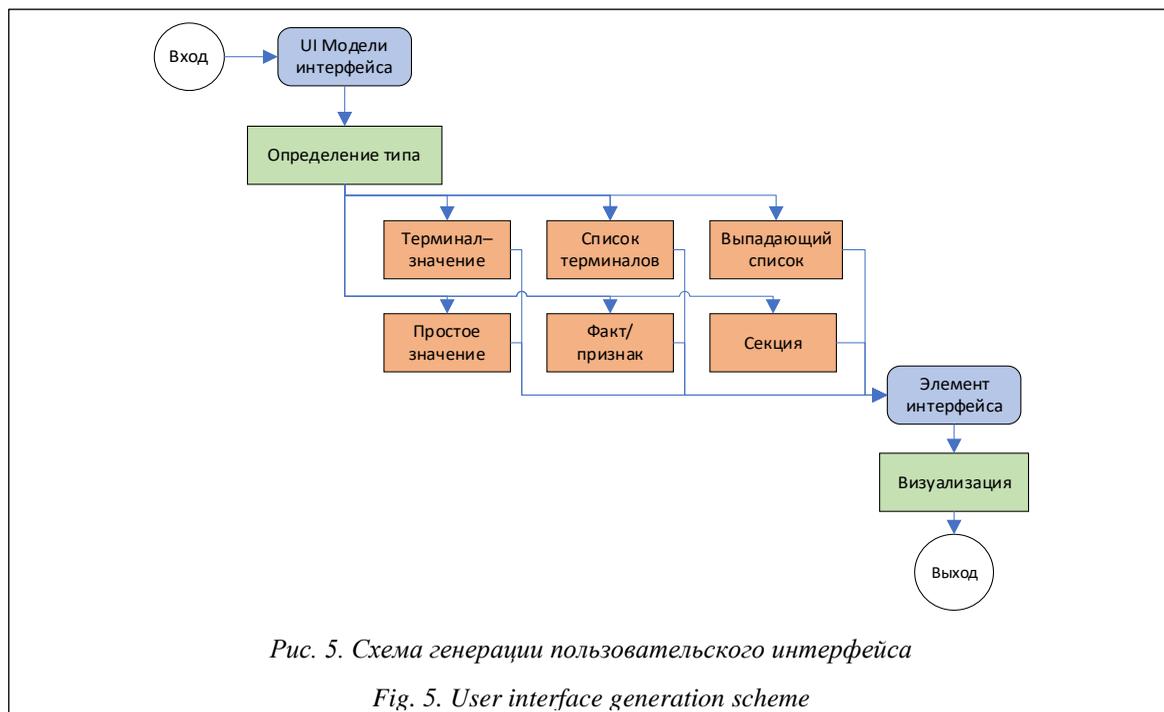


Рис. 5. Схема генерации пользовательского интерфейса

Fig. 5. User interface generation scheme

Мобильная оболочка по диагностике заболеваний

Подход и реализованная на платформе IASaaS оболочка для создания СППВР по диагностике заболеваний независимо от их группы описаны в работе [10]. С использованием этой оболочки реализованы различные системы и сервисы поддержки врачебных решений. В данной статье предложено расширение разработанного ранее продукта – мобильная версия оболочки.

Мобильная оболочка позволяет визуализировать интерфейс для диагностики заболеваний и для формализованного заполнения любых ресурсов, созданных по онтологии ЭМК. Приложения абсолютно не зависят от огромного разнообразия видов признаков заболеваний, фактов о пациенте, историй болезней и вариантов их диагностики.

В рамках тестирования реализованной мобильной оболочки были созданы два приложения по диагностике – вирусных и кардиозаболеваний. Оба приложения независимые и позволяют осуществлять удаленную диагностику ЭМК пациентов. Они опубликованы в магазине мобильных приложений RuStore и находятся в общем доступе.

Заключение

Внедрение СППВР в практическую медицину является современным трендом. Актив-

ное развитие мобильных технологий способствует адаптации СППВР к мобильным устройствам, чтобы врачи у постели больного могли оперативно вводить всю необходимую информацию и получать помощь при постановке диагноза и оценке состояния пациента.

При этом проблемами для мобильных устройств являются возможность вводить большой объем информации в формализованном виде и взаимодействие с ресурсоемкими компонентами СППВР (на удаленном сервере). Учитывая, что история болезни и ЭМК содержат много разделов с вложенной структурой, а количество потенциальных данных для ввода – десятки тысяч, методы их преодоления исключительно актуальны.

В работе описан подход к генерации адаптивных пользовательских интерфейсов, который генерируется на основе структуры ЭМК, а база медицинской терминологии является источником терминов для наполнения сведений через интерфейс. Структура ЭМК и база медицинской терминологии реализованы на том же удаленном сервере, где и компоненты СППВР, готовые к приему передаваемой из мобильного интерфейса информации.

Такой подход позволил добиться

- гибкости в создании интерфейса в приложении;
- независимости от медицинской терминологии;
- повторного использования интерфейса;
- дальнейшего легкого сопровождения;

– сохранения способности автоматической генерации и визуализации интерфейса при дополнении баз знаний (или расширении структуры онтологии).

Данный подход к генерации динамического пользовательского интерфейса значительно снижает нагрузку на разработчиков ПО в части создания и поддержки пользовательских интерфейсов. Реализованный интерфейс позволяет легко вводить данные в формализованном, пригодном для обработки виде даже в условиях развешиваемых словарей.

Описанная технология генерации пользовательского интерфейса успешно реализована в мобильной оболочке по диагностике заболеваний. Оболочка не зависит от группы заболеваний и позволит создавать множество сервисов диагностики.

Представленный подход к генерации динамического интерфейса может быть использован и в других сервисах, для которых характерны большие объемы вводимой информации и графовое представление исходных данных.

Список литературы

1. Гусев А.В., Зарубина Т.В. Поддержка принятия врачебных решений в медицинских информационных системах медицинской организации // Врач и информационные технологии. 2017. № 2. С. 60–72.
2. Мухаметшина В.Р. Системы поддержки принятия решений в медицине // Научные исследования молодых ученых: сб. ст. 2022. С. 22–24.
3. Вафин Р.Р., Насыров Р.В. Разработка интерфейса системы поддержки принятия врачебных решений для комплексного обеспечения медицинского технологического процесса // Методы компьютерной диагностики в биологии и медицине. 2020. С. 64–67.
4. Михеев А.Е., Фохт О.А., Хаткевич М.И. Один из подходов к формализации процесса внедрения МИС в медицинской организации // Врач и информационные технологии. 2018. № 5. С. 46–62.
5. Гончарова А.Б., Сергеева Е.И. Построение системы поддержки принятия решений в медицине // Интеграция наук. 2019. № 1. С. 272–274.
6. Фролов С.В., Куликов А.Ю., Остапенко О.А., Стрыгина Е.В. Системы поддержки врачебных решений в медицине // Научный журнал. 2018. № 9. С. 9–16.
7. Omboni S. Connected health in hypertension management. *Front. Cardiovasc. Med.*, 2019, vol. 6, art. 76. doi: 10.3389/fcvm.2019.00076.
8. Postel-Vinay N., Steichen O., Pebelier E. et al. Home blood pressure monitoring and e-Health: Investigation of patients' experience with the Hy-Result system. *Blood Pressure Monitoring*, 2020, vol. 25, no. 3, pp. 155–161. doi: 10.1097/MBP.0000000000000436.
9. Тимченко В.А., Грибова В.В., Федорищев Л.А., Москаленко Ф.М. Технология взаимодействия сервисов облачной платформы IASPaas с внешним программным обеспечением // Программные продукты и системы. 2018. Т. 31. № 2. С. 233–238. doi: 10.15827/0236-235X.122.233-238.
10. Грибова В.В., Федорищев Л.А. Онтологический подход к генерации адаптивных WIMP-интерфейсов редакторов баз знаний // КИИ-2018: тр. конф. 2018. Т. 1. С. 19–26.

Dynamic generation of a user interface for a medical mobile application based on an ontological approach

Valeriya V. Gribova ¹✉, Aleksandr Ya. Lifshits ¹, Philipp M. Moskalenko ¹,
Elena A. Shalfeeva ¹, Nikita E. Shevchenko ²

¹ Institute of Automation and Control Processes of the Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690041, Russian Federation

² Vladivostok State University, Vladivostok, 690014, Russian Federation

For citation

Gribova, V.V., Lifshits A.Ya., Moskalenko, Ph.M., Shalfeeva, E.A., Shevchenko, N.E. (2024) 'Dynamic generation of a user interface for a medical mobile application based on an ontological approach', *Software & Systems*, 37(3), pp. 431–439 (in Russ.). doi: 10.15827/0236-235X.142.431-439

Article info

Received: 19.12.2023

After revision: 21.02.2024

Accepted: 24.02.2024

Abstract. Medical decision support systems are designed to help doctors make timely and correct decisions. This is particularly important to use such systems remotely from medical centers and clinics, at the patient's bedside. The structure of an electronic medical record has many hierarchically presented sections, and input information for such records is assessed by tens of thousands of possible terms and their meanings. Taking this into account, it is relevant to develop new methods for generating mobile user interface for filling out electronic medical records. This paper presents a method for

generating an adaptive interface for filling out patient's electronic record in mobile applications based on an ontological approach. The approach assumes visualization by vertical layers selected by node levels in the hierarchical semantic structure of an electronic medical record document. The described generation technology is tested on a diagnostic mobile shell. The result is also a new technology of generating user interface and a software shell for combining diagnostic services for arbitrary groups of diseases for mobile devices. The practical significance of the technology consists in flexibility of the presented interface generation method: it can be used to create adaptive user interfaces in other subject areas and tasks, which are characterized by large volumes of input complex information.

Keywords: mobile application, knowledge base, medicine, interface, dynamic interface generation, ontology, intelligent systems

Acknowledgements. The work was supported by the Ministry of Education and Science of the Russian Federation (task no. FFWF-2021-0004)

References

1. Gusev, A.V., Zarubina, T.V. (2017) 'Clinical decisions support in medical information systems of a medical organization', *Medical Doctor and Information Technologies*, (2), pp. 60–72 (in Russ.).
2. Mukhametshina, V.R. (2022) 'Decision support systems in medicine', *Proc. Conf. Sci. Research of Young Scientists*, pp. 22–24 (in Russ.).
3. Vafin, R.R., Nasyrov, R.V. (2020) 'Development of the interface of the system for supporting medical decision-making for complex provision of medical technological process', *Computer Diagnostic Methods in Biology and Medicine*, pp. 64–67 (in Russ.).
4. Mikheev, A.E., Foght, O.A., Khatkevich, M.I. (2018) 'One of approaches to formalization of the HIS deployment in healthcare institution', *Medical Doctor and Information Technologies*, (5), pp. 46–62 (in Russ.).
5. Goncharova, A.B., Sergeeva, E.I. (2019) 'Medicine decision support system design', *Integration of Sci.*, (1), pp. 272–274 (in Russ.).
6. Frolov, S.V., Kulikov, A.Yu., Ostapenko, O.A., Strygina, E.V. (2018) 'Medical decision support systems in medicine', *Scientific J.*, (9), pp. 9–16 (in Russ.).
7. Omboni, S. (2019) 'Connected health in hypertension management', *Front. Cardiovasc. Med.*, 6, art. 76. doi: 10.3389/fcvm.2019.00076.
8. Postel-Vinay, N., Steichen, O., Pebelier, E. et al. (2020) 'Home blood pressure monitoring and e-Health: Investigation of patients' experience with the Hy-Result system', *Blood Pressure Monitoring*, 25(3), pp. 155–161. doi: 10.1097/MBP.0000000000000436.
9. Timchenko, V.A., Gribova, V.V., Fedorishchev, L.A., Moskalenko, F.M. (2018) 'Technology of interaction of IACPaaS cloud platform services with external software', *Software & Systems*, 31(2), pp. 233–238 (in Russ.). doi: 10.15827/0236-235X.122.233-238.
10. Gribova, V.V., Fedorishchev, L.A. (2018) 'An ontological approach to the generation of adaptive WIMP interfaces of knowledge base editors', *Proc. RCAI-2018*, 1, pp. 19–26 (in Russ.).

Авторы

Грибова Валерия Викторовна¹,

член-корреспондент РАН, д.т.н.,

зам. директора

по научной работе

gribova@iacp.dvo.ru

Лифшиц Александр Яковлевич¹,

к.т.н., ведущий инженер-программист,

mmvb@iacp.dvo.ru

Москаленко Филипп Михайлович¹,

к.т.н., старший научный сотрудник,

philipmm@iacp.dvo.ru

Шалфеева Елена Арэфьевна¹,

д.т.н., ведущий научный сотрудник,

shalf@iacp.dvo.ru

Шевченко Никита Евгеньевич²,

магистр, shev.nikita@mail.ru

Authors

Valeriya V. Gribova¹,

Corresponding Member

of RAS, Dr.Sci. (Engineering),

Deputy Director on Scientific Work,

gribova@iacp.dvo.ru

Aleksandr Ya. Lifshits¹, Cand. of Sci. (Engineering),

Leading Engineer-Programmer,

mmvb@iacp.dvo.ru

Philipp M. Moskalenko¹,

Cand. of Sci. (Engineering),

Senior Researcher, philipmm@iacp.dvo.ru

Elena A. Shalfeeva¹,

Dr.Sci. (Engineering),

Leading Researcher, shalf@iacp.dvo.ru

Nikita E. Shevchenko², Master of Sci.,

shev.nikita@mail.ru

¹ Институт автоматизации и процессов управления ДВО РАН, г. Владивосток, 690041, Россия

² Владивостокский государственный университет, г. Владивосток, 690014, Россия

¹ Institute of Automation and Control Processes of the Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690041, Russian Federation

² Vladivostok State University, Vladivostok, 690014, Russian Federation

Автоматизация проверки семантической составляющей текстовых ответов обучающихся в цифровой образовательной платформе

А.Г. Леонов^{1, 2, 3, 4}, Н.С. Мартынов², К.А. Машенко^{1, 2, 4}✉,
А.А. Холькина⁵, А.В. Шляхов²

¹ МГУ им. М.В. Ломоносова, г. Москва, 119991, Россия

² ФГУ ФНЦ НИИСИ РАН, г. Москва, 117218, Россия

³ Институт детства Московского педагогического государственного университета (МПГУ), г. Москва, 119435, Россия

⁴ Институт информационных систем Государственного университета управления, г. Москва, 109542, Россия

⁵ Национальный исследовательский технологический университет «МИСиС», г. Москва, 119049, Россия

Ссылка для цитирования

Леонов А.Г., Мартынов Н.С., Машенко К.А., Холькина А.А., Шляхов А.В. Автоматизация проверки семантической составляющей текстовых ответов обучающихся в цифровой образовательной платформе // Программные продукты и системы. 2024. Т. 37. № 3. С. 440–452. doi: 10.15827/0236-235X.142.440-452

Информация о статье

Группа специальностей ВАК: 2.3.5

Поступила в редакцию: 09.04.2024

После доработки: 04.05.2024

Принята к публикации: 14.05.2024

Аннотация. Цель настоящего исследования заключается в разработке и применении системы автоматизированной проверки ответов студентов на открытые текстовые вопросы в рамках цифровой образовательной платформы. Разработанная система предназначена для существенного снижения нагрузки на преподавателя и минимизации ошибок первого и второго рода в процессе проверки ответов студентов. Методология исследования включает использование современных методов извлечения информации, текстового анализа и сопоставления концепций. С помощью построения графа синтаксического разбора предложения формируются синтаксические деревья, после чего с учетом особенностей русского языка проводится анализ отношений между их узлами, например, исключение двойных отрицаний, извлечение именованных сущностей (NER) и распознавание лексических взаимосвязей. Для сравнения деревьев используется поиск оптимального пути редактирования, реализованный с применением динамического программирования. Результатом исследования является разработанная авторами система, способная оценивать ответы студентов с учетом смысловой глубины содержания текстов, что может обеспечить объективность и прозрачность оценки знаний. Кроме того, она позволяет снизить нагрузку на преподавателей, автоматизируя процесс проверки ответов, а возможность параметрической настройки сценария работы системы помогает подстроить ее под поставленные преподавателем специфические требования к ответам. Система способна игнорировать грамматические ошибки, так как ее фокус направлен на отслеживание расхождений или сходств в смысловой составляющей текстов.

Ключевые слова: цифровая образовательная платформа Мирера, смысловой анализ, семантический анализ, система оценки ответов, образовательные технологии, автоматизированная проверка

Благодарности. Работа выполнена в рамках госзадания ФГУ ФНЦ НИИСИ РАН по теме № FNEF-2024-0001 (1023032100070-3-1.2.1)

Введение. Повышение эффективности процесса обучения достигается не только за счет технических и дидактических средств, но и с помощью организации обратной связи с обучаемым, которая обеспечивается контролем знаний в интеллектуальных обучающих системах. Контроль знаний в таких системах, как правило, осуществляется автоматизированными тестирующими компонентами [1]. Тестовые задания, предлагаемые обучающимся, можно разделить на два типа – открытые и закрытые. Первые подразумевают конструируемый ответ на вопрос, а вторые – выбор одного или нескольких ответов из предложенного списка. Основная проблема закрытых вопросов заклю-

чается в том, что они проверяют, насколько хорошо обучающийся запомнил ответ, но не показывают процесс его размышлений. Проверить логику и степень понимания материала таким способом невозможно. Открытые вопросы, напротив, заставляют студента поразмышлять над вопросом, составить конструктивный ответ. Это наиболее полная и естественная форма контроля знаний. Проблемой открытых вопросов является огромная рутинная нагрузка на преподавателя при проверке студенческих ответов. Основная задача, которую в данном процессе должна выполнять интеллектуальная система, заключается в выявлении и проверке смысла в студенческом ответе.

Его правильность определяется путем сравнения этих мыслей с извлекаемыми из эталонного ответа преподавателя. Преимуществом автоматизации проверки является не только снижение нагрузки на преподавателя, но и прозрачность и справедливость результата, поскольку исключается фактор предвзятости.

Обзор существующих автоматизированных систем проверки ответов на открытые вопросы

В ходе анализа существующих автоматизированных систем проверки были выявлены четыре основные категории методов и подходов, на которые опираются архитектурные решения таких систем.

Первая категория использует метод извлечения информации. Он является разновидностью информационного поиска и заключается в выделении структурированной информации из неструктурированного текста. Описанный в работе [2] алгоритм один из самых первых в области проверки свободных ответов на английском языке с помощью извлечения информации. При проверке авторы строят графы разбора ответа обучающегося и эталонного ответа преподавателя, из которых впоследствии извлекается информация и сравнивается между собой. Важным исследованием в данной области является работа [3]. В ней авторы описывают систему, оценивающую ответ по сопоставлению фактов, извлеченных из текстов по определенным грамматическим правилам. Подход к сравнению построенных по извлеченной информации графов изложен в [4]. Рассматривая соединенные подграфы с графовым ядром, авторы получили низкоразмерное векторное представление исходного графа структуры. Затем сравнивались подграфы и измерялась их сопоставимость. Классические подходы текстовой репрезентации в виде графа представлены в [5].

Вторая категория опирается на сопоставление концепций. Метод основан на выделении списка минимальных концепций (ключевых понятий) из ответа обучающегося и эталона и на их последующем сравнении. Именно этим подходом пользовались при создании системы в работе [6]. С помощью фреймворка Generalised Phrase Structure Grammar описываются синтаксис и семантика языка, затем ответ обучающегося и эталонный ответ преподавателя разбиваются на списки минимальных концепций. Получившиеся концепции подсчитыва-

ются, и с учетом веса для каждой из них выставляется оценка.

Третья категория использует при проверке ответов лингвистические корпуса (цельный свод текстов для исследования языка). В работе [7] описываются основанные на корпусах классические подходы к нахождению схожести текстов. При работе с большими текстами преимущественно используются лингвистические корпуса. Для работы с маленькими текстами в качестве корпуса используют эталонные ответы преподавателя. При проверке этим методом составляются N-граммы (последовательности из N элементов) слов. В работе [8], где применялся данный метод, для поиска правильных отрывков из ответа обучающегося сначала используется степень зависимости N-грамм слов эталонного ответа в отрывке ответа обучающегося, а затем SVM-модель (модель, использующая метод опорных векторов) для улучшения ранжирования отрывков, включающая различные меры лексического, синтаксического и семантического сходства.

К числу быстро набирающих популярность относятся системы, использующие машинное обучение. Как правило, этот подход заключается в векторном представлении слов и использовании метрик для естественного языка, определяющих сходство текстов. В работе [9] предлагается подход, использующий различные методы обработки естественного языка и инструменты, такие как SCBOW, Wordnet, Word2vec (для векторного представления слов), косинусное сходство, мультиномиальный наивный Байес (MNB), частота слова и обратная частота документа (TF-IDF) (для определения меры сходства). Особый интерес исследователей в области обработки естественного языка вызывают языковые модели, в которых применяются рекуррентные нейронные сети и трансформеры. За счет своих архитектур эти нейронные сети понимают тексты на семантическом уровне, что позволяет им отвечать на вопросы человека, вести конструктивный диалог. Несмотря на высокие результаты таких языковых моделей, как GPT, BLOOM, LoRA, они регулярно подвергаются критике [10]. В этой связи использование подхода, основанного на языковых моделях, ставится под сомнение в системах, где критически важна цена ошибки. Помимо этого, в русском языке малейшее изменение даже порядка слов может привести к принципиально разным по смыслу ответам, что не обнаруживается моделями, основанными на машинном обучении. В работе [11] авторы

предлагают модель, работающую на графовых сверточных нейронных сетях (GCNS) и использующую деревья синтаксических зависимостей для получения семантической репрезентации предложений. С помощью этой модели слова в предложениях кодируются в семантическое векторное представление.

Зачастую в модели используется смешение обозначенных выше подходов. Например, в работе [12] авторы применяют совмещенную концепцию графа знаний (KG), иллюстрирующего отношения между концепциями, и больших лингвистических корпусов. В работе [13] описана система, совмещающая в себе составление векторных малоразмерных представлений и алгоритм Siamese Manhattan LSTM (MaLSTM) [14]. Работа этой системы была проверена авторами на открытых студенческих ответах в цифровой образовательной платформе. К комплексным относится модель из [15], использующая совместно текстовую сегментацию, определение частей речи, построение семантической матрицы и графа синтаксического разбора. В [16] рассмотрен алгоритм, в котором из текста выделяются ключевые слова с помощью анализа графлетов, а после этого применяется MNB.

В ходе изучения опыта создания систем автоматизированной проверки было отмечено, что только описанная в [3] система создавалась для обработки русскоязычного текста.

Постановка задачи и определение требований

Перед авторами данного исследования стояла задача разработки системы автоматизированной проверки любых по длине ответов на открытые вопросы. Преподавателю необходимо создать набор коротких односложных предложений с мыслями, которые должны присутствовать в ответе студента. Предложения в ответе студента могут быть любой сложности. Разрабатываемая система должна уметь анализировать ответы обучающихся на русском языке и на основе сравнения с эталонным ответом преподавателя выставлять предварительную оценку. При сравнении необходимо учитывать не только полностью совпадающие слова, но и синонимы, антонимы, должны игнорироваться формы и склонения слов, учитываться различные возможные написания дат и имен собственных, а также использование местоимений. Помимо этого, должны учитываться обычные и многоуровневые отрицания

и их влияние на итоговый смысл слов и предложения в целом. За использование антонимов и отрицаний, инвертирующих смысл написанного, при расчете оценки должны начисляться штрафы. Для автоматического выставления «задание зачтено» и «задание не зачтено» преподавателю следует определять пороги отсеечения. Так, если уровень правильности меньше нижнего порога, решение считается неверным. Если доля правильности больше верхнего порога, решение автоматически засчитывается. Если набранная оценка оказалась между двумя порогами, то требуется ручная проверка преподавателем, о чем ему приходит уведомление.

Система развернута на базе цифровой образовательной платформы Мирера, обеспечивающей хранение и сдачу решений, отображение результатов, рассылку уведомлений и пр. [17, 18].

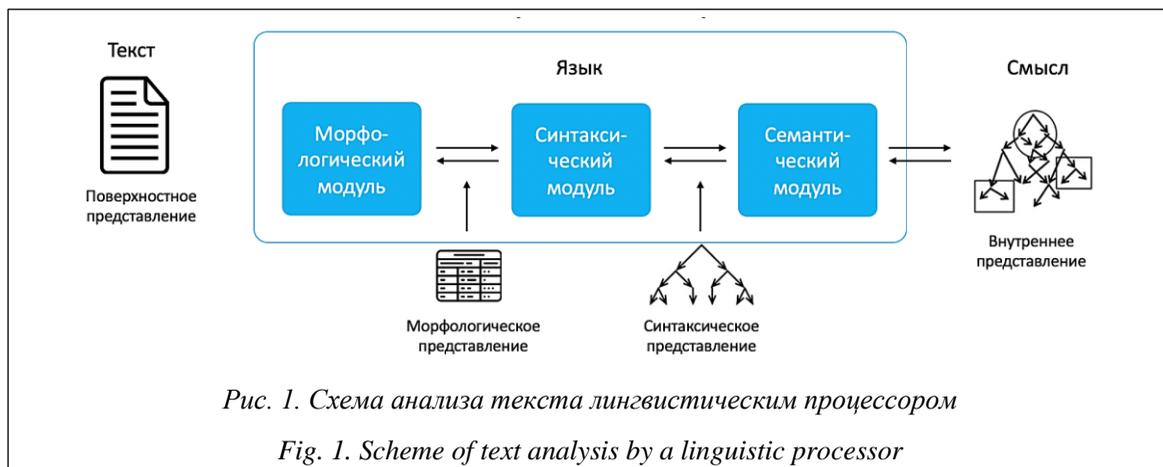
Построение графа синтаксического разбора

Решать поставленную задачу предлагается с помощью информационного поиска, комбинируя подходы построения графов разбора и сопоставления фактов. Таким образом, решение задачи сводится к синтаксическому анализу, результатом которого являются графы синтаксического разбора, а также к построению на основе этих графов деревьев и к их сравнению.

Под синтаксическим анализом в лингвистике и информатике подразумевают процесс сопоставления линейной последовательности лексем (слов, токенов) естественного или формального языка с его формальной грамматикой. Для программной реализации построения графов синтаксического разбора применяются лингвистические процессоры, основными задачами которых являются извлечение смысла из текста и его интерпретация набором формальных средств, понятных компьютеру. В процессе анализа текста лингвистический процессор выполняет морфологический и синтаксический анализ. Результатом работы лингвистического процессора является синтаксическое представление предложения в виде графа зависимостей, которое можно семантически анализировать. Схема работы лингвистического процессора графически представлена на рисунке 1.

При программной реализации компонентов лингвистического процессора используются два подхода – метод, основанный на правилах, и машинное обучение с учителем.

Основная идея первого метода заключается в создании набора правил, которые опреде-



ляют, как проставлять связи в предложении. Русский язык насчитывает огромное количество таких правил. В лингвистические процессоры, основанные на этом методе, например, в Томита-парсер, необходимо вносить эти правила самостоятельно, что требует немалых усилий и отдельной команды лингвистов. Авторы работы [3], использующие этот подход, отмечают, что заранее определить полный набор правил нельзя. По этой причине при анализе сложных предложений могут возникать проблемы.

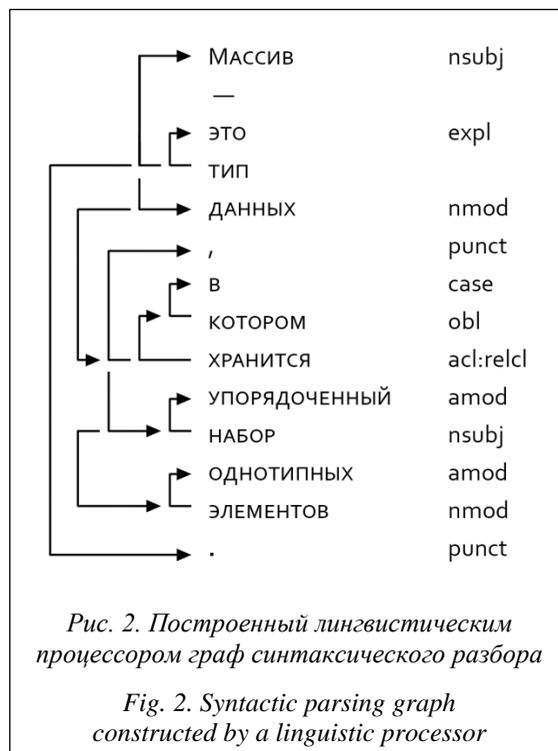
Метод, основанный на машинном обучении, заключается в обучении моделей на лингвистических корпусах, где размечена синтаксическая структура. Один из популярных проектов, содержащий в себе модели для морфологического и синтаксического анализа, добился высоких результатов в автоматическом структурировании текстов на русском языке. Речь идет о проекте Natasha (набор Python-библиотек для обработки текстов на естественном русском языке (<https://natasha.github.io/>)), программные инструменты которого позволяют решать все базовые задачи обработки естественного языка: сегментация на токены и предложения, морфологический и синтаксический анализ, лемматизация, извлечение именованных сущностей (NER) и анализ слов на принадлежность к определенным частям речи (POS-tagging). Эти модели позволили авторам собрать гибкий и эффективный лингвистический процессор, не требующий дополнительных действий от преподавателя по сравнению с инструментами, использующими метод, основанный на правилах. Отмечается проект CoreNLP (<https://stanfordnlp.github.io/CoreNLP/>) для работы с иностранными текстами.

В результате работы лингвистического процессора имеется граф синтаксического разбора

(рис. 2), в котором связи между словами определяются согласно межъязыковой типологии Universal Dependency Relations (<https://universaldependencies.org/u/dep/>).

Построение синтаксического дерева на основе графа синтаксического разбора

Сравнение синтаксических графов является вычислительно сложным алгоритмом. Это обусловлено тем, что в русском языке порядок слов нефиксированный. Так, слова могут идти абсолютно в разной последовательности, при этом смысл сказанного может как не измениться вообще, так и измениться на противоположный. Помимо этого, в синтаксическом



графе предложений русского языка могут быть циклы, а в графе одного предложения может быть несколько компонент связности. Все это приводит к тому, что даже графы похожих по своему устройству предложений могут быть абсолютно разными. Поэтому необходимо, во-первых, очистить исходный граф от частей речи и слов, не влияющих на общий смысл предложения, и, во-вторых, привести граф к дереву, что позволит применять вычислительно более быстрые алгоритмы определения степени их различия. Для работы с графами и, в частности, с деревьями авторы пользуются библиотекой NetworkX – Python-библиотека для работы с графами и сетями (<https://networkx.org/>).

Очистка циклов и объединение в одну компоненту связности

В достаточно сложных и длинных предложениях в русском языке модель построения графа синтаксического разбора из библиотеки Natasha может вернуть граф, в котором будут циклы. Чтобы удалить ребро в цикле наилучшим образом, выделяются части речи, лексически наиболее значимые в русском языке. Такими являются глаголы, существительные, имена собственные и местоимения. Поэтому для устранения циклов в структуре взаимосвязей внутри графа самым приоритетным является удаление ребра, входящего в узел, обладающий наивысшей значимостью. Если оно находится, то происходит разрыв этой взаимосвязи. Однако если такое ребро найти не удается, при возвращении в стартовую вершину удаляется последняя найденная взаимосвязь.

Таким образом, гарантируется исчезновение всех циклов внутри графа. Для предложений со сложной составной структурой несвязный граф, состоящий из нескольких деревьев, преобразуется в односвязное дерево. С этой целью выбирается главный корень, и все остальные корни становятся его наследниками. Впоследствии, если в студенческом ответе составные части сложного предложения будут указаны в другом порядке, алгоритм сравнения деревьев сможет перенаправить это ребро в другую сторону без понижения вероятности совпадения.

Наглядно продемонстрируем преобработку графа с помощью процедур избавления от циклов, соединения корней для получения односвязности и очистки графа от лишних частей речи, которые определены как лексически малозначительные. Рассматривается следующая строка: «Движения макроскопических тел

при скорости, много меньшей скорости света». На рисунке 3а показан предварительно построенный граф для данного текста, а на рисунке 3б – преобразованный исходный граф после применения к нему алгоритма предобработки. Следует отметить, что в данном примере в структуре графа был разорван цикл, соединены между собой две компоненты связности и удалены лексически малозначительные части речи.

Предобработка отрицаний

Важной частью работы являются поиск и анализ лексической отрицательности узлов графа. Отрицание может распространяться как на родительский узел (например, «совсем не яркий» подразумевает, что «не» является наследником у слова «яркий»), так и на узел наследника (например, «нет места» влечет связь, в которой «нет» будет родительским узлом для «места»). Если лексическая отрицательность распространяется от наследника к родителю, то выполняется алгоритм удаления наследника, который заключается в удалении узла отрицания с дальнейшим переприсвоением наследников. После перестроения графовой структуры родителю добавляется метка

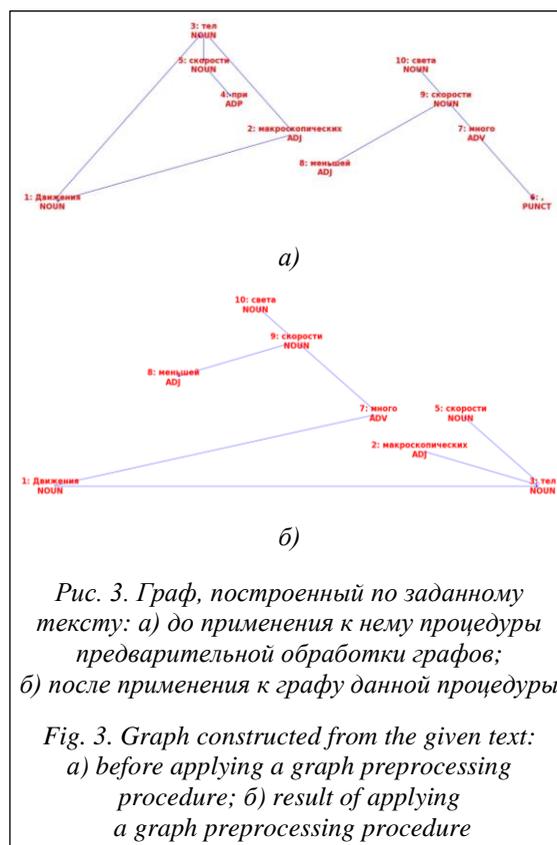


Рис. 3. Граф, построенный по заданному тексту: а) до применения к нему процедуры предварительной обработки графов; б) после применения к графу данной процедуры

Fig. 3. Graph constructed from the given text: а) before applying a graph preprocessing procedure; б) result of applying a graph preprocessing procedure

отрицательного смысла. В случае обратной ситуации, когда отрицание распространяется от родительского узла к наследнику, добавляется метка отрицательного смысла для всех наследников без удаления корневого слова.

Хорошо иллюстрирует работу алгоритма с отрицаниями граф, построенный по следующему предложению: «Я не мог не согласиться, что нет никаких противоречий» (<http://www.swsys.ru/uploaded/image/2024-3/19.jpg>).

Узлы синтаксических деревьев

Чтобы правильно сравнивать узлы синтаксического дерева между собой, требуется корректно определять способы их сравнения. Для достижения этой цели выделим три различных типа узлов, которые получаются в результате построения синтаксического дерева.

1. «Имена». После соответствующей предобработки текста из нескольких элементов дерева в один узел собирается все, что относится к имени, и сохраняется в едином универсальном формате. При сравнении узлов, которым присвоен тип имени, производится поиск частичных совпадений. Например, в случае отсутствия отчества в одном из узлов и его присутствия в другом, а также при условии совпадения остальных частей имени узлы в данной паре признаются совпадающими (узлы «А.С. Пушкин», «Александр Пушкин» и «Пушкин» совпадают).

2. «Даты». Аналогично с обработкой имен все составные части, относящиеся к одной дате, собираются в единый узел с соответствующим присвоением типа узла. Процедура сравнения такого типа узлов определяет даты совпадающими в случае их частичного совпадения («12 декабря 2023», «12.12.2023» и «12 декабря» совпадают).

3. «Текст». В этот тип узлов включаются все остальные вершины графа. Сравнение пары узлов такого типа будет описано детально позднее в силу того, что эта процедура является куда более нетривиальной, чем для остальных типов узлов.

Для наглядности рассмотрим обработку пары следующих предложений: 1) «С сентября 1830 года Михаил Юрьевич Лермонтов учился в Московском университете»; 2) «С 01.09.1830 М.Ю. Лермонтов учился в Московском университете». В данных предложениях даты и имена написаны в разных форматах, однако они в любом случае выделяются в единый узел. Кроме того, при дальнейшей обработке графов

и их последующем сравнении даты и имена в этом примере будут распознаны совпадающими (<http://www.swsys.ru/uploaded/image/2024-3/20.jpg>).

Виды отношений на множестве узлов синтаксических деревьев

Введем отношения на парах узлов.

Назовем узлы совпадающими, если они в точности совпадают после их последовательных преобразований к стандартной форме или являются синонимами и одновременно имеют или не имеют метку отрицательности. Также узлы будут совпадающими, если являются антонимами друг к другу и один из узлов имеет метку отрицательности. Примером таких преобразований для текстовых узлов могут послужить процедуры лемматизации (приведение слов к нормальной форме) и стемминга (нахождение основы слова).

Назовем узлы противоположными, если они в точности совпадают после их последовательных преобразований к стандартной форме или являются синонимами и один из узлов имеет метку отрицательности. Также узлы будут противоположными, если они являются антонимами друг к другу и оба узла одновременно имеют или не имеют метку отрицательности. Для определения того, что слова – синонимы или антонимы, авторы используют открытый репозиторий, в котором синонимы и антонимы представлены в виде графовой структуры (https://github.com/ahmados/rusynonyms/blob/main/ru_synonyms/synonyms.py).

Назовем узлы наложившимися или наложенными друг на друга, если они являются совпадающими или противоположными.

Для узлов, по типу являющихся датами или именами, алгоритмом сравнения будет поиск частичных совпадений. Наибольший интерес представляет алгоритм сравнения текстовых узлов. Следует отметить, что в русском языке местоимения могут употребляться вместо имен существительных, прилагательных или числительных. Поэтому важно при сравнении текстового узла, в котором содержится местоимение, с другим узлом определить его соответствующее отношение к узлу, необязательно являющемуся местоимением. Для таких слов нужно установить совпадение по гендерной принадлежности, определение которой возможно с помощью сравнения текста до применения к нему процедуры стемминга.

Алгоритм сравнения синтаксических деревьев, построенных на основе графов синтаксического разбора

Так как процесс сравнения синтаксических деревьев является вычислительно сложной процедурой, перед его началом необходимо убедиться, что вообще имеет смысл сравнивать графы между собой столь ресурсозатратным алгоритмом. Для этого предварительно определяется степень возможной изоморфности деревьев, построенных на основе эталонного и студенческого ответов, а затем сравниваются все узлы одного графа со всеми узлами другого. В случае получения в качестве результата предварительной проверки процентного соотношения наложившихся узлов меньшего, чем заданная нижняя граница автоматического отклонения, в качестве ответа системы возвращаются заведомо нулевые метрики, означающие несовпадение синтаксических деревьев студенческого и эталонного ответов.

Далее степень изоморфности графов определяется с помощью алгоритма для нахождения оптимального пути между двумя графами с использованием динамического программирования, представленного в работе [19].

1. Инициализируется матрица динамического программирования D значениями $D(i, 0) = i$ и $D(0, j) = j$, где i и j – индексы, представляющие i -й и j -й узлы в двух графах. Объясняется это тем, что первая строка и столбец представляют случай, когда один из графов пуст, поэтому расстояние редактирования между любым узлом непустого графа и пустым графом равно количеству узлов в непустом графе.

2. Вычисляется оптимальный путь редактирования. Для каждой ячейки (i, j) в таблице D вычисляется минимальное расстояние редактирования одного подграфа в другой с помощью возможных элементарных операций над узлами (вставки, удаления, замены). Стоимость каждой операции добавляется к стоимости предыдущего преобразования подграфа, и минимальное значение сохраняется в $D(i, j)$. Оптимальный путь между двумя графами получается обратным проходом через таблицу динамического программирования, начиная с нижней правой ячейки $D(n, m)$, где n и m – количество узлов в двух сравниваемых графах.

3. Применяются оптимизации. Алгоритм включает в себя несколько оптимизаций для повышения его эффективности, таких как сокращение пространства поиска и использование эвристик для управления поиском.

Таким образом, с помощью элементарных преобразований можно искать оптимальные наложения одного графа на другой, содержащие в себе наибольшее количество наложенных узлов. На вход данному алгоритму подается функция, которая определяет узлы, считающиеся наложенными друг на друга. Алгоритм на выходе возвращает пары узлов, оказавшихся наложенными по методу кратчайшего расстояния редактирования.

После получения наилучшего варианта наложения вершин друг на друга проводится анализ отношений на полученных парах узлов. Выявляются сначала совпадающие узлы, а затем распознающиеся как противоположные. В конце обрабатываются вершины, не относящиеся ни к тем, ни к другим, но наложившиеся друг на друга в ходе работы алгоритма, а также не получившие пары вовсе. Для сравнения по смыслу таких аберрантных пар в дальнейшем применяется кросс-проверка с целью обнаружения совпадающих узлов.

Обработка множественных отрицаний

Найденные противоположные вершины подвергаются дополнительной постобработке на предмет выявления и удаления сдвоенных противоположностей. Критерий для их распознавания в рекурсивной обработке – наличие у пары противоположных вершин наследников, которые тоже являются парой противоположных вершин. Анализируется глубина подобной рекурсии, и в дальнейшем четным образом происходит удаление распознанных данных способом пар узлов – остается либо одна самая первая пара противоположных узлов, либо ни одной (<http://www.swsys.ru/uploaded/image/2024-3/13.jpg>). Так, в предложениях «Я не мог не согласиться с этим утверждением» и «Я мог согласиться с этим утверждением» пары противоположностей и их противоположных наследников после обработки выглядят так: «мог (отрицательно) – мог» и «согласиться (отрицательно) – согласиться». Нахождение таких пар отражается при подсчете вероятности в виде совпадений.

Подсчет метрик

При вычислении метрик совпадения деревьев студенческого и эталонного ответов необходимо учитывать различия в найденных парах узлов: совпадающие, противоположные и совпадающие после кросс-проверки вершины.

В работе выделены три информативных показателя.

1. Доля наложения. Она измеряет то, насколько предложения являются подходящими одно другому. В нем учитываются количество наложенных вершин, вычисленное после применения апробационного алгоритма для выявления изоморфности графов, и количество наложений, полученных кросс-проверкой. Все это сравнивается с графом эталонного ответа, в котором из-за простоты предполагается меньшее количество узлов.

2. Доля совпадения. Отличие от доли наложения в том, что вместо количества наложенных вершин рассматривается количество именно совпадающих вершин. Для корректного учета число удаленных двойных противоположностей должно добавляться к количеству совпадающих вершин.

3. Число противоположных вершин. Такие противоположные с эталонным ответом вершины свидетельствуют о наличии лексической ошибки в ответе ученика, поэтому в метрику они поступают в виде штрафа.

Демонстрация работы системы на показательных примерах

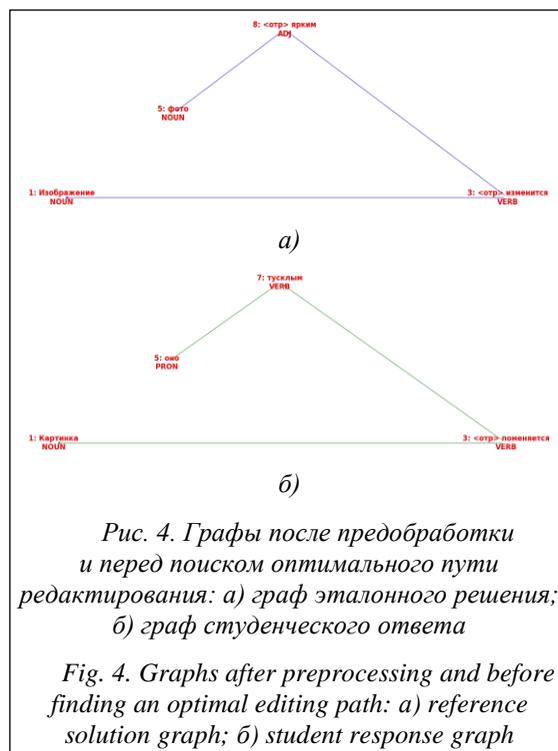
Рассмотрим значение анализа синонимов и местоимений на примере сравнения предложений «Изображение не изменится, фото будет менее ярким» и «Картинка не поменяется, оно будет тусклым». На рисунке 4 можно отследить взаимосвязи между дальнейшими наложениями вершин друг на друга. Слова «изображение» и «картинка» являются синонимами, поэтому они отмечаются в виде совпадающих. Аналогично в роли синонимов выступают слова «изменится» и «поменяется», поэтому их отрицания совпадают. Обработка местоимений позволяет соотнести между собой «фото» и «оно» в качестве совпадающих узлов. Наконец, «ярким (отриц.)» и «тусклым» совпадают, так как «ярким» и «тусклым» являются антонимами, а одному из слов приписана метка отрицательности. В конечном счете метрики в виде доли наложения и доли совпадения принимают единичное значение, что свидетельствует о полном совпадении графов студенческого и эталонного решений.

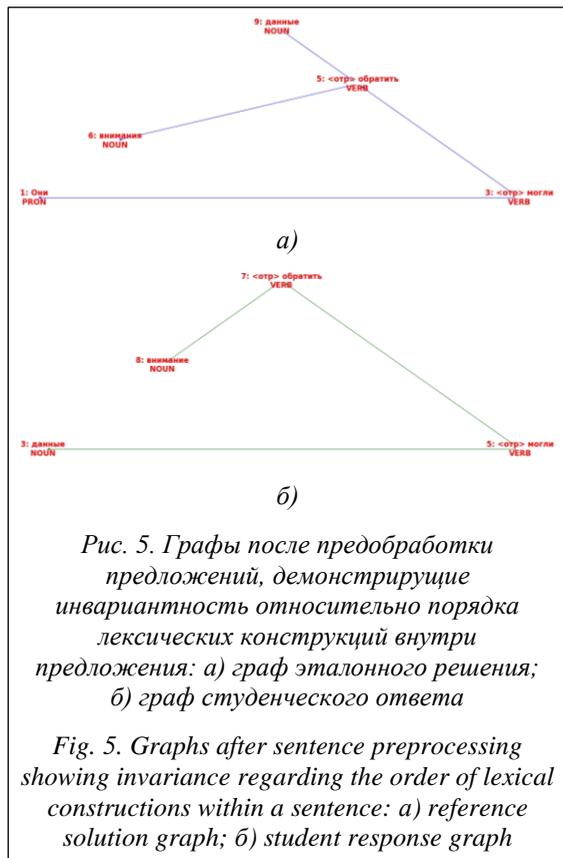
Точно таким же образом обрабатываются обратные случаи, в которых накладываются узлы, являющиеся комбинациями отрицаний, синонимов и антонимов и получающие статус противоположностей.

Продемонстрируем независимость результата от порядка расположения лексически вер-

ных логико-грамматических конструкций внутри предложения. Для этого рассмотрим следующую пару строк: «Они не могли не обратить внимания на эти данные» и «На эти данные не могли не обратить внимание». Графы для данных предложений после предобработки изображены на рисунке 5. Полученные в результате работы системы пары совпадающих вершин и метрики для доли наложения и доли совпадения свидетельствуют о независимости результата ее работы от порядка конструкций внутри предложения. Акцент данного алгоритма сделан на том, чтобы уловить смысловую составляющую текстов. Отметим важность в данном примере кросс-проверки, так как именно с ее помощью пара совпадающих вершин «данные» и «данные» находят друг друга, хотя поиск оптимального пути не поставил их в соответствие из-за отсутствия действующего лица «они» в тексте студенческого ответа.

Аналогично можно рассмотреть пару сложных предложений, в которых основы представлены в разном порядке. Например, «Мы были восхищены, а они едва подавали признаки жизни» и «Они едва подавали признаки жизни, а мы были восхищены» (<http://www.swsys.ru/uploaded/image/2024-3/14.jpg>). Полученные пары совпадающих вершин и единичные метрики свидетельствуют о независимости результата работы системы от порядка расположения основ предложения и связанных с ними оборотов.





Научная новизна исследования

В данной работе выделяется разработка системы автоматизированной проверки ответов на основе метода извлечения информации, что позволяет оценивать студенческие ответы со смысловой стороны текста, игнорируя многие грамматические неточности. Это обеспечивает прозрачность, справедливость и исключает предвзятость в оценке результатов, а также позволяет уменьшить нагрузку на преподавателя. Система автоматизированной проверки текстов на основе извлечения информации является более интерпретируемой для пользователей, чем на основе нейросетевых алгоритмов, так как она руководствуется правилами и алгоритмами, специфичными для языка, что делает ее более надежной в обработке студенческих ответов, поскольку проверяется реальный смысл ответа, а не его общая тематика. Отмечается возможность гибкой настройки системы с помощью задаваемых пользователем параметров обработки. Например, настраиваются верхние и нижние границы метрик, по которым проводится классификация итогового результата и которые могут быть специфичными для различных заданий, а также имеется возможность изменять внутреннюю систему штрафов и поощрений для каждой метрики.

Тестирование системы

Для проверки работоспособности системы были проведены контрольные мероприятия для студентов. Их ответы собирались и использовались в качестве тестовой выборки. Всего в выборке участвовали 456 студенческих ответов. После получения ответа система автоматизированной проверки принимала его на вход и, руководствуясь заданными преподавателем порогами отсечения по метрике доли наложения, относила к одному из трех классов: правильный ответ, неправильный ответ или неопределенный ответ, нуждающийся в дополнительной ручной проверке преподавателем по причине неуверенности системы после попадания результирующей метрики работы алгоритма внутрь интервала, ограниченного порогами отсечения, заданными преподавателем (выше границы признания ответа неверным и ниже границы признания его верным). Далее все те же студенческие ответы отдавались преподавателю, который независимо от решения системы автоматизированной проверки размечал их на правильные и неправильные. В данном тестировании пороги отсечения по метрике настраивались вручную и отличались от задания к заданию на основании аналитической оценки преподавателем специфики конкретного задания. В дальнейших исследованиях планируется разработка ассистента автоматической подстройки пороговых значений под статистические особенности каждого отдельно взятого задания. После получения результатов проверок составлялась матрица несоответствий и высчитывались следующие ключевые метрики для полученных результатов: количество истинно положительных, ложноотрицательных, ложноположительных, истинно отрицательных, неопределенных положительных и неопределенных отрицательных результатов (табл. 1).

Отмечается, что пороги отсечения настраиваются преподавателем вручную и вопрос их автоматического подбора по текущим статистическим характеристикам является открытым для дальнейших исследований. Важно настраивать их таким образом, чтобы сохранить отсутствие ошибок первого и второго рода, но при этом повысить метрику доли автоматической проверки, что будет выражаться в уменьшении доли ответов, отправленных на дополнительную проверку преподавателем в связи с неуверенностью системы. Стоит упомянуть о невозможности эмпирического под-

бора стандартных порогов, единых для всех заданий, так как задания в общем случае неоднородные и сами метрики являются параметрически варьируемыми. Это означает наличие необходимого требования именно динамической подстройки порогов под статистически значимую выборку для каждого задания.

Таблица 1

Матрица несоответствий для проведенного эксперимента оценки студенческих ответов

Table 1

Inconsistency matrix for the conducted experiment evaluating student responses

Матрица несоответствий		Оценка тестирующей системы		
		Ответ верен	Ответ неверен	Не определено
Оценка преподавателя	Ответ верен	311	0	57
	Ответ неверен	0	11	77

Немаловажной метрикой является доля распознанных ответов с помощью системы автоматической проверки, которая составила 70,61 %. Заметно, что в данном эксперименте система проявляла весьма щадящее по отношению к студенческим ответам поведение, о чем свидетельствует низкое соотношение истинно отрицательных к истинно положительным результатам, равное 0,035. Это достигается за счет ручного выбора достаточно малого нижнего порога отсечения.

Сравнение результатов с современными аналогами

Как уже отмечалось, система из [3] работает с русскоязычными текстами. Приведем для сравнения метрики, полученные в результате ее работы на аналогичном эксперименте со студенческими ответами (табл. 2). Результирующие метрики: точность – 96,64 %, полнота – 83,65 %, F-мера с единичным параметром $\beta = 0,8968$.

В силу отсутствия в работе [3] данных по порогам отсечения в целях сравнения можно считать, что пороги настраиваются по наилучшим показателям метрик. В таком случае в качестве преимущества системы, рассмотренной в настоящей работе, можно выделить отсутствие ошибок первого и второго рода. Это достигается благодаря тому, что при возникновении сомнений система отправляет преподава-

телю студенческий ответ на дополнительную проверку. Именно снижение нагрузки на преподавателей при возможности минимизации ошибочных классификаций является значимым преимуществом системы.

Таблица 2

Матрица несоответствий для проведенного эксперимента оценки студенческих ответов

Table 2

Inconsistency matrix for the experiment evaluating student responses conducted

Матрица несоответствий		Оценка тестирующей системы	
		Ответ верен	Ответ неверен
Оценка преподавателя	Ответ верен	747	146
	Ответ неверен	26	526

Заключение

Авторами была разработана система автоматизированной проверки ответов на открытые вопросы на основе метода извлечения информации. Система имеет возможность настройки преподавателем порогов отсечения для оценки. Немаловажную роль играет их корректная подстройка под статистические результаты с учетом минимизации ошибок распознающей системы и адаптации под различные метрики, настраиваемые преподавателем, что является одним из преимущественных направлений для дальнейших исследований.

В подавляющем количестве случаев результат работы системы совпадает с ожиданием преподавателей. Однако подобное соответствие ожиданиям имеет высокую корреляцию с качеством эталонного решения, которое посылается со стороны преподавателя, а также с четкостью формулировки задания. В связи с этим актуальным направлением продолжения работы является автоматизированная помощь преподавателю в составлении правильного эталонного решения.

Акцент в оценивании ответов смещен на анализ смысловой составляющей всего текста, что позволяет игнорировать некоторые грамматические ошибки: словообразовательные, морфологические или синтаксические. Однако нарушение лексических норм отслеживается данной системой, и в случае обнаружения подобного нарушения в результирующую метрику вносится штраф.

Список литературы

1. Леонов А.Г., Дьяченко М.С., Машенко К.А., Орловский А.Е., Райко И.Г., Райко М.В. Новые подходы к автоматизации проверки заданий в цифровых курсах // Информатизация образования и методика электронного обучения: цифровые технологии в образовании: матер. VI Междунар. науч. конф. 2022. Ч. 3. С. 173–178.
2. Mitchell T., Russell T., Broomhead P., Aldridge N. Towards robust computerised marking of free-text responses. Proc. 6th CAA Conf., 2002, pp. 233–249.
3. Кожевников В.А., Сабинин О.Ю. Система автоматической проверки ответов на открытые вопросы на русском языке // Информатика, телекоммуникации и управление. 2018. Т. 11. № 3. С. 57–72.
4. Gao Ju., Gao Ji. A similarity measurement method based on graph kernel for disconnected graphs. Proc. IJCAI, 2019, pp. 6430–6431.
5. Osman A.H., Barukub O.M. Graph-based text representation and matching: A review of the state of the art and future challenges. IEEE Access, 2020, vol. 8, pp. 87562–87583. doi: 10.1109/ACCESS.2020.2993191.
6. Callear D., Jerrams-Smith J., Soh V. CAA of short non-MCQ answers. Proc. 5th Computer Assisted Assessment Conf., 2001, pp. 1–14.
7. Goma W.H., Fahmy A.A. A survey of text similarity approaches. IJCA, 2013, vol. 68, no. 13, pp. 13–18. doi: 10.5120/11638-7118.
8. Othman N., Faiz R. Question answering passage retrieval and re-ranking using N-grams and SVM. Computación y Sistemas, 2016, vol. 20, no. 3, pp. 483–494. doi: 10.13053/CyS-20-3-2470.
9. Bashir F., Arshad H., Javed A. et al. Subjective answers evaluation using machine learning and natural language processing. IEEE Access, 2021, vol. 9, pp. 158972–158983. doi: 10.1109/ACCESS.2021.3130902.
10. Нурутдинов А.Р., Латыпов Р.Х. Перспективы биоинспирированного подхода в разработке систем искусственного интеллекта (обзор тенденций) // Ученые записки Казанского университета. Сер. Физико-математические науки. 2022. Т. 164. № 2-3. С. 244–265. doi: 10.26907/2541-7746.2022.2-3.244-265.
11. Marcheggiani D., Titov I. Encoding sentences with graph convolutional networks for semantic role labeling. Proc. EMNLP, 2017, pp. 1506–1515. doi: 10.18653/v1/D17-1159.
12. Sawant U., Garg S., Chakrabarti S., Ramakrishnan G. Neural architecture for question answering using a knowledge graph and Web corpus. Inf. Retrieval J., 2019, vol. 22, no. 3, pp. 324–349. doi: 10.1007/s10791-018-9348-8.
13. Bahel V., Thomas A. Text similarity analysis for evaluation of descriptive answers. 2021.arXiv preprint arXiv:2105.02935.
14. Othman N., Faiz R., Smaïli K. Manhattan Siamese LSTM for question retrieval in community question answering. In: LNPSE. Proc. OTM, 2019, vol. 11877, pp. 661–677. doi: 10.1007/978-3-030-33246-4_41.
15. Yang J., Li Y., Ga C., Zhang Y. Measuring the short text similarity based on semantic and syntactic information. Future Generation Computer Systems, 2021, vol. 114, pp. 169–180. doi: 10.1016/j.future.2020.07.043.
16. Alqaryouti O., Khwileh H., Farouk T., Nabhan A., Shaalan K. Graph-based keyword extraction. In: SCI. Intelligent Natural Language Processing: Trends and Applications, 2018, vol. 740, pp. 159–172. doi: 10.1007/978-3-319-67056-0_9.
17. Бахтеев О.Ю., Гафаров Ф.М., Гриншкун В.В., Дятлова О.В. и др. Цифровая платформа образования // Вестн. РФФИ. 2022. № 1. С. 87–103. doi: 10.22204/2410-4639-2022-113-01-87-103.
18. Васильев И.А., Кушниренко А.Г., Леонов А.Г., Машенко К.А., Холькина А.А., Шляхов А.В. Цифровая образовательная платформа Мирера – основа цифровой трансформации образовательного процесса // Новые образовательные стратегии в современном информационном пространстве: сб. матер. 2023. С. 140–144.
19. Abu-Aisheh Z., Raveaux R., Ramel J.-Y., Martineau P. An exact graph edit distance algorithm for solving pattern recognition problems. Proc. ICPRAM, 2015, vol. 1, pp. 271–278. doi: 10.5220/0005209202710278.

**Automation of semantic analysis for textual responses of students
in a digital educational platform**

**Aleksandr G. Leonov^{1,2,3,4}, Nikolay S. Martynov², Kirill A. Mashchenko^{1,2,4}✉,
Anna A. Kholkina⁵, Artem V. Shlyakhov¹**

¹M.V. Lomonosov Moscow State University, Moscow, 119991, Russian Federation

²Federal State Institution SRISA RAS, Moscow, 117218, Russian Federation

³Moscow State Pedagogical University, Moscow, 119435, Russian Federation

⁴State University of Management, Moscow, 109542, Russian Federation

⁵National University of Science Research Technological "MISIS",
Moscow, 119049, Russian Federation

For citation

Leonov, A.G., Martynov, N.S., Mashchenko, K.A., Kholkina, A.A., Shlyakhov, A.V. (2024) 'Automation of semantic analysis for textual responses of students in a digital educational platform', *Software & Systems*, 37(3), pp. 440–452 (in Russ.). doi: 10.15827/0236-235X.142.440-452

Article info

Received: 09.04.2024

After revision: 04.05.2024

Accepted: 14.05.2024

Abstract. The purpose of this study is to develop and apply a system of automated checking of students' answers to open-ended text questions within the framework of a digital educational platform. The developed system enables significant reducing of teacher workload and minimizing both type I and type II errors during evaluating students' answers. The research methodology involves modern methods of information extraction, text analysis and concept comparison. Constructing a sentence-parsing graph forms syntactic trees. After that, taking into account the peculiarities of the Russian language, the relations between their nodes are analyzed. For example, there might be double negation elimination, named entity extraction and lexical relationship recognition. To compare trees with each other, the authors use the search for the optimal edit path implemented using dynamic programming. The result of the research is a system that is able to evaluate students' answers taking into account the semantic depth of texts' content, which can ensure objectivity and transparency of knowledge evaluation. In addition, it reduces teacher workload by automating the process of checking answers. The possibility of parametric setting of the system operation scenario helps to adjust it to specific requirements to the answers set by a teacher. The system is able to ignore grammatical errors, as it focuses on tracking differences or similarities in the semantic component of texts. The practical significance of the work is in the possibility of effective and fair control of students' knowledge taking into account a text semantic component, in reducing teachers' routine and accelerating the process of evaluating students' answers. Also, it adapts to teachers' specific requirements and provides prompt feedback. This study is relevant and significant for educational practice, as it corresponds to modern educational trends and technologies.

Keywords: Mirera digital learning platform, meaning analysis, semantic analysis, response evaluation system, educational technology, automated verification

Acknowledgements. The work was carried out within the framework of the state assignment of the Federal State Institution SRISA RAS on topic no. FNEF-2024-0001 (1023032100070-3-1.2.1)

References

1. Leonov, A.G., Dyachenko, M.S., Mashchenko, K.A., Orlovsky, A.E., Raiko, I.G., Raiko, M.V. (2022) 'New approaches to automation of testing tasks in digital courses', *Proc. Int. Sci. Conf. Informatization of Education and E-learning Methods: Digital Technologies in Education*, 3, pp. 173–178 (in Russ.).
2. Mitchell, T., Russell, T., Broomhead, P., Aldridge, N. (2002) 'Towards robust computerised marking of free-text responses', *Proc. 6th CAA Conf.*, pp. 233–249.
3. Kozhevnikov, V.A., Sabinin, O.Yu. (2018) 'System of automatic verification of answers to open questions in Russian', *Computing, Telecommunications and Control*, 11(3), pp. 57–72 (in Russ.).
4. Gao, Ju., Gao, Ji. (2019) 'A similarity measurement method based on graph kernel for disconnected graphs', *Proc. IJCAI*, pp. 6430–6431.
5. Osman, A.H., Barukub, O.M. (2020) 'Graph-based text representation and matching: A review of the state of the art and future challenges', *IEEE Access*, 8, pp. 87562–87583. doi: 10.1109/ACCESS.2020.2993191.
6. Callear, D., Jerrams-Smith, J., Soh, V. (2001) 'CAA of short non-MCQ answers', *Proc. 5th Computer Assisted Assessment Conf.*, pp. 1–14.
7. Gomaa, W.H., Fahmy, A.A. (2013) 'A survey of text similarity approaches', *IJCA*, 68(13), pp. 13–18. doi: 10.5120/11638-7118.
8. Othman, N., Faiz, R. (2016) 'Question answering passage retrieval and re-ranking using N-grams and SVM', *Computación y Sistemas*, 20(3), pp. 483–494. doi: 10.13053/CyS-20-3-2470.
9. Bashir, F., Arshad, H., Javed, A. et al. (2021) 'Subjective answers evaluation using machine learning and natural language processing', *IEEE Access*, vol. 9, pp. 158972–158983. doi: 10.1109/ACCESS.2021.3130902.
10. Nurutdinov, A.R., Latypov, R.Kh. (2022) 'Potentials of the Bio-inspired approach in the development of artificial intelligence systems (trends review)', *Proc. of Kazan University. Phys. and Math. Ser.*, 164(2-3), pp. 244–265 (in Russ.). doi: 10.26907/2541-7746.2022.2-3.244-265.
11. Marcheggiani, D., Titov, I. (2017) 'Encoding sentences with graph convolutional networks for semantic role labeling', *Proc. EMNLP*, pp. 1506–1515. doi: 10.18653/v1/D17-1159.
12. Sawant, U., Garg, S., Chakrabarti, S., Ramakrishnan, G. (2019) 'Neural architecture for question answering using a knowledge graph and Web corpus', *Inf. Retrieval J.*, 22(3), pp. 324–349. doi: 10.1007/s10791-018-9348-8.
13. Bahel, V. and Thomas, A. (2021) 'Text similarity analysis for evaluation of descriptive answers', *arXiv preprint arXiv:2105.02935*.
14. Othman, N., Faiz, R., Smaïli, K. (2019) 'Manhattan Siamese LSTM for question retrieval in community question answering', in *LNPSE. Proc. OTM*, 11877, pp. 661–677. doi: 10.1007/978-3-030-33246-4_41.
15. Yang, J., Li, Y., Ga, C., Zhang, Y. (2021) 'Measuring the short text similarity based on semantic and syntactic information', *Future Generation Computer Systems*, 114, pp. 169–180. doi: 10.1016/j.future.2020.07.043.
16. Alqaryouti, O., Khwileh, H., Farouk, T., Nabhan, A., Shaaan, K. (2018) 'Graph-based keyword extraction', in *SCI. Intelligent Natural Language Processing: Trends and Applications*, 740, pp. 159–172. doi: 10.1007/978-3-319-67056-0_9.
17. Bakhteev, O.Yu., Gafarov, F.M., Grinshkun, V.V., Dyatlova, O.V. et al. (2022) 'Digital Education Platform', *Vestn. RFFI*, (1), pp. 87–103 (in Russ.). doi: 10.22204/2410-4639-2022-113-01-87-103.

18. Vasilev, I.A., Kushniренко, A.G., Leonov, A.G., Mashchenko, K.A., Kholkina, A.A., Shlyakhov, A.V. (2023) 'Mirera digital educational platform – the basis of digital transformation of the educational process', *Proc. NESinODS*, pp. 140–144 (in Russ.).

19. Abu-Aisheh, Z., Raveaux, R., Ramel, J.-Y., Martineau, P. (2015) 'An exact graph edit distance algorithm for solving pattern recognition problems', *Proc. ICPRAM*, 1, pp. 271–278. doi: 10.5220/0005209202710278.

Авторы

Леонов Александр Георгиевич^{1, 2, 3, 4}

к.ф.-м.н., доцент, ведущий научный сотрудник,
зав. кафедрой, профессор,
dr.l@math.msu.su

Николай Сергеевич Мартынов², инженер,

nikolai.martynov@math.msu.ru

Мащенко Кирилл Алексеевич^{1, 2, 4},

младший научный сотрудник,

kirill010399@vip.niisi.ru

Анна Александровна Холькина⁵,

студент, kholkina.a2021@gmail.com

Артем Вячеславович Шляхов²,

младший научный сотрудник,

shlyakhov@vip.niisi.ru

Authors

Aleksandr G. Leonov^{1, 2, 3, 4},

Cand. of Sci. (Physics and Mathematics),
Associate Professor, Professor, Head of Chair,
Leading Researcher, dr.l@math.msu.su

Nikolay S. Martynov², Engineer,

nikolai.martynov@math.msu.ru

Kirill A. Mashchenko^{1, 2, 4},

Junior Researcher,

kirill010399@vip.niisi.ru

Anna A. Kholkina⁵,

Student, kholkina.a2021@gmail.com

Artem V. Shlyakhov¹,

Junior Researcher,

shlyakhov@vip.niisi.ru

¹ МГУ им. М.В. Ломоносова,

г. Москва, 119991, Россия

² ФГУ ФНЦ НИИСИ РАН,

г. Москва, 117218, Россия

³ Институт детства Московского педагогического

государственного университета (МПГУ),

г. Москва, 119435, Россия

⁴ Институт информационных систем

Государственного университета управления,

г. Москва, 109542, Россия

⁵ Национальный исследовательский

технологический университет «МИСиС»,

г. Москва, 119049, Россия

¹ M.V. Lomonosov Moscow State University,

Moscow, 119991, Russian Federation

² Federal State Institution SRISA RAS,

Moscow, 117218, Russian Federation

³ Moscow State Pedagogical University,

Moscow, 119435,

Russian Federation

⁴ State University of Management,

Moscow, 109542,

Russian Federation

⁵ National University of Science

Research Technological "MISIS",

Moscow, 119049, Russian Federation

СОДЕРЖАНИЕ

Катулин М.С., Решетников А.Г., Рябов А.Р., Ульянов С.В. Оптимизация процессов интеллектуального управления в реальном времени на физической модели робота-манипулятора с помощью генетического алгоритма.....	301
Шайманов Н.Ю., Авраамов В.П., Иванов А.А., Куксенко С.П. Использование метода валидации выделением особенностей для сопоставления наборов экспериментальных или смоделированных данных	310
Коробейников А.Г. Применение реализованных в MATLAB методов Big Data в задаче выделения геомагнитных пульсаций типа Pc3 из данных INTERMAGNET	318
Дородных Н.О., Юрин А.Ю. Программная реализация алгоритмов для создания прототипов баз знаний на основе визуального моделирования и трансформаций.....	324
Коростелев Д.А., Подвесовский А.Г., Захарова А.А. Организация хранения данных функционирования объектов киберфизических систем	334
Максимов А.Е., Снетков П.П., Иванов А.А., Куксенко С.П. Применение MATLAB/GNU Octave для расчета погонных параметров многопроводных линий передачи методом моментов.....	344
Зяблова Е.Р. Модификация алгоритма поиска кратчайших путей GH-графа для анализа сложных технических систем.....	354
Татарникова Т.М., Боженко В.В. Основные принципы работы обобщенной регрессионной нейронной сети при заполнении пропущенных значений в наборах данных	364
Долгов Н.В., Ильин В.А. Функциональное моделирование двух сложных конкурирующих систем для обоснования выбора рациональной структуры исследуемой системы	369
Гуличева А.А., Рыбаков А.А. Реберная раскраска кубического графа в задаче распараллеливания расчетов на неструктурированной поверхностной расчетной сетке	374
Ай Мин Тайк, Лупин С.А., Телегин П.Н., Шабанов Б.М. Ускорение расчета критериальной функции в задаче размещения всенаправленных антенн	384
Татарникова Т.М., Булгаков Д.А. Визуальная оценка качества работы генератора псевдослучайных чисел для решения криптографических задач.....	393
Абу-Абед Ф.Н., Жиронкин С.А. Компоненты и функциональность программного средства моделирования структуры импортозамещения.....	402
Вяткин С.И., Долговесов Б.С. Компьютерное моделирование сложных взаимодействий отдельных элементов гранулированных материалов с применением функционально заданных объектов.....	412

Антонов М.О., Темкин И.О.

Распознавание и отслеживание дефектов дорожного полотна в реальном времени на основе комплексного использования стандартных вычислительных процедур и глубоких нейронных сетей..... 421

Грибова В.В., Лифшиц А.Я., Москаленко Ф.М., Шалфеева Е.А., Шевченко Н.Е.

Динамическая генерация пользовательского интерфейса мобильного медицинского приложения на основе онтологического подхода 431

Леонов А.Г., Мартынов Н.С., Мащенко К.А., Холькина А.А., Шляхов А.В.

Автоматизация проверки семантической составляющей текстовых ответов обучающихся в цифровой образовательной платформе 440

C O N T E N T

<i>Katulín, M.S., Reshetnikov, A.G., Ryabov, A.R., Ulyanov, S.V.</i> Physical model of a robot manipulator: Optimization of real-time intelligent control processes using a genetic algorithm.....	301
<i>Shaymanov, N.Yu., Avraamov, V.P., Ivanov, A.A., Kuksenko, S.P.</i> Applying the feature selective validation method to compare experimental or simulated datasets.....	310
<i>Korobeynikov, A.G.</i> Identifying Pc3 type geomagnetic pulsations from INTERMAGNET data: MATLAB Big Data.....	318
<i>Dorodnykh, N.O., Yurin A.Yu.</i> Software implementation of algorithms for knowledge base prototyping with visual modelling and transformations	324
<i>Korostelyov, D.A., Podvesovskii, A.G., Zakharova, A.A.</i> Cyber-physical system objects: Organizing operation data storage	334
<i>Maksimov, A.E., Snetkov, P.P., Ivanov, A.A., Kuksenko, S.P.</i> MATLAB/GNU Octave for calculating per-unit-length parameters of multiconductor transmission lines using the method of moments	344
<i>Zyablova, E.R.</i> Modifying a GH-graph shortest path search algorithm for analyzing complex technical systems	354
<i>Tatarnikova, T.M., Bozhenko, V.V.</i> Basic principles of generalized regression neural network in filling missing values in datasets.....	364
<i>Dolgov, N.V., Ilin, V.A.</i> Justifying the choice of a rational structure of the system under study: Functional modeling of two complex competing systems	369
<i>Gulicheva, A.A., Rybakov, A.A.</i> Cubic graph edge coloring in the problem of calculation parallelization on an unstructured surface computational grid.....	374
<i>Aye Min Thike, Lupin, S.A., Telegin, P.N., Shabanov, B.M.</i> Accelerating criterion function calculation in an omnidirectional antenna placement problem.....	384
<i>Tatarnikova, T.M., Bulgakov, D.A.</i> Visual quality assessment of the pseudorandom number generator for solving cryptographic tasks.....	393
<i>Abu-Abed, F.N., Zhironkin, S.A.</i> Components and functionality of a software tool for modelling an import substitution structure	402
<i>Vyatkin, S.I., Dolgovesov, B.S.</i> Computer modeling of complex interactions of distinct elements of granular materials using functionally specified objects	412
<i>Antonov, M.O., Temkin, I.O.</i> Real-time recognition and tracking of road surface defects based on complex using of standard computational procedures and deep neural networks	421

<i>Gribova, V.V., Lifshits A.Ya., Moskalenko, Ph.M., Shalfeeva, E.A., Shevchenko, N.E.</i>	
Dynamic generation of a user interface for a medical mobile application based on an ontological approach.....	431
<i>Leonov, A.G., Martynov, N.S., Mashchenko, K.A., Kholkina, A.A., Shlyakhov, A.V.</i>	
Automation of semantic analysis for textual responses of students in a digital educational platform	440

Внимание подписчиков

Научный журнал
«Программные продукты и системы»
распространяется по подписке
через подписное агентство «УП УРАЛ-ПРЕСС»
Индекс издания
в электронном каталоге агентства 70799

Телефон редакции: (4822) 39-91-49
Электронная почта: red@cps.tver.ru
Сайт журнала: www.swsys.ru