

www.aurora-group.eu

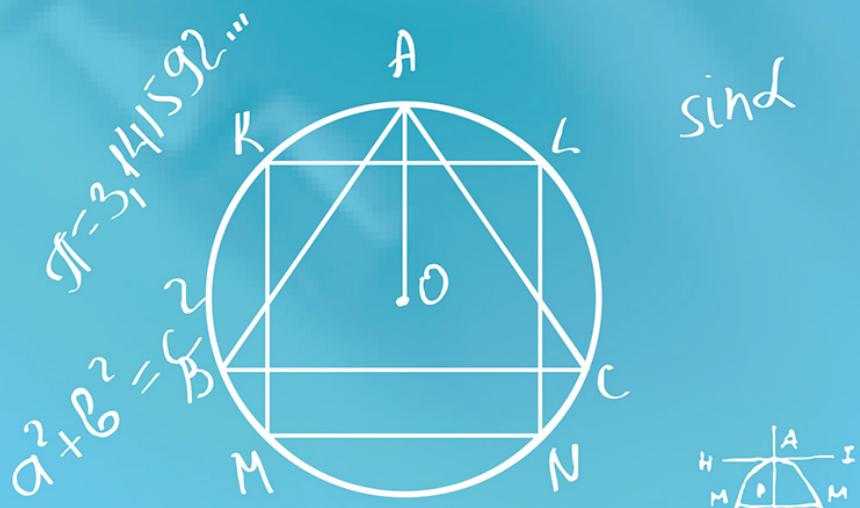
www.nbpublish.com

ПРОГРАММНЫЕ СИСТЕМЫ

и

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ

научный журнал



AURORA Group s.r.o.

nota bene

Выходные данные

Номер подписан в печать: 05-10-2024

Учредитель: Даниленко Василий Иванович, w.danilenko@nbpublish.com

Издатель: ООО <НБ-Медиа>

Главный редактор: Морозов Михаил Николаевич, кандидат технических наук,
mikhail.n.morozov@gmail.com

ISSN: 2454-0714

Контактная информация:

Выпускающий редактор - Зубкова Светлана Вадимовна

E-mail: info@nbpublish.com

тел.+7 (966) 020-34-36

Почтовый адрес редакции: 115114, г. Москва, Павелецкая набережная, дом 6А, офис 211.

Библиотека журнала по адресу: http://www.nbpublish.com/library_tariffs.php

Publisher's imprint

Number of signed prints: 05-10-2024

Founder: Danilenko Vasiliy Ivanovich, w.danilenko@nbpublish.com

Publisher: NB-Media ltd

Main editor: Morozov Mikhail Nikolaevich, kandidat tekhnicheskikh nauk,
mikhail.n.morozov@gmail.com

ISSN: 2454-0714

Contact:

Managing Editor - Zubkova Svetlana Vadimovna

E-mail: info@nbpublish.com

тел.+7 (966) 020-34-36

Address of the editorial board : 115114, Moscow, Paveletskaya nab., 6A, office 211 .

Library Journal at : http://en.nbpublish.com/library_tariffs.php

Редакционный совет

Гельман Виктор Яковлевич – доктор технических наук, профессор, профессор кафедры медицинской информатики и физики ФГБОУ ВО «Северо-Западный государственный медицинский университет им. И.И.Мечникова», 191015, Россия, г. Санкт-Петербург, ул. Кирочная, д.41, gelm@sg2104.spb.edu

Поляков Виктор Павлович – доктор педагогических наук, профессор, Главный научный сотрудник лаборатории психолого-педагогического и учебно-методического обеспечения развития информатизации образования Центра информатизации образования Федерального государственного бюджетного научного учреждения «Институт управления образованием Российской академии образования», 105062, г. Москва, ул. Макаренко, д. 5/16, стр. 1Б, polvikpal@mail.ru

Гармаев Баир Заятуевич – кандидат физико-математических наук, научный сотрудник, Институт физического материаловедения Сибирского Отделения РАН, 670000, Россия, республика Бурятия, г. Улан-Удэ, ул. Сахьяновой, 6, каб. 313

Клименко Анна Борисовна – кандидат технических наук, научный сотрудник Научно-исследовательского института многопроцессорных вычислительных систем имени академика А.В. Каляева Южного федерального университета (НИИ МВС ЮФУ), 347935, Россия, Ростовская область, г. Таганрог, ул. 8 Переулок, 15

Лютикова Лариса Адольфовна – кандидат физико-математических наук, заведующая отделом Нейроинформатики и машинного обучения, Институт прикладной математики и автоматизации Кабардино-Балкарского научного центра РАН – филиал Кабардино-Балкарского научного центра РАН (ИПМА КБНЦ РАН), 360000, Россия, Республика Кабардино-Балкария, г. Нальчик, ул. Шортанова, 89а

Мустафаев Арслан Гасанович – доктор технических наук, профессор, Государственное автономное образовательное учреждение высшего образования "Дагестанский государственный университет народного хозяйства", кафедра «Информационные технологии и информационная безопасность», 367015, Россия, Республика Дагестан, г. Махачкала, ул. Атаева, 5, каб. 4.5

Шестаков Александр Валентинович – кандидат технических наук, доцент Южный Федеральный университет, кафедра вычислительной техники, 347902, Россия, Ростовская область, г. Таганрог, ул. Свободы, 24/2

Сидоркина Ирина Геннадьевна - доктор технических наук, профессор, декан факультета Информатики и вычислительной техники Поволжского государственного технологического университета, Йошкар-Ола, Россия E-mail: dekan_fivt@mail.ru

Екатерина Прасолова-Forland - PhD, Норвежский университет науки и технологии (NTNU), Тронхейм, Норвегия E-mail: Ekaterina.Prasolova-Forland@idi.ntnu.no

Голенков Владимир Васильевич - доктор технических наук, профессор, заведующий кафедрой Интеллектуальных информационных технологий Белорусского государственного университета информатики и радиоэлектроники, г. Минск, Республика Беларусь E-mail: golen@bsuir.by

Домошницкий Александр Исаакович - кандидат физико-математических наук, декан естественно-научного факультета Университетского центра в г.Ариэль, Израиль, Самария E-mail: adom@ariel.ac.il Department of Mathematics and Computer Sciences, The Ariel

University Center of Samaria, 44837 Ariel, ISRAEL

Коробейников Анатолий Григорьевич - доктор технических наук, профессор «Институт земного магнетизма, ионосферы и распространения радиоволн РАН (ИЗМИРАН)», Санкт-Петербургский филиал E-mail: korobeynikov_a_g@mail.ru

Заболеева-Зотова Алла Викторовна, доктор технических наук, профессор Волгоградского технического университета, Волгоград, Россия E-mail: zabzot@gmail.com

Бенкевич Леонид Владимирович - кандидат физических наук и инженерной физики, научный сотрудник Массачусетского Технологического Института (MIT), обсерватория Хэйстек, Бостон, США E-mail: lbenkey@gmail.com

Морозов Михаил Николаевич - кандидат технических наук, профессор, руководитель лаборатории мультимедиа, заведующий кафедрой Информатики и системного программирования Поволжского государственного технологического университета, Йошкар-Ола, Россия E-mail: mikhail.n.morozov@gmail.com

Олзоева Сэсэг Ивановна - доктор технических наук, профессор, Восточно-Сибирский государственный университет технологий и управления (г. Улан-Удэ) E-mail: sseseg@yandex.ru

Курейчик Владимир Викторович - доктор технических наук, профессор, заведующий кафедрой Систем автоматизации проектирования Технологического института «Южного федерального университета» в г.Таганрог, Россия E-mail: vkur@tsure.ru

Филатова Наталья Николаевна - доктор технических наук, профессор, Тверской государственный технический университет, Тверь, Россия E-mail: nfilatova99@mail.ru

Песошин Валерий Андреевич - член-корреспондент Академии наук Республики Татарстан, заслуженный деятель науки Республики Татарстан и Российской Федерации, доктор технических наук, профессор, заслуженный деятель науки и техники Республики Татарстан. Заведующий кафедрой Компьютерных систем Казанского национальный исследовательский университет им. А.Н. Туполева, Казань, Россия E-mail: pesoshin@evm.kstu-kai.ru

Краснов Сергей Викторович - доктор технических наук, профессор, проректор по научно-исследовательской работе, заведующий кафедрой Информатика и системы управления Волжского университета им. Татищева, Тольятти, Россия E-mail: krasnovlt@mail.ru

Горохов Алексей Витальевич - доктор технических наук, профессор кафедры Прикладной математики и информационных технологий Поволжского государственного технологического университета, Йошкар-Ола, Россия E-mail: agv64@mail.ru

Галанина Наталья Андреевна - доктор технических наук, профессор, Чувашский государственный университет им. И.Н.Ульянова, Чебоксары, Россия E-mail: galaninacheb@mail.ru

Сюзев Владимир Васильевич - доктор технических наук, профессор, заведующий кафедрой Компьютерные системы и сети Московского государственного технического университета им. Н. Э. Баумана, Москва, Россия E-mail: v.suzev@bmstu.ru

Леухин Анатолий Николаевич - доктор физико-математических наук, профессор, заведующий кафедрой Информационной безопасности Поволжского государственного технологического университета, Йошкар-Ола, Россия E-mail: code@volgatech.net

Гвинианидзе Темур Николаевич - Доктор технических наук, профессор,
Государственный университет им. Ак. Церетели Грузия, г. Кутаиси, пр. Тамар-мепе 59.
П.и 4600. temuri1951@mail.ru

Council of Editors

Gelman Viktor Yakovlevich – Doctor of Technical Sciences, Professor, Professor of the Department of Medical Informatics and Physics of the I.I.Mechnikov Northwestern State Medical University, 41 Kirochnaya Str., St. Petersburg, 191015, Russia, gelm@sg2104.spb.edu

Polyakov Viktor Pavlovich – Doctor of Pedagogical Sciences, Professor, Chief Researcher of the Laboratory of Psychological, Pedagogical and Educational methodological support for the development of Informatization of Education of the Center for Informatization of Education of the Federal State Budgetary Scientific Institution "Institute of Education Management of the Russian Academy of Education", 105062, Moscow, Makarenko str., 5/16, p. 1B, polvikpal@mail.ru

Garmaev Bair Zayatuevich – Candidate of Physical and Mathematical Sciences, Researcher, Institute of Physical Materials Science of the Siberian Branch of the Russian Academy of Sciences, 670000, Russia, Republic of Buryatia, Ulan-Ude, Sakhyanova str., 6, room 313

Klimenko Anna Borisovna – Candidate of Technical Sciences, Researcher at the Research Institute of Multiprocessor Computing Systems named after Academician A.V. Kalyaev of the Southern Federal University (Research Institute of the Ministry of Internal Affairs of the Southern Federal University), 347935, Russia, Rostov region, Taganrog, ul. 8 Lane, 15

Lyutikova Larisa Adolfovna – Candidate of Physical and Mathematical Sciences, Head of the Department of Neuroinformatics and Machine Learning, Institute of Applied Mathematics and Automation of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences - branch of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences (IPMA KBSC RAS), 360000, Russia, Republic of Kabardino-Balkaria, Nalchik, 89a Shortanova str.

Mustafayev Arslan Hasanovich – Doctor of Technical Sciences, Professor, State Autonomous Educational Institution of Higher Education "Dagestan State University of National Economy", Department of "Information Technologies and Information Security", 367015, Russia, Republic of Dagestan, Makhachkala, Ataeva str., 5, office 4.5

Alexander V. Shestakov – Candidate of Technical Sciences, Associate Professor, Southern Federal University, Department of Computer Engineering, 24/2 Svobody str., Taganrog, Rostov Region, 347902, Russia

Sidorkina Irina Gennadievna - Doctor of Technical Sciences, Professor, Dean of the Faculty of Computer Science and Computer Engineering of the Volga State Technological University, Yoshkar-Ola, Russia E-mail: dekan_fivt@mail.ru

Ekaterina Prasolova-Forland - PhD, Norwegian University of Science and Technology (NTNU), Trondheim, Norway E-mail: Ekaterina.Prasolova-Forland@idi.ntnu.no

Golenkov Vladimir Vasilyevich - Doctor of Technical Sciences, Professor, Head of the Department of Intelligent Information Technologies of the Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus E-mail: golen@bsuir.by

Domoshnitsky Alexander Isaakovich - Candidate of Physical and Mathematical Sciences, Dean of the Faculty of Natural Sciences of the University Center in Ariel, Israel, Samaria E-mail: adom@ariel.ac.il Department of Mathematics and Computer Sciences, The Ariel University Center of Samaria, 44837 Ariel, ISRAEL

Korobeynikov Anatoly Grigorievich - Doctor of Technical Sciences, Professor, Institute of

Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences (IZMIRAN), St. Petersburg Branch E-mail: korobeynikov_a_g@mail.ru

Zaboleeva-Zotova Alla Viktorovna, Doctor of Technical Sciences, Professor of Volgograd Technical University, Volgograd, Russia E-mail: zabzot@gmail.com

Leonid V. Benkevich - Candidate of Physical Sciences and Engineering Physics, Researcher at the Massachusetts Institute of Technology (MIT), Haystack Observatory, Boston, USA E-mail: ibenkev@gmail.com

Mikhail N. Morozov - Candidate of Technical Sciences, Professor, Head of the Multimedia Laboratory, Head of the Department of Computer Science and System Programming of the Volga State Technological University, Yoshkar-Ola, Russia E-mail: mikhail.n.morozov@gmail.com

Olzoeva Seseg Ivanovna - Doctor of Technical Sciences, Professor, East Siberian State University of Technology and Management (Ulan-Ude) E-mail: sseseg@yandex.ru

Kureychik Vladimir Viktorovich - Doctor of Technical Sciences, Professor, Head of the Department of Design Automation Systems of the Technological Institute of the Southern Federal University in Taganrog, Russia E-mail: vkur@tsure.ru

Natalia Filatova - Doctor of Technical Sciences, Professor, Tver State Technical University, Tver, Russia E-mail: nfilatova99@mail.ru

Pesoshin Valery Andreevich - Corresponding member of the Academy of Sciences of the Republic of Tatarstan, Honored Scientist of the Republic of Tatarstan and the Russian Federation, Doctor of Technical Sciences, Professor, Honored Worker of Science and Technology of the Republic of Tatarstan. Head of the Department of Computer Systems of Kazan National Research University named after A.N. Tupolev, Kazan, Russia E-mail: pesoshin@evm.kstu-kai.ru

Krasnov Sergey Viktorovich - Doctor of Technical Sciences, Professor, Vice-Rector for Research, Head of the Department of Computer Science and Control Systems of the Volga State University. Tatishcheva, Togliatti, Russia E-mail: krasnovtlt@mail.ru

Gorokhov Alexey Vitalievich - Doctor of Technical Sciences, Professor of the Department of Applied Mathematics and Information Technologies of the Volga State Technological University, Yoshkar-Ola, Russia E-mail: agv64@mail.ru

Galanina Natalia Andreevna - Doctor of Technical Sciences, Professor, I.N.Ulyanov Chuvash State University, Cheboksary, Russia E-mail: galaninacheb@mail.ru

Vladimir V. Syuzev - Doctor of Technical Sciences, Professor, Head of the Department of Computer Systems and Networks of the Bauman Moscow State Technical University, Moscow, Russia E-mail: v.suzev@bmstu.ru

Leukhin Anatoly Nikolaevich - Doctor of Physical and Mathematical Sciences, Professor, Head of the Department of Information Security of the Volga State Technological University, Yoshkar-Ola, Russia E-mail: code@volgattech.net

Gvinianidze Temur Nikolaevich - Doctor of Technical Sciences, Professor, Ak. Tsereteli State University Georgia, Kutaisi, 59 Tamar-mepe Ave., and 4600. temuri1951@mail.ru

Требования к статьям

Журнал является научным. Направляемые в издательство статьи должны соответствовать тематике журнала (с его рубрикатором можно ознакомиться на сайте издательства), а также требованиям, предъявляемым к научным публикациям.

Рекомендуемый объем от 12000 знаков.

Структура статьи должна соответствовать жанру научно-исследовательской работы. В ее содержании должны обязательно присутствовать и иметь четкие смысловые разграничения такие разделы, как: предмет исследования, методы исследования, апелляция к оппонентам, выводы и научная новизна.

Не приветствуется, когда исследователь, трактуя в статье те или иные научные термины, вступает в заочную дискуссию с авторами учебников, учебных пособий или словарей, которые в узких рамках подобных изданий не могут широко излагать свое научное воззрение и заранее оказываются в проигрышном положении. Будет лучше, если для научной полемики Вы обратитесь к текстам монографий или докторских диссертаций работ оппонентов.

Не превращайте научную статью в публицистическую: не наполняйте ее цитатами из газет и популярных журналов, ссылками на высказывания по телевидению.

Ссылки на научные источники из Интернета допустимы и должны быть соответствующим образом оформлены.

Редакция отвергает материалы, напоминающие реферат. Автору нужно не только продемонстрировать хорошее знание обсуждаемого вопроса, работ ученых, исследовавших его прежде, но и привнести своей публикацией определенную научную новизну.

Не принимаются к публикации избранные части из докторских диссертаций, книг, монографий, поскольку стиль изложения подобных материалов не соответствует журнальному жанру, а также не принимаются материалы, публиковавшиеся ранее в других изданиях.

В случае отправки статьи одновременно в разные издания автор обязан известить об этом редакцию. Если он не сделал этого заблаговременно, рискует репутацией: в дальнейшем его материалы не будут приниматься к рассмотрению.

Уличенные в плагиате попадают в «черный список» издательства и не могут рассчитывать на публикацию. Информация о подобных фактах передается в другие издательства, в ВАК и по месту работы, учебы автора.

Статьи представляются в электронном виде только через сайт издательства <http://www.enotabene.ru> кнопка "Авторская зона".

Статьи без полной информации об авторе (соавторах) не принимаются к рассмотрению, поэтому автор при регистрации в авторской зоне должен ввести полную и корректную информацию о себе, а при добавлении статьи - о всех своих соавторах.

Не набирайте название статьи прописными (заглавными) буквами, например: «ИСТОРИЯ КУЛЬТУРЫ...» — неправильно, «История культуры...» — правильно.

При добавлении статьи необходимо прикрепить библиографию (минимум 10–15 источников, чем больше, тем лучше).

При добавлении списка использованной литературы, пожалуйста, придерживайтесь следующих стандартов:

- [ГОСТ 7.1-2003 Библиографическая запись. Библиографическое описание. Общие требования и правила составления.](#)
- [ГОСТ 7.0.5-2008 Библиографическая ссылка. Общие требования и правила составления](#)

В каждой ссылке должен быть указан только один диапазон страниц. В теле статьи ссылка на источник из списка литературы должна быть указана в квадратных скобках, например, [1]. Может быть указана ссылка на источник со страницей, например, [1, с. 57], на группу источников, например, [1, 3], [5-7]. Если идет ссылка на один и тот же источник, то в теле статьи нумерация ссылок должна выглядеть так: [1, с. 35]; [2]; [3]; [1, с. 75-78]; [4]....

А в библиографии они должны отображаться так:

[1]
[2]
[3]
[4]....

Постраничные ссылки и сноски запрещены. Если вы используете сноска, не содержащую ссылку на источник, например, разъяснение термина, включите сноска в текст статьи.

После процедуры регистрации необходимо прикрепить аннотацию на русском языке, которая должна состоять из трех разделов: Предмет исследования; Метод, методология исследования; Новизна исследования, выводы.

Прикрепить 10 ключевых слов.

Прикрепить саму статью.

Требования к оформлению текста:

- Кавычки даются углками (« ») и только кавычки в кавычках — лапками (“ ”).
- Тире между датамидается короткое (Ctrl и минус) и без отбивок.
- Тире во всех остальных случаяхдается длинное (Ctrl, Alt и минус).
- Даты в скобках даются без г.: (1932–1933).
- Даты в тексте даются так: 1920 г., 1920-е гг., 1540–1550-е гг.
- Недопустимо: 60-е гг., двадцатые годы двадцатого столетия, двадцатые годы XX столетия, 20-е годы ХХ столетия.
- Века, король такой-то и т.п. даются римскими цифрами: XIX в., Генрих IV.
- Инициалы и сокращения даются с пробелом: т. е., т. д., М. Н. Иванов. Неправильно: М.Н. Иванов, М.Н. Иванов.

ВСЕ СТАТЬИ ПУБЛИКУЮТСЯ В АВТОРСКОЙ РЕДАКЦИИ.

По вопросам публикации и финансовым вопросам обращайтесь к администратору Зубковой Светлане Вадимовне
E-mail: info@nbpublish.com
или по телефону +7 (966) 020-34-36

Подробные требования к написанию аннотаций:

Аннотация в периодическом издании является источником информации о содержании статьи и изложенных в ней результатах исследований.

Аннотация выполняет следующие функции: дает возможность установить основное

содержание документа, определить его релевантность и решить, следует ли обращаться к полному тексту документа; используется в информационных, в том числе автоматизированных, системах для поиска документов и информации.

Аннотация к статье должна быть:

- информативной (не содержать общих слов);
- оригинальной;
- содержательной (отражать основное содержание статьи и результаты исследований);
- структурированной (следовать логике описания результатов в статье);

Аннотация включает следующие аспекты содержания статьи:

- предмет, цель работы;
- метод или методологию проведения работы;
- результаты работы;
- область применения результатов; новизна;
- выводы.

Результаты работы описывают предельно точно и информативно. Приводятся основные теоретические и экспериментальные результаты, фактические данные, обнаруженные взаимосвязи и закономерности. При этом отдается предпочтение новым результатам и данным долгосрочного значения, важным открытиям, выводам, которые опровергают существующие теории, а также данным, которые, по мнению автора, имеют практическое значение.

Выводы могут сопровождаться рекомендациями, оценками, предложениями, гипотезами, описанными в статье.

Сведения, содержащиеся в заглавии статьи, не должны повторяться в тексте аннотации. Следует избегать лишних вводных фраз (например, «автор статьи рассматривает...», «в статье рассматривается...»).

Исторические справки, если они не составляют основное содержание документа, описание ранее опубликованных работ и общеизвестные положения в аннотации не приводятся.

В тексте аннотации следует употреблять синтаксические конструкции, свойственные языку научных и технических документов, избегать сложных грамматических конструкций.

Гонорары за статьи в научных журналах не начисляются.

Цитирование или воспроизведение текста, созданного ChatGPT, в вашей статье

Если вы использовали ChatGPT или другие инструменты искусственного интеллекта в своем исследовании, опишите, как вы использовали этот инструмент, в разделе «Метод» или в аналогичном разделе вашей статьи. Для обзоров литературы или других видов эссе, ответов или рефератов вы можете описать, как вы использовали этот инструмент, во введении. В своем тексте предоставьте prompt - командный вопрос, который вы использовали, а затем любую часть соответствующего текста, который был создан в ответ.

К сожалению, результаты «чата» ChatGPT не могут быть получены другими читателями, и хотя невосстановимые данные или цитаты в статьях APA Style обычно цитируются как личные сообщения, текст, сгенерированный ChatGPT, не является сообщением от человека.

Таким образом, цитирование текста ChatGPT из сеанса чата больше похоже на совместное использование результатов алгоритма; таким образом, сделайте ссылку на автора алгоритма записи в списке литературы и приведите соответствующую цитату в тексте.

Пример:

На вопрос «Является ли деление правого полушария левого полушария реальным или метафорой?» текст, сгенерированный ChatGPT, показал, что, хотя два полушария мозга в некоторой степени специализированы, «обозначение, что люди могут быть охарактеризованы как «левополушарные» или «правополушарные», считается чрезмерным упрощением и популярным мифом» (OpenAI, 2023).

Ссылка в списке литературы

OpenAI. (2023). ChatGPT (версия от 14 марта) [большая языковая модель].
<https://chat.openai.com/chat>

Вы также можете поместить полный текст длинных ответов от ChatGPT в приложение к своей статье или в дополнительные онлайн-материалы, чтобы читатели имели доступ к точному тексту, который был сгенерирован. Особенno важно задокументировать созданный текст, потому что ChatGPT будет генерировать уникальный ответ в каждом сеансе чата, даже если будет предоставлен один и тот же командный вопрос. Если вы создаете приложения или дополнительные материалы, помните, что каждое из них должно быть упомянуто по крайней мере один раз в тексте вашей статьи в стиле APA.

Пример:

При получении дополнительной подсказки «Какое представление является более точным?» в тексте, сгенерированном ChatGPT, указано, что «разные области мозга работают вместе, чтобы поддерживать различные когнитивные процессы» и «функциональная специализация разных областей может меняться в зависимости от опыта и факторов окружающей среды» (OpenAI, 2023; см. Приложение А для полной расшифровки). .

Ссылка в списке литературы

OpenAI. (2023). ChatGPT (версия от 14 марта) [большая языковая модель].
<https://chat.openai.com/chat> Создание ссылки на ChatGPT или другие модели и программное обеспечение ИИ

Приведенные выше цитаты и ссылки в тексте адаптированы из шаблона ссылок на программное обеспечение в разделе 10.10 Руководства по публикациям (Американская психологическая ассоциация, 2020 г., глава 10). Хотя здесь мы фокусируемся на ChatGPT, поскольку эти рекомендации основаны на шаблоне программного обеспечения, их можно адаптировать для учета использования других больших языковых моделей (например, Bard), алгоритмов и аналогичного программного обеспечения.

Ссылки и цитаты в тексте для ChatGPT форматируются следующим образом:

OpenAI. (2023). ChatGPT (версия от 14 марта) [большая языковая модель].
<https://chat.openai.com/chat>

Цитата в скобках: (OpenAI, 2023)

Описательная цитата: OpenAI (2023)

Давайте разберем эту ссылку и посмотрим на четыре элемента (автор, дата, название и

источник):

Автор: Автор модели OpenAI.

Дата: Дата — это год версии, которую вы использовали. Следуя шаблону из Раздела 10.10, вам нужно указать только год, а не точную дату. Номер версии предоставляет конкретную информацию о дате, которая может понадобиться читателю.

Заголовок. Название модели — «ChatGPT», поэтому оно служит заголовком и выделено курсивом в ссылке, как показано в шаблоне. Хотя OpenAI маркирует уникальные итерации (например, ChatGPT-3, ChatGPT-4), они используют «ChatGPT» в качестве общего названия модели, а обновления обозначаются номерами версий.

Номер версии указан после названия в круглых скобках. Формат номера версии в справочниках ChatGPT включает дату, поскольку именно так OpenAI маркирует версии. Различные большие языковые модели или программное обеспечение могут использовать различную нумерацию версий; используйте номер версии в формате, предоставленном автором или издателем, который может представлять собой систему нумерации (например, Версия 2.0) или другие методы.

Текст в квадратных скобках используется в ссылках для дополнительных описаний, когда они необходимы, чтобы помочь читателю понять, что цитируется. Ссылки на ряд общих источников, таких как журнальные статьи и книги, не включают описания в квадратных скобках, но часто включают в себя вещи, не входящие в типичную рецензируемую систему. В случае ссылки на ChatGPT укажите дескриптор «Большая языковая модель» в квадратных скобках. OpenAI описывает ChatGPT-4 как «большую мультимодальную модель», поэтому вместо этого может быть предоставлено это описание, если вы используете ChatGPT-4. Для более поздних версий и программного обеспечения или моделей других компаний могут потребоваться другие описания в зависимости от того, как издатели описывают модель. Цель текста в квадратных скобках — кратко описать тип модели вашему читателю.

Источник: если имя издателя и имя автора совпадают, не повторяйте имя издателя в исходном элементе ссылки и переходите непосредственно к URL-адресу. Это относится к ChatGPT. URL-адрес ChatGPT: <https://chat.openai.com/chat>. Для других моделей или продуктов, для которых вы можете создать ссылку, используйте URL-адрес, который ведет как можно более напрямую к источнику (т. е. к странице, на которой вы можете получить доступ к модели, а не к домашней странице издателя).

Другие вопросы о цитировании ChatGPT

Вы могли заметить, с какой уверенностью ChatGPT описал идеи латерализации мозга и то, как работает мозг, не ссылаясь ни на какие источники. Я попросил список источников, подтверждающих эти утверждения, и ChatGPT предоставил пять ссылок, четыре из которых мне удалось найти в Интернете. Пятая, похоже, не настоящая статья; идентификатор цифрового объекта, указанный для этой ссылки, принадлежит другой статье, и мне не удалось найти ни одной статьи с указанием авторов, даты, названия и сведений об источнике, предоставленных ChatGPT. Авторам, использующим ChatGPT или аналогичные инструменты искусственного интеллекта для исследований, следует подумать о том, чтобы сделать эту проверку первоисточников стандартным процессом. Если источники являются реальными, точными и актуальными, может быть лучше прочитать эти первоисточники, чтобы извлечь уроки из этого исследования, и перефразировать или процитировать эти статьи, если применимо, чем использовать их интерпретацию модели.

Материалы журналов включены:

- в систему Российского индекса научного цитирования;
- отображаются в крупнейшей международной базе данных периодических изданий Ulrich's Periodicals Directory, что гарантирует значительное увеличение цитируемости;
- Всем статьям присваивается уникальный идентификационный номер Международного регистрационного агентства DOI Registration Agency. Мы формируем и присваиваем всем статьям и книгам, в печатном, либо электронном виде, оригинальный цифровой код. Префикс и суффикс, будучи прописанными вместе, образуют определяемый, цитируемый и индексируемый в поисковых системах, цифровой идентификатор объекта — digital object identifier (DOI).

[Отправить статью в редакцию](#)

Этапы рассмотрения научной статьи в издательстве NOTA BENE.



Содержание

Аллатов А.Н., Терлоев Э.З., Матчин В.Т. Архитектура трёхмерной свёрточной нейронной сети для детектирования факта фальсификации видеоряда	1
Зеленский А.А., Грибков А.А. Конфигурирование память-ориентированной системы управления движением	12
Трофимова В.С., Каршиева П.К., Рахманенко И.А. Метод трансферного обучения для дообучения нейронных сетей под особенности набора данных в задаче верификации диктора	26
Лукичев Р.В. Эволюция технологий семантического веба: проблемы и перспективы	37
Тиханычев О.В. Об уточнении понятия «доверенности» систем искусственного интеллекта	44
Скачёва Н.В. Анализ идиом в нейронном машинном переводе: набор данных	55
Скляр А.Я. Численные методы нахождения корней многочленов с действительными и комплексными коэффициентами	64
Англоязычные метаданные	77

Contents

Alpatov A.N., Terloev E.Z., Matchin V.T. Architecture of a three-dimensional convolutional neural network for detecting the fact of falsification of a video sequence	1
Zelenskii A.A., Gribkov A.A. Configuration of memory-oriented motion control system	12
Trofimova V.S., Karshieva P.K., Rakhmanenko I.A. Fine-tuning neural networks for the features of a dataset in the speaker verification task using transfer learning	26
Lukichev R.V. The evolution of the Semantic Web technologies: problems and prospects	37
Tikhanychev O.V. On clarifying the concept of "power of attorney" of artificial intelligence systems	44
Skacheva N.V. Analysis of Idioms in Neural Machine Translation: A Data Set	55
Sklyar A.Y. Numerical methods for finding the roots of polynomials with real and complex coefficients	64
Metadata in english	77

Программные системы и вычислительные методы*Правильная ссылка на статью:*

Аллатов А.Н., Терлоев Э.З., Матчин В.Т. Архитектура трёхмерной свёрточной нейронной сети для детектирования факта фальсификации видеоряда // Программные системы и вычислительные методы. 2024. № 3. DOI: 10.7256/2454-0714.2024.3.70849 EDN: MNOVWB URL: https://nbpublish.com/library_read_article.php?id=70849

Архитектура трёхмерной свёрточной нейронной сети для детектирования факта фальсификации видеоряда**Аллатов Алексей Николаевич**

ORCID: 0000-0001-8624-1662

доцент; кафедра ИиППО; МИРЭА - Российский технологический университет

119454, Россия, г. Москва, пр-т Вернадского, 78

✉ aleksej01-91@mail.ru

Терлоев Эмиль Зияудиновичаспирант; кафедра инструментального и прикладного программного обеспечения; МИРЭА —
Российский технологический университет

119454, Россия, г. Москва, пр-т Вернадского, 78

✉ emil199@yandex.ru

Матчин Василий Тимофеевичстарший преподаватель; институт информационных технологий; МИРЭА — Российский
технологический университет

119454, Россия, г. Москва, пр-т Вернадского, 78

✉ matchin@mirea.ru

Статья из рубрики "Базы знаний, интеллектуальные системы, экспертные системы, системы поддержки
принятия решений"

DOI:

10.7256/2454-0714.2024.3.70849

EDN:

MNOVWB

Дата направления статьи в редакцию:

26-05-2024

Дата публикации:

10-06-2024

Аннотация: В статье отражено использование нейросетевых технологий для определения фактов фальсификации содержимого видеорядов. В современном мире новые технологии стали неотъемлемой частью мультимедийной среды, однако их распространение также создало новую угрозу - возможность неправомерного использования для фальсификации содержимого видеорядов. Это приводит к возникновению серьезных проблем, таких как распространение фейковых новостей, дезинформация общества. В научной статье рассматривается данная проблема и определяется необходимость использования нейронных сетей для ее решения. В сравнении с другими существующими моделями и подходами, нейронные сети обладают высокой эффективностью и точностью в обнаружении фальсификации видеоданных благодаря своей способности к извлечению сложных признаков и обучению на больших объемах исходных данных, что особо важно при снижении разрешения анализируемого видеоряда. В рамках данной работы представлена математическая модель идентификации фальсификации аудио и видеоряда в видеозаписи, а также модель на основе трехмерной свёрточной нейронной сети для определения факта фальсификации видеоряда, путём анализа содержимого отдельных кадров. В рамках данной работы было предложено рассмотреть задачу идентификации фальсификаторов в видеозаписи, как совместное решение двух задач: идентификации фальсификации аудио- и видеоряда, а сама результирующая задача, была преобразована в классическую задачу классификации. Любая видеозапись может быть отнесена к одной из четырёх групп, описанных в работе. Только видеозаписи, относящиеся к первой группе, считаются аутентичными, а все остальные - сфабрикованными. Для повышения гибкости модели, были добавлены вероятностные классификаторы, что позволяет учитывать степень уверенности в предсказаниях. Особенность полученного решения состоит в возможности настройки пороговых значений, что позволяет адаптировать модель к различным уровням строгости в зависимости от задачи. Для определения сфабрикованных видеорядов предложена архитектура трёхмерной свёрточной нейронной сети, включающей слой предобработки и нейросетевой слой. Полученная модель обладает достаточной степенью точности определения фальсифицированных видеорядов, с учетом значительного понижения разрешения кадров. Апробация модели на тренировочном наборе данных показала долю корректного определения фальсификации видеорядов выше 70%, что заметно лучше угадывания. Несмотря на достаточную точность модель может быть доработана для более существенного увеличения доли корректных предсказаний.

Ключевые слова:

машинальное обучение, нейронные сети, свёрточные нейронные сети, фальсификация видео, дипфейки, детектирование дипфейков, фальсификация аудио, предобработка данных, обнаружение аномалий, пакетная нормализация

Введение

В настоящее время большую популярность стали набирать нейронные сети, основной целью которых является генерация изображений и голосовых аудиозаписей. Высокая степень доступности для обычного пользователя делает их более популярными.

Самыми популярными сервисами являются DALL-E от openAI, midjourney, stable diffusion, FaceApp, FaceSwap и подобные им [\[1\]](#) [\[2\]](#). Для генерации голосовых аудиозаписей используются такие популярные сервисы, как elevenlabs, Microsoft custom neural voice и speechify [\[3\]](#).

В большинстве случаев данные утилиты используются в безобидных целях, для представления получившихся изображений друзьям и знакомым, для публикации на своей странице в социальной сети, для ускорения рабочего процесса в области дизайна или для ускорения процесса создания аудио книг. Несложно представить значительное упрощение рабочих процессов в художественных сферах, в том числе и киноиндустрии. Помимо этого, возможно «воскрешение» умерших актеров при помощи инструментов генерации голосовых аудиозаписей, а также инструментов переноса лица [\[4\]](#).

С другой стороны, данные технологии ставят под вопрос необходимость актеров и художников в кинематографе, а большая доступность делает их более привлекательными инструментами для злоумышленников [\[5\]](#). Среди сценариев использования инструментов нейросетевой генерации фото и аудио рядов, возможно создание видеозаписи, в котором популярная политическая или медийная личность делает спорное заявление, способное нанести большой репутационный урон. Также возможна кража личности и дальнейшие преступные действия [\[6\]](#).

Примером кражи личности при помощи нейросетевых технологий является случай, произошедший весной 2022-го года, когда на видеохостинговом сервисе YouTube начали появляться трансляции, с участием нейросетевой копии Илона Маска, предлагающей зрителям передать ему свои криптовалютные вложения для получения их обратно с процентами. Пример трансляций представлен на рисунке 1. [\[7\]](#)



Рисунок 1 – Трансляции с участием нейросетевой копией Илона Маска на платформе YouTube [\[7\]](#)

С одной стороны, мошенническая схема достаточно очевидна. С другой стороны, незнающий пользователь может не придать этому значения, так как трансляцию ведет достаточно популярная личность, что повышает доверие пользователя к получаемой информации.

Описание модели идентификации факта фальсификации видеозаписи

Задачу идентификации фальсификации аудио и видеоряда, в рамках данной работы, сведена к классической задачи классификации.

Любая видеозапись будет определена одной из 4-х групп:

1. Фоторяды и аудио ряд видеозаписи аутентичны;
2. Фоторяды видеозаписи сфабрикован, аудио ряд видеозаписи аутентичен;
3. Фоторяды видеозаписи аутентичны, аудио ряд видеозаписи сфабрикован;
4. Фоторяды и аудио ряд видеозаписи сфабрикованы.

Видеозаписи, относящиеся к первой группе, считаются аутентичными, а не относящиеся к ней видеозаписи считаются сфабрикованными.

Обозначим через X фоторяды видеозаписи, а через A - аудиоряд видеозаписи. Выделим два классификатора C_X и C_A . Первый классификатор определяет подлинность фоторяда (аутентичен или сфабрикован), а классификатор C_A , определяет подлинность аудиоряда (аутентичен или сфабрикован). Тогда $C_X(X) = 1$, если фоторяды аутентичны, иначе $C_X(X) = 0$, то есть фоторяды сфабрикованы. В случае аудиорядов $C_A(A) = 1$, если аудиоряды аутентичны, иначе $C_A(A) = 0$, то есть аудиоряды сфабрикованы. Тогда аутентичность видеозаписи может быть определена как

$$\Phi_{authenticity}(X, A) = \begin{cases} 1, & \text{если } C_X(X) = 1 \text{ и } C_A(A) = 1 \\ 0 & \text{иначе} \end{cases}.$$

Тогда процедуру определения группы видеозаписи можно обобщить. Для этого, вначале, определим значения $C_X(X)$ и $C_A(A)$ с помощью классификаторов. Далее сравним результаты с возможными комбинациями для соотнесения с каждой группой. Так, если $(C_X(X), C_A(A)) = (1, 1)$, то видеозапись относится к первой группе, а значит она аутентична. Если $(C_X(X), C_A(A)) = (0, 1)$, то видеозапись не аутентична, так как сфабрикован видеоряд (относится ко второй группе). Если $(C_X(X), C_A(A)) = (1, 0)$, то видеозапись не аутентична, так как сфабрикована аудиодорожка(относится к третьей группе). Иначе, если $(C_X(X), C_A(A)) = (0, 0)$, то видеозапись не аутентична, так как сфабрикована аудиодорожка и видеоряд(относится к четвёртой группе). Однако такой «жёсткий» порог для классификации, требует уверенности в том, что модель верна в своих предсказаниях. Для повышения гибкости модифицируем предложенную модель, добавив вероятностные классификаторы.

Пусть $P(C_X(X)) = 1|X$ обозначает вероятность того, что фоторяды аутентичны, и $P(C_A(A)) = 1|A$. Далее Используем байесовский подход для определения совместной вероятности подлинности фоторядов и аудиоряда. Пусть θ_X и θ_A являются пороговыми значениями для определения аутентичности фоторядов и аудиоряда соответственно. Тогда подлинность фоторядов и аудиоряда на основе пороговых значений определяется через $\hat{C}_X = 1$, если $P(C_X(X) = 1|X) \geq \theta_X$, иначе $\hat{C}_X = 0$. Аналогично, $\hat{C}_A = 1$, если $P(C_A(A) = 1|A) \geq \theta_A$, иначе $\hat{C}_A = 0$. Тогда аутентичность видеозаписи может быть определена

$$\Phi_{authenticity}(X, A) = \begin{cases} 1, & \text{если } P(C_X(X) = 1|X) \geq \theta_X \text{ и } P(C_A(A) = 1|A) \geq \theta_A \\ 0 & \text{иначе} \end{cases}.$$

Возможность настройки пороговых значений θ_X и θ_A позволяет адаптировать модель к различным уровням строгости в зависимости от задачи. Например, в критически важных случаях, можно установить высокие пороги для минимизации ложных положительных результатов, что делает предложенную модель более настраиваемой. Это особенно полезно в ситуациях, где данные могут быть неоднозначными или шумными. В свою

очередь, такой подход позволяет улучшить надёжность системы, так как решения принимаются на основе распределения вероятностей, а не на основе одиночного детерминированного результата.

Технически, выявление сфабрикованных видеорядов возможно при помощи анализа кадров и поиска аномалий, при помощи анализа аудио ряда на предмет аномалий или при комбинированном анализе. В данной работе будет рассматриваться только анализ фоторядов.

Для определения сфальфицированных фоторядов можно воспользоваться трёхмерной свёрточной нейронной сетью. Свёрточные слои в нейронной сети позволяют уменьшить размерность входа, тем самым ускоряя процесс обучения. Трёхмерный свёрточный слой имеет размерность $N \times M \times K$,

где:

N — количество кадров во временной оси,

M и K — пространственные размерности (высота и ширина кадра).

Одиночный трёхмерный слой в данном случае будет декомпозирован на слой с размерностью $1 \times M \times K$, называемым пространственной свёрткой, и слой с размерностью $N \times 1 \times 1$, называемым временной свёрткой. Таким образом достигается уменьшение количества обучаемых параметров, по сравнению с использованием обычного трёхмерного слоя с размерностью $N \times M \times K$, а также показывает лучший результат при определении действий на видео [\[9\]](#).

Обозначим входное видеоданное как $X \in \mathbb{R}^{N \times M \times K \times C}$, где C — количество каналов (например, 3 для RGB-видео). Пространственная свёртка, в данном случае, применяется для обработки пространственных признаков каждого кадра, то есть двумерные свёртки применяются к каждому кадру независимо. Тогда, пусть f_{spat} — операция пространственной свёртки с ядром $f_{spat}: \mathbb{R}^{N \times M \times K \times C} \rightarrow \mathbb{R}^{N' \times M' \times K' \times C'}$

Новые высота и ширина каждого кадра после свёртки будут зависеть от размера ядра $H \times W$, шага свёртки (англ.stride) и паддинга (англ.padding). Конкретные значения M' и K' можно вычислить следующим образом (если $stride = 1$ и $padding = 0$)

$$K' = K - W + 1$$

Временная же свёртка применяется для обработки временных признаков, то есть одномерные свёртки применяются вдоль временной оси. Пусть f_{temp} — операция временной свёртки с ядром $N' = N - T + 1$

Визуализация такой свёртки показана на рисунке 2.

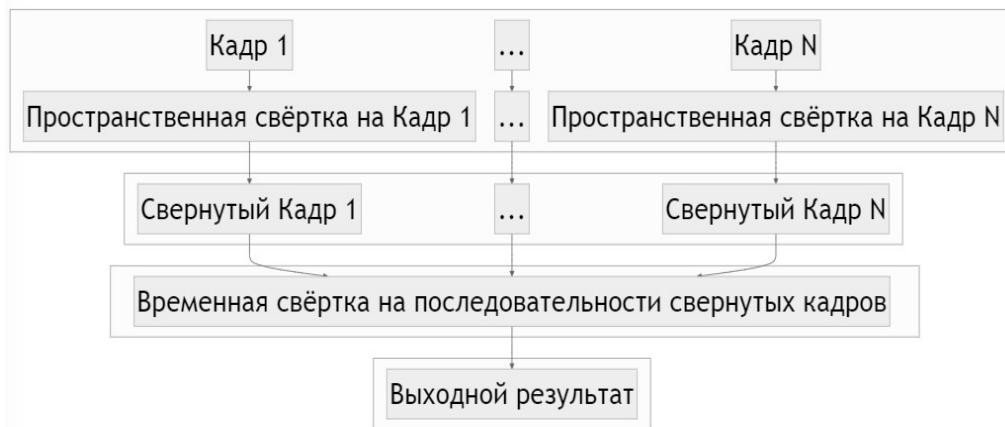


Рисунок 2 – Визуализация трехмерной свертки с декомпозицией на пространственную и временную

Архитектура искусственной нейронной сети

В качестве набора данных был взят набор ZF DeepFake Dataset^[10]. Данный набор состоит из коротких видеозаписей, из которых 199 видеозаписей фальсифицированы и 176 аутентичны (на момент написания статьи).

В представленное технологическое решение состоит из слоя предобработки видеозаписи и нейросетевого слоя. В слое предобработки видео разбивается на 10 кадров и размер каждого кадра уменьшается или увеличивается до размера 224 на 224 пикселей. Входная размерность каждого вектора видеозаписи получается $10 \times 224 \times 224 \times 3$, где последняя размерность – цветовые каналы: красный, зеленый и синий.

После предобработки видео векторы передаются в нейросетевую модель. Модель машинного обучения состоит входного слоя; слоя трехмерной свертки с декомпозицией сверток на пространственную и временную с набивкой до выходного размера, 16-ю фильтрами и размером ядра $3 \times 7 \times 7$; слоя пакетной нормализации, слоя кусочно-линейной функции активации (ReLU)^[11]; слоя изменения размерности кадров до 112×112 ; остаточного слоя с 32-я фильтрами и ядром размером $3 \times 3 \times 3$; слоя снижения размерности кадров до 64×64 ; остаточного слоя с 64-я фильтрами и ядром размером $3 \times 3 \times 3$; трехмерного слоя субдискретизации на основе среднего значения^[12], слоя выравнивания (flatten) и полносвязного слоя с 10-ю выходами. Функция ошибки модели – категориальная перекрестная энтропия с оптимизатором Adam и скоростью обучения 0.0001.

Набор данных для обучения модели состоит из 100 видеозаписей, из которых 50 фальсифицированы и 50 аутентичны. Наборы тестирования и валидации состоят из 40 записей, в каждый набор входят по 20 аутентичных и 20 фальсифицированных. Каждая конкретная видеозапись в наборах не используется более чем в одном из наборов одновременно. Обучение проводится в течение 10 эпох. Структура нейронной сети представлена на рисунке 3.

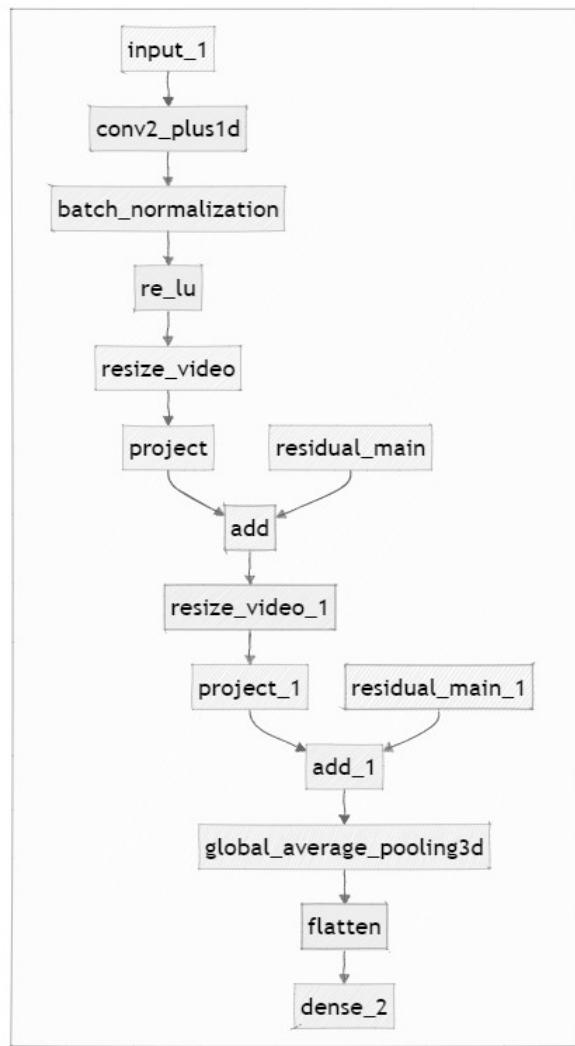


Рисунок 3 – Структура нейронной сети в виде блок-схемы

Тестирование решения

Доля корректных предсказаний (аккуратность) тренировочного набора в последнюю эпоху обучения составляла 75%. Изменение значения функции ошибки с течением обучения представлено на рисунке 4. Изменение значения аккуратности с течением обучения представлено на рисунке 5.

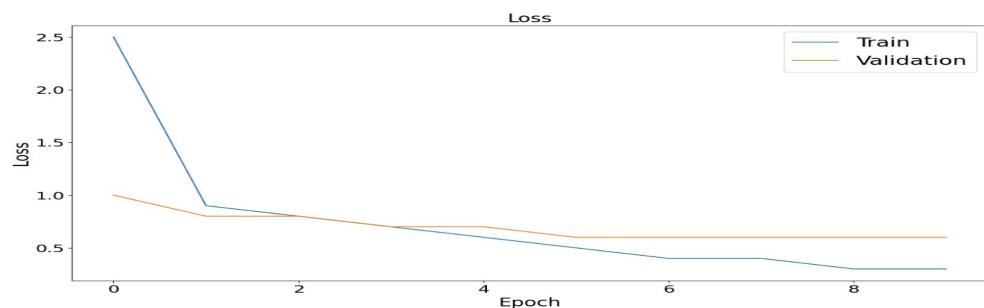


Рисунок 4 – График изменения значения функции ошибки с течением обучения для тренировочного и валидационного набора данных

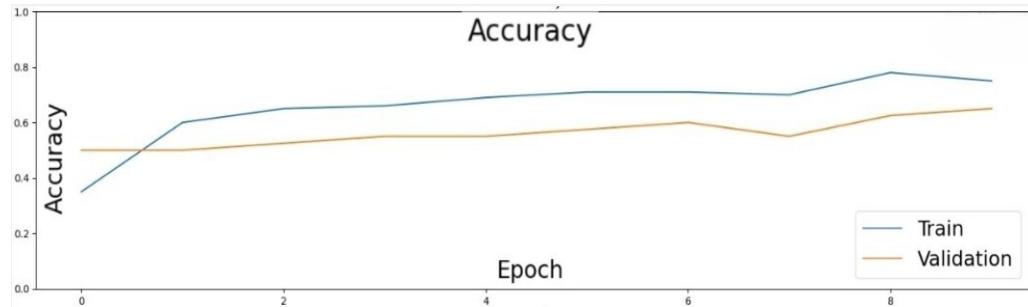


Рисунок 5 – График изменения значения аккуратности с течением обучения для тренировочного и валидационного набора данных

Несмотря на заметное увеличение аккуратности предсказаний на обучающем наборе, на валидационном наборе данных видно только незначительное улучшение. Матрица несоответствий обучающего набора данных, представлена на рисунке 6.

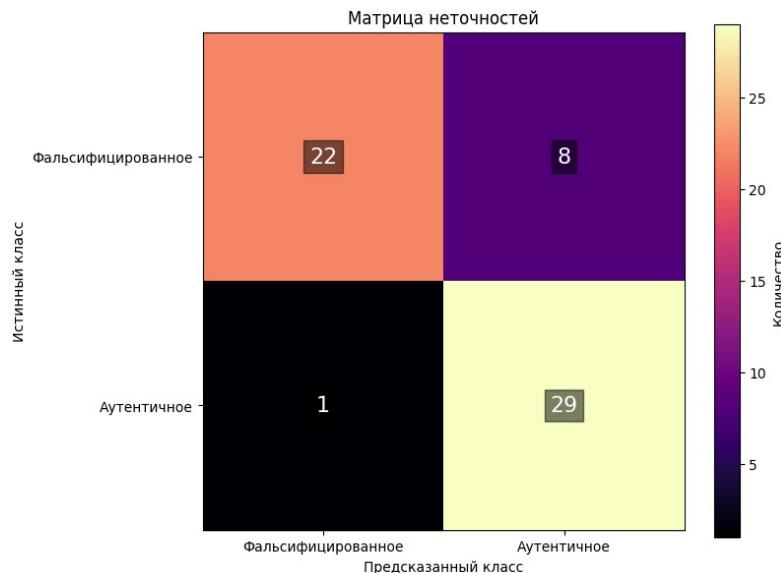


Рисунок 6 – Матрица несоответствий обучающего набора

Исходя из матрицы несоответствий видно, что модель чаще определяет видео, как аутентичное, из-за чего возникает много ложноотрицательных предсказаний. Матрица несоответствий для тестового набора данных представлена на рисунке 7.

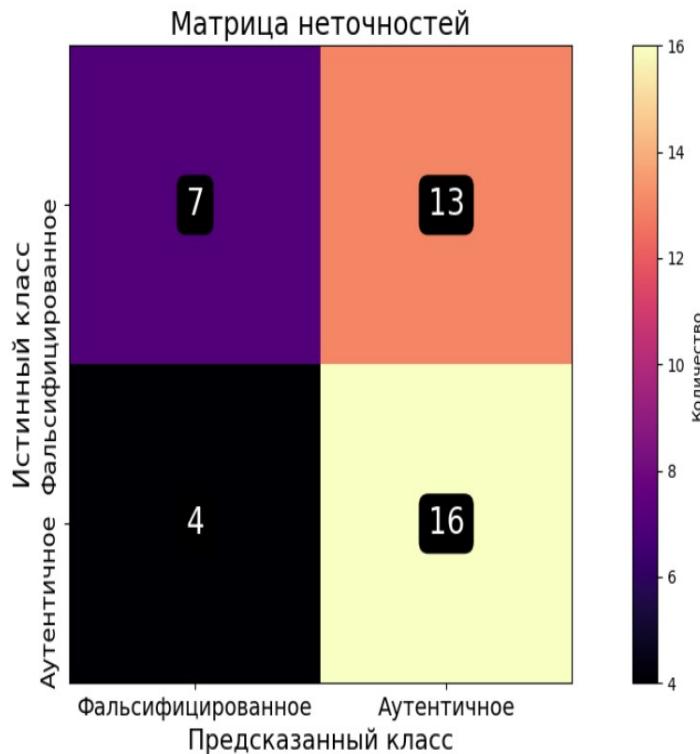


Рисунок 7 – Матрица несоответствий тестового набора

Значения точности, полноты, и F1-меры для возможных классов представлены в таблице 1.

Таблица 1 – Значения точности, полноты, и F1-меры для предсказываемых классов

Класс/метрика	Аутентичное	Фальсифицированное
Точность	0.552	0.6364
Полнота	0.8	0.35
F1-мера	0.653	0.451

Заключение

В настоящей работе представлена нейронная сеть для определения факта фальсификации видеорядов с существенной долей корректного определения. Несмотря на это, модель может быть значительно улучшена путем дополнительного наполнения обучающего набора данных и последующего увеличения доли тренировочного набора; сокращения рабочей области, путем выделения и последующего анализа конкретных зон возможной фальсификации; изменения структуры нейронной сети.

Дальнейшая работа над проблемой может также быть направлена на разработку метода определения факта фальсификации без использования моделей машинного обучения, с целью снижения риска возможных проблем с переобучением и снижения доли корректных предсказаний в случае изменений в технологии фальсификации видеорядов с использованием нейронных сетей.

Библиография

1. Beyan E.V. P., Rossy A.G.C. A review of AI image generator: influences, challenges, and future prospects for architectural field // Journal of Artificial Intelligence in Architecture. 2023. V. 2. №. 1. Рр. 53-65.
2. Huang Y. F., Lv S., Tseng K.K., Tseng P.J., Xie, X., Lin, R.F.Y. Recent advances in artificial intelligence for video production system // Enterprise Information Systems. 2023.

- V. 17. №. 11. Pp. 2246188.
3. Albert V. D., Schmidt H. J. AI-based B-to-B brand redesign: A case study // Transfer. 2023. P. 47.
4. Алиев Э. В. Проблемы использования цифровых технологий в киноиндустрии // European Journal of Arts. 2023. No1. C. 33–37. DOI: <https://doi.org/10.29013/EJA-23-1-33-37>
5. Chow, P. S. Ghost in the (Hollywood) machine: Emergent applications of artificial intelligence in the film industry // NECSUS_European Journal of Media Studies. 2020. V. 9. №. 1. Pp. 193-214.
6. Лемаи́ка С. В. Проблемы противодействия использования дипфейков в преступных целях // Юристъ-Правоведъ. 2023. No 2(105). C. 143–148.
7. Vakilinia I. Cryptocurrency giveaway scam with youtube live stream // 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). 2022. Pp. 0195-0200.
8. Tran D., Wang H., Torresani L., Ray J., LeCunY., Paluri M. A closer look at spatiotemporal convolutions for action recognition // Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2018. Pp. 6450-6459.
9. Naik K. J., Soni A. Video classification using 3D convolutional neural network // Advancements in Security and Privacy Initiatives for Multimedia Images. IGI Global. 2021. Pp. 1-18.
10. ZF DeepFake Dataset [Электронный ресурс] URL: <https://www.kaggle.com/datasets/zfturbo/zf-deepfake-dataset> (дата обращения: 20.01.2024).
11. Garbin C., Zhu X., Marques O. Dropout vs. batch normalization: an empirical study of their impact to deep learning // Multimedia tools and applications. 2020. V. 79. №. 19. Pp. 12777-12815.
12. Zhou D. X. Theory of deep convolutional neural networks: Downsampling // Neural Networks. 2020. V. 124. Pp. 319-327.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

В статье рассматривается разработка и тестирование модели трёхмерной свёрточной нейронной сети (3D CNN) для детектирования факта фальсификации видеоряда. Целью исследования является создание системы, способной эффективно распознавать аутентичные и сфабрикованные видеофайлы.

Методология включает использование 3D CNN, где свёрточные слои декомпозируются на пространственные и временные, что позволяет уменьшить количество обучаемых параметров и улучшить результаты при анализе видео. Датасет ZF DeepFake был использован для обучения и тестирования модели, что обеспечивает достаточную достоверность результатов. Модель обучалась и тестиировалась на различных наборах данных, включающих аутентичные и фальсифицированные видеозаписи.

С развитием технологий нейронных сетей и их доступностью для широких масс увеличивается риск использования этих технологий в мошеннических целях, таких как создание дипфейков. Актуальность исследования подчеркивается необходимостью разработки надёжных методов распознавания фальсификаций, что может помочь предотвратить преступления и сохранить репутацию публичных лиц.

Научная новизна работы заключается в предложении усовершенствованной архитектуры

3D CNN для детектирования фальсификаций видеоряда, а также в использовании вероятностного подхода для повышения точности классификации. Предложенная модель позволяет гибко настраивать пороговые значения для различных задач, что делает её универсальной и более точной.

Статья написана в научном стиле с чёткой структурой и логичным изложением материала. Введение подробно описывает текущие проблемы и цели исследования. Описание методологии и архитектуры модели дано детально, что позволяет понять ключевые аспекты работы. Тестирование модели и обсуждение результатов выполнены с использованием графиков и таблиц, что делает выводы прозрачными и понятными.

В заключении авторы подчеркивают эффективность предложенной модели и необходимость дальнейшего её совершенствования. Указывается на возможность улучшения модели за счёт увеличения объёма данных для обучения и изменения архитектуры нейросети. Дальнейшее исследование также может быть направлено на разработку методов детектирования фальсификаций без использования машинного обучения, что может снизить риск переобучения.

Статья будет интересна исследователям в области искусственного интеллекта, компьютерного зрения и информационной безопасности. Представленные результаты могут найти применение в различных областях, включая медиаиндустрию, правовую сферу и кибербезопасность.

Для дальнейшего развития работы предлагаю увеличить объем данных для обучения. Это включает расширение датасета за счет использования большего объема данных для обучения и тестирования модели. Важно рассмотреть использование различных источников данных, включая публичные датасеты и собственные сборы видеозаписей. Также следует диверсифицировать данные, включив различные типы фальсификаций, что позволит более полно представить все возможные сценарии.

Статья представляет собой важный вклад в область детектирования фальсификаций видеоряда и рекомендуется к публикации. Представленные результаты демонстрируют высокий потенциал предложенной модели и её применимость в реальных условиях.

Маленькое замечание: в предложении «Значения точности, полноты, и F1-меры для возможных классов представлены ...» перед «и» запятая не нужна.

Программные системы и вычислительные методы*Правильная ссылка на статью:*

Зеленский А.А., Грибков А.А. Конфигурирование память-ориентированной системы управления движением // Программные системы и вычислительные методы. 2024. № 3. DOI: 10.7256/2454-0714.2024.3.71073 EDN: TTQBVA URL: https://nbpublish.com/library_read_article.php?id=71073

Конфигурирование память-ориентированной системы управления движением

Зеленский Александр Александрович

ORCID: 0000-0002-3464-538X

кандидат технических наук

ведущий научный сотрудник; Научно-производственный комплекс "Технологический центр"

124498, Россия, г. Москва, пл. Шокина, 1, строение 7

[✉ zelenskyaa@gmail.com](mailto:zelenskyaa@gmail.com)**Грибков Андрей Армович**

ORCID: 0000-0002-9734-105X

доктор технических наук

ведущий научный сотрудник; Научно-производственный комплекс "Технологический центр"

124498, Россия, г. Москва, пл. Шокина, 1, строение 7

[✉ andarmo@yandex.ru](mailto:andarmo@yandex.ru)[Статья из рубрики "Параллельные алгоритмы решения задач вычислительной математики"](#)**DOI:**

10.7256/2454-0714.2024.3.71073

EDN:

TTQBVA

Дата направления статьи в редакцию:

19-06-2024

Дата публикации:

25-07-2024

Аннотация: В статье исследуются возможности конфигурирования цикла управления, т.е. определения распределения интервалов времени, необходимых для выполнения

отдельных операций управления, по потокам исполнения, обеспечивающего реализуемость управления. Параллельное выполнение операций управления, там, где это допускается алгоритмом управления, в случае успешного конфигурирования цикла управления позволяет существенно снизить его длительность. Объектом исследования в данной статье являются системы управления с объектно-ориентированной архитектурой, предполагающей комбинированную вертикально-горизонтальную интеграцию функциональных блоков и модулей, распределяющих между собой все задачи управления. Данная архитектура реализуется посредством акторной инструментальной модели с использованием метапрограммирования. Такие системы управления наилучшим образом обеспечивают сокращение длительности цикла управления за счет параллельного выполнения вычислительных и других операций управления. Рассматриваются несколько подходов к конфигурированию цикла управления: без оптимизации, с комбинаторной оптимизацией по времени, с комбинаторной оптимизацией по ресурсам системы. Также достижение конфигурации, близкой к оптимальной, может быть достигнуто за счет использования адаптивного конфигурирования. Исследования показывают, что задача конфигурирования цикла системы управления имеет несколько вариантов решения. Практическое получение решения задачи конфигурирования в случае комбинаторной оптимизации связано с существенными сложностями, обусловленными высокой алгоритмической сложностью задачи и большим объемом потребных вычислений, быстро растущим по мере увеличения числа операций на этапах цикла управления. Возможным средством преодоления этих сложностей является использование стохастических методов, резко снижающих потребный объем вычислений. Также существенное снижение сложности задачи конфигурирования цикла системы управления можно добиться при использовании адаптивного конфигурирования, имеющего два варианта реализации. Первый вариант - это конфигурирование цикла системы управления в реальном времени. Второй вариант - это определение квазиоптимальной конфигурации на основе многократного конфигурирования с разными исходными данными и последующего сравнения получаемых результатов.

Ключевые слова:

система управления, память-ориентированная, конфигурирование, оптимизация, цикл, элементы, операции управления, потоки исполнения, адаптивный, методы сортировки

Исследование выполнено при поддержке Российского научного фонда по гранту № 24-19-00692, <http://rscf.ru/project/24-19-00692/>

Введение

Проблематика управления движением в реальном времени, необходимого для современных промышленных роботов, станков и другого технологического оборудования, складывается из нескольких взаимосвязанных составляющих. Центральной из них является проблема реализуемости управления, заключающаяся в определении (и обеспечении) возможности управления движением с быстродействием, необходимым для обеспечения заданной точности воспроизведения скорости и траектории движения. Исследования, проведенные авторами, показали [1, 2], что проблема реализуемости управления обуславливает крайне жесткие требования к быстродействию систем управления сложными объектами. К таким сложным объектам управления, в частности, относятся прецизионное технологическое оборудование с

большим числом одновременно управляемых осей (пять осей и более).

Одной из специфических особенностей управления движением в реальном времени является требование высокого быстродействия системы управления, а не высокой производительности. Производительность системы управления (как и любой вычислительной машины) определяется как среднее число вычислительных операций (или иных операций управления) за единицу времени, причем этот расчет обычно выполняется для достаточно больших объемов вычислений (операций управления) за значительный интервал времени.

Управление в реальном времени в большинстве случаев реализуется в виде сравнительно небольшого набора операций, однако этот набор операций должен быть выполнен за чрезвычайно малый интервал времени – цикл управления, не превышающий единицы миллисекунд, а для наиболее сложных объектов управления сокращающийся до микросекунд. В результате критичным для реализуемости управления становится обеспечение параметра быстродействия системы управления – величины, обратной длительности цикла управления. В более точной формулировке быстродействие – это количество единиц сложности объекта, обслуживаемых системой управления за единицу времени.

Согласно методологии, разработанной авторами с соавторами, сложность объекта управления зависит от числа типов элементов и среднего количества элементов одного типа в системе, числа типов связей и среднего количества связей одного типа в системе, числа контролируемых параметров, посредством которых описывается состояние отдельного элемента системы, а также числа отслеживаемых состояний контролируемого параметра [\[1\]](#).

Быстродействие системы управления движением сложным объектом ограничено действием трех групп факторов, получивших образные названия стена мощности (the power wall), стена памяти (the memory wall) и стена частоты (the frequency wall) [\[3\]](#). Для практического преодоления этих трех «стен» и повышения в результате быстродействия системы управления движения должны быть решены три задачи, имеющие решение за счет совершенствования архитектуры вычислительных машин: во-первых, необходимо уменьшить объем обрабатываемого потока данных за счет использования параллельных вычислений; во-вторых, необходимо увеличить скорость передачи данных между элементами вычислительной машины за счет применения технологий обработки в памяти (PIM) [\[4\]](#) или обработки вблизи памяти (NMC) [\[5\]](#); в-третьих, необходимо устранить очереди при одновременном обращении к одной памяти нескольких вычислительных устройств за счет физического разделения памяти между устройствами.

Решать указанные три задачи позволяет память-ориентированная архитектура системы управления движением, при которой данные в процессе вычислений не перемещаются между процессором и памятью, а остаются в памяти, в которую интегрируется процессор. Ключевым отличием память-ориентированной архитектуры от традиционной (процессорно-ориентированной) заключается в отказе от организации управления, при которой оно разделяется на несколько уровней (например, стратегический, тактический и исполнительный). Вместо этого имеет место комбинированная вертикально-горизонтальная интеграция функциональных блоков и модулей, распределяющих между собой все задачи управления.

На программно-алгоритмическом уровне реализация память-ориентированной архитектуры обеспечивается акторной инструментальной моделью, в которой каждый

актор (элемент акторной модели) соответствует элементу системы управления движением, который находит свое отражение в виде задержки и цикле системы управления. Причем этот элемент может быть как реальным, соответствующим функциональному модулю в составе системы управления движения, так и виртуальным, формируемым в памяти (общей или локальной для отдельного функционального модуля) для выполнения частной вычислительной задачи, преобразования данных или генерации управляющих команд для периферийных устройств. Условию максимальной автономности акторов, асинхронно обменивающихся данными в процессе совместной реализации функции управления, соответствует программная реализация системы управления движением на основе метапрограммирования [6]. В этом случае каждый из акторов генерируется по необходимости в процессе управления в виде отдельной программы, запускаемой в общей памяти или в локальной памяти отдельного функционального модуля.

Практическое решение проблемы реализуемости управления движением сложного объекта сводится к распараллеливанию выполнения операций управления и, соответственно, к определению оптимального взаимного распределения затрат времени на отдельные операции с учетом установленных алгоритмических ограничений по последовательности их выполнения. Заметим, что именно с наличием таких ограничений связано существование закона Амдала [7].

Задачей данной статьи является решение указанной задачи оптимального распределения затрат времени на отдельные операции в рамках цикла управления. Кроме того, необходимо определить возможные алгоритмы практической реализации оптимизационных расчетов в том случае, если их объем окажется существенным.

Подходы к конфигурированию цикла системы управления

На рис. 1 приведена диаграмма оптимизированного цикла системы управления промышленного робота [2]. Определение оптимального взаимного распределения затрат времени на отдельные операции представляет собой задачу конфигурирования цикла системы управления движением. Данная задача может быть решена на основе трех основных подходов, каждый из которых, по мнению авторов, имеет право на существование.

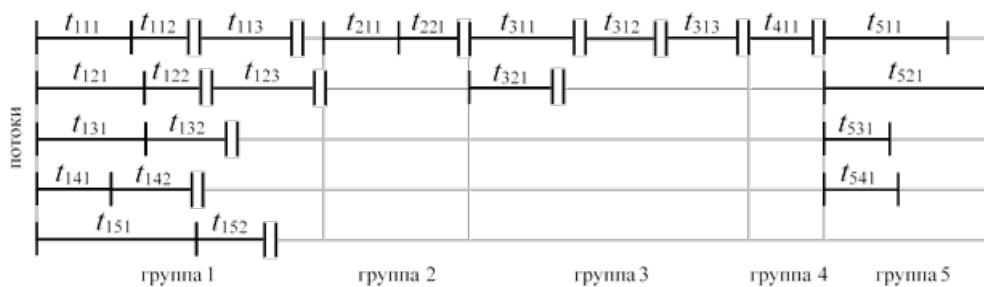


Рисунок 1. Диаграмма оптимизированного цикла СУ промышленного робота:

$t_{111}, t_{121}, t_{131}, t_{141}, t_{151}$ – задержки коммуникационной сети при передаче данных от датчиков и сенсоров в модули очувствления; t_{112} – задержка системы технического зрения, СТЗ (без ИНС); t_{113} – задержка искусственной нейронной сети (ИНС) при обработке данных для СТЗ; t_{122} – задержка системы распознавания речи, СРР (без ИНС); t_{123} – задержка ИНС при обработки данных для СРР; $t_{132}, t_{142}, t_{152}$ – задержки

модулей обработки данных от датчиков положения, скорости и др.; t_{211}, t_{221} — задержки ядра: интерпретатора (не в каждом цикле) и задержка при обработке данных от модулей очувствления; t_{311} — задержка модуля трансформации; t_{312} — задержка модуля интерполяции; t_{313} — задержка модуля эквидистантной коррекции; t_{321} — задержка модуля предварительного просмотра; t_{411} — задержка модуля разгона-торможения; $t_{511}, t_{521}, t_{531}, t_{541}$ — задержки регуляторов исполнительных устройств; $||$ — задержка записи/чтения данных в общей памяти.

Первый подход, назовем его конфигурацией без оптимизации, основан на воспроизведении заданной конфигурации цикла управления с известным (определенным на основе экспертной оценки разработчика) распределением элементов (задержек отработки операций) по группам и потокам исполнения. Примером реализации такого подхода является приведенная на рис. 1 диаграмма оптимизированного цикла системы управления промышленного робота.

Если (для получения более простого представления) включить задержки чтения/записи данных в общей памяти в задержки соответствующих элементов, то длительность цикла системы управления в данном случае определяется следующим образом:

$$\tau = \sum_{u=1}^{u_{\max}} \max \left(\bigcup_{p=1}^{p_{u\max}} \sum_{j=1}^{m_{up}} t_{uj}^{(p)} \right), \quad (1)$$

где u_{\max} — число групп операций в цикле управления; $p_{u\max}$ — количество параллельных потоков обработки данных в u -ой группе; m_{up} — количество элементов в p -ом параллельном потоке обработки данных в u -ой группе.

Ресурсы, потребные для реализации данной конфигурации, должны быть не меньше суммарных ресурсов, необходимых всем элементам во всех потоках исполнения в каждый момент времени цикла управления.

Второй подход, назовем его конфигурацией с комбинаторной оптимизацией по времени, основан на конфигурировании цикла управления из множества элементов, разбитых на группы (определяемые последовательностью и взаимосвязанностью элементов в рамках алгоритма управления), но не распределенных по потокам исполнения. В процессе оптимизации устанавливается такое распределение по потокам исполнения, при котором длительность цикла управления будет минимальной.

Рассмотрим порядок оптимизации цикла управления по времени.

u -я группа операций формируется из множества элементов $A_u = \{a_{u1}, a_{u2}, \dots, a_{uj}, \dots, a_{un(u)}\}$, каждый из которых $a_{ui} = \{t_{ui}, w_{ui}\}$ характеризуется временем t_{ui} , необходимым для выполнения операции управления, соответствующей данному элементу, и ресурсами w_{ui} , необходимыми для реализации данного элемента. В качестве ресурсов элемента может рассматриваться потребный объем памяти и/или вычислительная мощность (число элементарных операций, необходимых для выполнения рассматриваемой операции управления, за единицу времени ее выполнения).

Элементы a_{ui} не могут быть задействованы в произвольном порядке. Порядок определение порядка задействования элементов задается в виде матрицы отношений

$$K_u = \begin{pmatrix} k_{11} & k_{12} & \dots & k_{1n_u} \\ k_{21} & k_{22} & \dots & k_{2n_u} \\ \dots & \dots & \dots & \dots \\ k_{n_u 1} & k_{n_u 2} & \dots & k_{nn_u} \end{pmatrix}, \quad (2)$$

в которой каждый элемент $k_{ij} \in \{-1, 0, 1\}$ определяет последовательность элементов i и j : «-1» – j потом i , «0» – порядок не имеет значения, «1» – i потом j .

Из элементов u -ой группы формируется выборка для отдельного p -го потока, в которой каждый ее элемент отожествляется с элементом из множества u -й группы:

$$a_{uj}^{(p)} \equiv a_{ui}, \quad (3)$$

причем элементы не повторяются

$$\forall h \neq j : a_{uh}^{(p)} \not\equiv a_{ui} \quad (4)$$

и реализуются в правильном порядке относительно всех предшествующих по времени элементов во всех (ранее определенных) потоках, включая свой,

$$\forall p, d : K_u \left(i(a_{u(j_d-1)}^{(d)}), i(a_{uj}^{(p)}) \right) \neq -1, \quad (5)$$

где величина j_d для потока d определяется исходя из условия

$$\sum_{q=1}^{j_d} t_{uq}^{(d)} > \sum_{q=1}^{j_p} t_{uq}^{(p)} > \sum_{q=1}^{j_d-1} t_{uq}^{(d)},$$

$i(a_{uj}^{(p)})$ – значение номера элемента в исходном множестве u -ой группы, соответствующее элементу $a_{uj}^{(p)}$ в выборке для p -го потока.

Если $K_u = -1$, то все элементы, уже включенные в выборку (от 1 до j), переставляются исходя из условия (5).

Каждый элемент в выборке для p -го потока, как и в исходном множестве u -ой группы, определяется потребным временем и ресурсами:

$$a_{uj}^{(p)} = \left\{ t_{uj}^{(p)}, w_{uj}^{(p)} \right\}. \quad (6)$$

Состав выборки для p -го потока u -ой группы представляет собой множество из m_{up} элементов

$$A_u^{(p)} = \left\{ a_{u1}^{(p)}, \dots, a_{uj}^{(p)}, \dots, a_{um_{up}}^{(p)} \right\}, \quad (7)$$

причем сумма элементов всех потоков равна числу элементов nu в исходном множестве u -ой группы и элементы в различных потоках (множествах) не повторяются:

$$\sum_{p=1}^{P_{u \max}} m_{up} = n_u, \forall d \neq p : \nexists (A_u^{(p)} \wedge A_u^{(d)}). \quad (8)$$

Последнее условие обеспечивается автоматически ввиду последовательного (без пропусков) включения элементов из исходного множестве u -ой группы в выборки для 1-

го, 2-го и последующих потоков.

Для реализации дальнейшей оптимизации необходимо задать функцию суммарных ресурсов, задействованных элементами во всех потоках u -ой группы в заданный момент времени:

$$w_u(t) = \sum_{p=1}^{P_{u\max}} w_{uj}^{(p)}, \quad (9)$$

где номер соответствующего элемента j определяется из выражения

$$\sum_{q=1}^j t_{uj}^{(p)} > t.$$

Теперь можно определить длительность части цикла управления, соответствующей выполнению операций u -ой группы в виде множества A_u с заданной последовательностью элементов:

$$\tau_u = \max \left(\bigcup_{p=1}^{P_{u\max}} \sum_{j=1}^{m_{up}} t_{uj}^{(p)} \right), \quad (10)$$

причем должно выполняться условие

$$w_u(t \leq \tau_u) \leq w_{u\max}.$$

Следует обратить внимание, что зависимости (9) и (10) не предполагают возможности произвольных задержек между элементами для изменения сочетаний элементов разных потоков для минимизации потребных ресурсов.

Для получения минимального времени отработки операций u -ой группы необходимо провести аналогичный анализ для множества A_u с другими последовательностями элементов (число вариантов перестановок $N_u = n_u!$) и другими значениями числа элементов в потоках.

Число вариантов перестановок быстро растет по мере роста числа элементов в множестве. Например, если при $n_u = 10$ (реализуемом при объединении некоторых связанных элементов, см. рис. 1) имеем приемлемое значение $N_u = n_u! = 3.6 \times 10^6$, то при $n_u = 15$ имеем $N_u = n_u! = 1.3 \times 10^{12}$, а при $n_u = 20$ имеем $N_u = n_u! = 2.4 \times 10^{18}$.

Значения числа элементов в потоках для группы u могут быть заданы в виде множества

$$M_{uj} = \left\{ m_{u1}^{(j)} \quad m_{u2}^{(j)} \quad \dots \quad m_{up}^{(j)} \quad \dots \quad m_{up_{\max}}^{(j)} \right\}, \quad (11)$$

где p_{\max} – число потоков в u -ой группе, причем

$$\sum_{p=1}^{P_{\max}} m_{up}^{(j)} = n_u.$$

Число вариантов множества M_{uj} представляет собой количество упорядоченных разбиений числа n_u элементов в исходном множестве u -ой группы и равно $j_{\max} = 2^{n_u-1}$, что в случае $n_u = 10$ дает $j_{\max} = 512$, в случае $n_u = 15$ – $j_{\max} = 1.6 \times 10^4$, а в случае $n_u = 20$ – $j_{\max} = 5.2 \times 10^5$.

Минимальное время отработки операций u -ой группы:

$$\tau_u^{\min} = \min \left[\bigcup_{i=1, j=1}^{N_u, j_{\max}} \tau_u(A_{ui}, M_{uj}) \right], \quad (12)$$

где M_{uj} – j -й (из j_{\max}) вариант множества M_{uj} , A_{uj} – i -я (из N_u) перестановка исходного множества A_u .

Исходя из приведенных выше зависимостей, число вариантов, задаваемых $\tau_u(A_{ui}, M_{ui})$, равно $N_u \cdot j_{\max} = n_u! \cdot 2^{n_u-1}$ и составляет в зависимости от числа элементов в исходном множестве u -ой группы от 1.8×10^9 при $n_u = 10$ до 2.1×10^{16} при $n_u = 15$ и 1.2×10^{24} при $n_u = 20$. Последние числа вариантов слишком велики и не могут быть на практике проанализированы.

Максимальная сложность данного оптимизационного алгоритма соответствует нотации $O(n! \cdot 2^{n-1} \cdot n^2)$, где n^2 – сомножитель, связанный с сортировкой выборки перебором из исходного множества u -ой группы. Данный характер сложности оказывает критическое негативное влияние на практическую реализуемость полного перебора всех возможных вариантов. Такая ситуация является типичной для задач комбинаторной оптимизации.

Средством решения данной задачи комбинаторной оптимизации при большом числе элементов в группе является использование стохастических методов со случайной генерацией последовательности элементов в исходном множестве A_u для u -ой группы и множества M_{uj} значений числа элементов в потоках этой группы. Для повышения эффективности случайной выборки можно целенаправленно обеспечивать существенное различие последовательностей элементов, разбиение групп на потоки с сопоставимым числом элементов, ограничивать вариативность числа потоков исходя из доступных ресурсов для всей группы и средних значений ресурсов отдельных элементов, собирать потоки из элементов исходного множества не последовательно, а параллельно (поочередно добавляя элементы в разные потоки) и т.д. В результате N_u и j_{\max} не будут зависеть от n_u (будут примерно постоянными, устанавливаемыми исходя из экспертной оценки разработчика), сложность алгоритма существенно снизится и может соответствовать сложности сортировки, имеющей в случае простого перебора нотацию $O(n^2)$.

Критерием оценки репрезентативности случайной выборки может служить монотонность изменений расчетных значений вблизи минимального значения длительности цикла управления.

Длительность цикла управления определяется следующим образом:

$$\tau = \sum_{u=1}^{u_{\max}} \tau_u^{\min}. \quad (13)$$

Третий подход, назовем его конфигурацией с комбинаторной оптимизацией по ресурсам, также, как и второй, основан на конфигурировании цикла управления из множества элементов, разбитых на группы, но не распределенных по потокам исполнения. В процессе оптимизации устанавливается такое распределение по потокам исполнения, при котором длительность цикла управления не превысит заданную, а потребные ресурсы будут минимальными.

Рассмотрим порядок оптимизации цикла управления по ресурсам.

Для данного варианта оптимизации справедливы формулы (2-10), полученные нами ранее для варианта оптимизации по времени.

Ресурсы, потребные для реализации всех операций u -ой группы цикла управления:

$$w_u = \max \left[w_u(t=0 \dots \tau_u) \right]. \quad (14)$$

При реализации оптимальной по ресурсам конфигурации цикла управления (минимальные) потребные ресурсы составят

$$w = \max \left(\bigcup_{u=1}^{u_{\max}} \min \left[\bigcup_{i=1, j=1}^{N_u, j_{\max}} w_u(A_{ui}, M_{uj}) \right] \right). \quad (15)$$

причем длительность цикла управления не должна превысить заданное максимальное значение:

$$\tau = \sum_{u=1}^{u_{\max}} \left[\max \left(\bigcup_{p=1}^{P_u \max} \sum_{j=1}^{m_{up}} t_{uj}^{(p)} \right) \right] \leq \tau_{\max}.$$

Конфигурация цикла управления с оптимизацией по времени или ресурсам – оптимизационные задачи, относящаяся к группе NP-полных задач. Ограничениями задачи конфигурирования цикла управления с комбинаторной оптимизацией являются, соответственно, допустимые (вычислительные) ресурсы системы управления или длительность цикла управления.

Ввиду высокой алгоритмической сложности рассматриваемых оптимизационных задач ни один из типовых оптимизационных алгоритмов, используемых для комбинаторной оптимизации, не может в полной мере ее воспроизвести. В частности, не удается построить оптимизационные алгоритмы на взвешенных ориентированных графах, например алгоритме Декстры, алгоритме Флойда-Уоршелла и др. Необходимость использования взвешенных ориентированных графов обусловлена тем, что элементы конфигурации цикла управления имеют веса и ограничения по последовательности. Проблема в построении оптимизационных алгоритмов связана с тем, что в известных алгоритмах вес элементов (ребер графа) влияет на построение маршрута локально, определяя последовательность пары вершин, между тем в нашей оптимизационной задаче ребра графа (время или ресурсы элемента) хотя и имеют вес, но он оказывается лишь в сочетании с весами других ребер графа. В результате один и тот же вес ребра в разных позициях конфигурации цикла управления оказывает разное влияние на длительность цикла управления или потребные для управления ресурсы.

Существующие алгоритмы оптимизации, однако, могут быть использованы для решения локальных задач в рамках комбинаторной оптимизации. В частности, использование алгоритмов сравнения при упорядочивании выборки для p -го потока u -ой группы позволяет снизить время выполнения алгоритма сортировки с $O(n^2)$ до $O(n \log n)$ [8, с. 206-207].

К числу лучших по времени и памяти алгоритмов сортировки относятся [9]: merge sort (сортировка слиянием [8, с. 181-192]), heapify и heap sort (пирамидальная сортировка [10]) – лучшее, среднее и худшее время $O(n \log n)$, память $O(n)$; quick sort (сортировка Хоара) – лучшее и среднее время $O(n \log n)$, худшее время $O(n^2)$, память $O(\log n)$; tree sort (сортировка с помощью двоичного дерева) – лучшее время $O(n)$, среднее и худшее

время $O(n \log n)$, память $O(n)$; **timsort** (гибридный алгоритм сортировки, сочетающий сортировку вставками и сортировку слиянием) и **binary insertion sort** (гибридный алгоритм сортировки, сочетающий сортировку вставками с бинарным поиском места вставки [11]) – лучшее время $O(n)$, среднее и худшее время $O(n \log n)$, память $O(n)$.

Адаптивное конфигурирование цикла системы управления

Наряду с рассмотренными вариантами комбинаторной оптимизации конфигурации цикла управления по времени или ресурсам, также может быть использовано адаптивное конфигурирование цикла управления. Это конфигурирование, как и рассмотренные варианты оптимизации, применяется к группам элементов, а итоговые результаты получаются после объединения полученных для них результатов. При этом в каждый момент времени задействуются максимально доступные ресурсы, обеспечивая тем самым минимизацию времени исполнения всех элементов группы.

В каждый момент времени нам известны элементы, реализуемые на каждом потоке u -ой группы $\{a_u^{(1)}, a_u^{(2)}, \dots, a_u^{(p)}, \dots, a_u^{(p_{\max})}\}$, причем каждый элемент взят из исходного множества элементов u -ой группы, после чего в исходном множестве он помечается как использованный. Это реализуется заданием для элементов a_{ui} в исходном множестве, наряду со временем и ресурсами, параметра $z_{ui} \in \{1, 0\}$ «использован» / «не использован».

Когда время выполнения какого-либо из элементов заканчивается, на его место (в тот же поток) назначается неиспользованный элемент из исходного множества, для которого выполняются два условия. Согласно первому условию, назначаемый q -ый элемент должен быть задан к выполнению не позже любого из оставшихся незадействованных элементов, т.е.

$$\forall 1 \leq j \leq n_u, j \neq i(a_u^{(q)}), z_{uj} = 0 : K_u(i(a_u^{(q)}), j) \neq -1, \quad (16)$$

а согласно второму условию, назначаемый q -ый элемент должен обладать максимальной потребностью в ресурсах из всех оставшихся неиспользованными в исходном множестве, но при этом суммарные потребные ресурсы всех потоков (включая поток назначаемого элемента) не должны превышать допустимого максимального значения, т.е.

$$w_u^{(q)} = \max \left(\bigcup_{i=1}^{n_u} w_{ui}, z_{ui} = 0, \sum_{p=1}^{p_{\max}} w_u^{(p)} \leq w_u^{\max} \right). \quad (17)$$

Если после назначения нового элемента в системе остаются ресурсы, то по тому же алгоритму назначается еще один элемент, для которого добавляется новый поток.

После завершения выполнения элементов 1-ой группы запускается выполнение элементов 2-ой группы и т.д. Цикл управления завершается тогда, когда заканчиваются элементы во множествах всех групп.

Запуск адаптивного конфигурирования осуществляется с заданных экспертной оценкой разработчика стартовых элементов (т.е. элементов, выполняемых, согласно алгоритмы управления, первыми). Потоки добавляются пока задействованные во всех потоках ресурсы меньше имеющихся в наличии в системе.

Для реализации адаптивного конфигурирования, которое в общем случае не гарантирует минимизации длительности цикла управления, могут быть использованы указанные выше алгоритмы оптимизации сортировки. Они применимы как для реализации условия (16),

так и (17).

Адаптивное конфигурирование может быть использовано для решения двух различающихся задач: во-первых, для распределения элементов цикла управления в процессе управления, т.е. конфигурирования в реальном времени; во-вторых, для определения квазиоптимальной конфигурации цикла управления на основе многократного адаптивного конфигурирования и сравнения его результатов.

В случае адаптивного конфигурирования в реальном времени частота переопределения конфигурации ограничена интервалом времени, необходимым для его расчета. Вероятно, этот интервал времени много больше длительности цикла управления, т.е. переопределение конфигурации цикла управления происходит один раз за сотни или даже тысячи циклов управления. По мнению авторов, это является допустимым, поскольку в том случае, когда вследствие корректирования алгоритма управления требуется переопределение времени выполнения элементов, оно происходит за время, соответствующее (например, для систем управления движением) существенным изменениям объекта управления (например, перемещениям рабочего органа).

Многократное адаптивное конфигурирование цикла управления предполагает его выполнение с разными перестановками исходного множества элементов заданной группы.

По результатам j -го адаптивного конфигурирования формируется множество

$$B_{uj} = \{b_{uj1}, \dots b_{ujn}, \dots b_{ujn_u}\}, \quad (18)$$

образованное элементами $b_{ujn} = \{u, i, p\}$, где $\{u, i, p\}$ – номер группы, номер элемента в исходном множестве элементов u -ой группы и номер потока для n -го элемента множества B_{uj} .

Интервал времени, необходимый для выполнения элементов p -го потока u -ой группы, определяется из условия

$$\forall b_{ujn} = \{u, i, p\}, 1 \leq n \leq n_u : \tau_{uj}^{(p)} = \sum_{i=1}^{n_u} t_{ui}. \quad (19)$$

Длительность части цикла управления, за которую выполняются все элементы u -ой группы, соответствующая j -ой реализации адаптивного конфигурирования, определяется из условия

$$\tau_{uj} = \max \left(\bigcup_{p=1}^{p_{\max}} \tau_{uj}^{(p)} \right). \quad (20)$$

Наименьшая длительность части цикла управления, за которую выполняются все элементы u -ой группы, из всех j_{\max} реализаций адаптивного конфигурирования:

$$\tau_u = \min \left(\bigcup_{j=1}^{j_{\max}} \tau_{uj} \right). \quad (21)$$

Наименьшая длительность цикла управления, получаемая по результатам многократного адаптивного конфигурирования:

$$\tau = \sum_{u=1}^{u_{\max}} \tau_u. \quad (22)$$

Выводы

Резюмируем проведенное в статье исследование:

1. Управление сложными объектами в реальном времени требует для обеспечения своей реализуемости высокого быстродействия системы управления. Высокое быстродействие предполагает возможность выполнения ограниченного (сравнительно небольшого) набора операций управления за малый интервал времени – цикл управления, не превышающий единицы миллисекунд или даже (для наиболее сложных объектов управления) микросекунд.
2. Требуемое сокращение длительности цикла управления может быть достигнуто за счет параллельного выполнения вычислительных и других операций управления. Наилучшим образом это обеспечивается при использовании память-ориентированной архитектуры системы управления, реализуемой посредством акторной инструментальной модели, программируемой средствами метапрограммирования.
3. На практике задача обеспечения реализуемости управления сводится к конфигурированию цикла управления, т.е. определению взаимного распределения затрат времени на различные операции управления по потокам исполнения согласно заданным ограничениям по их последовательности и доступным вычислительным ресурсам системы.
4. Возможны три подхода к конфигурированию цикла управления: без оптимизации с конфигурацией, заданной на основе экспертной оценки разработчика, а также конфигурирование с комбинаторной оптимизацией по времени или по ресурсам системы. Комбинаторная оптимизация (в обоих случаях) характеризуется крайне высокой алгоритмической сложностью и на практике должна быть упрощена введением стохастических методов.
5. Менее точным и достоверным подходом к конфигурированию цикла системы управления является адаптивное конфигурирование, которое может реализоваться в реальном времени, либо, при многократном использовании, в качестве инструмента определения квазиоптимальной конфигурации цикла системы управления.

Библиография

1. Зеленский А.А., Кузнецов А.П., Илюхин Ю.В., Грибков А.А. Реализуемость управления движением промышленных роботов, станков с ЧПУ и мехатронных систем. Часть 1 // Вестник машиностроения. 2022. №11. С. 43-51
2. Зеленский А.А., Кузнецов А.П., Илюхин Ю.В., Грибков А.А. Реализуемость управления движением промышленных роботов, станков с ЧПУ и мехатронных систем. Часть 2 // Вестник машиностроения. 2023. №3. С. 213-220
3. Cell Broadband Engine Programming Tutorial. Version 2.0. IBM Systems and Technology Group, December 15, 2006. URL:
https://arcb.csc.ncsu.edu/~mueller/cluster/ps3/CBE_Tutorial_v2.0_15December2006.pdf
4. Ghose S., Hsieh K., Boroumand A., Ausavarungnirun R., Mutlu O. Enabling the Adoption of Processing-in-Memory: Challenges, Mechanisms, Future Research Directions. 2018. URL:
<https://arxiv.org/abs/1802.00320>
5. Singh G., Chelini L., Corda S., Awan A.J., Stuijk S., Jordans R., Corporaal H., Boonstraaz A. Near-Memory Computing: Past, Present, and Future. August 2019, Microprocessors and Microsystems 71. URL: https://www.researchgate.net/publication/335028505_Near-Memory_Computing_Past_Present_and_Future
6. Зеленский А.А., Ивановский С.П., Илюхин Ю.В., Грибков А.А. Программирование

- доверенной память-центрической системы управления движением робототехнических и мехатронных систем // Вестник Московского авиационного института. 2022. Т. 29. №4. С. 197-210
7. Juurlink B., Meenderinck C. Amdahl's law for predicting the future of multicores considered harmful // ACM SIGARCH Computer Architecture News, 40 (2012), 2. pp.1-9
8. Кнут Д.Э. Искусство программирования, том 3. Сортировка и поиск. М.: ООО "И.Д. Вильямс", 2018. 832 с.
9. Time Complexities of all Sorting Algorithms. Geeks for Geeks, 2023. URL: <https://www.geeksforgeeks.org/time-complexities-of-all-sorting-algorithms/>
10. Heap Algorithms. Massachusetts Institute of Technology, 2010. URL: <https://courses.csail.mit.edu/6.006/fall10/handouts/recitation10-8.pdf>
11. Binary Insertion Sort. Geeks for Geeks, 2023. URL: <https://www.geeksforgeeks.org/binary-insertion-sort/>

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Статья посвящена проблематике управления движением в реальном времени для промышленных роботов и другого технологического оборудования. Авторы фокусируются на разработке и оптимизации память-ориентированной архитектуры системы управления, обеспечивающей высокое быстродействие и точность управления. В статье предложена методология, направленная на разработку и оптимизацию системы управления движением промышленных роботов с использованием память-ориентированной архитектуры. Центральная идея этой методологии заключается в том, что данные в процессе вычислений не перемещаются между процессором и памятью, а остаются в памяти, в которую интегрируется процессор. Это позволяет уменьшить задержки при передаче данных и повысить общую производительность системы за счет использования технологий обработки в памяти или обработки вблизи памяти.

Одним из ключевых аспектов методологии является акторная инструментальная модель, где каждый элемент системы управления движением представлен актором, соответствующим функциональному модулю системы. Акторы генерируются по необходимости в процессе управления в виде отдельных программ, запускаемых в общей памяти или в локальной памяти отдельного функционального модуля. Для программной реализации системы управления движением используется метапрограммирование, что позволяет генерировать и исполнять программы на основе текущих потребностей системы.

Конфигурирование цикла управления заключается в определении взаимного распределения затрат времени на различные операции управления по потокам исполнения согласно заданным ограничениям по их последовательности и доступным вычислительным ресурсам системы. Предложены три подхода к конфигурированию: без оптимизации, с комбинаторной оптимизацией по времени и с комбинаторной оптимизацией по ресурсам. В первом подходе воспроизводится заданная конфигурация цикла управления с известным распределением элементов по группам и потокам исполнения. Во втором подходе формируются циклы управления из множества элементов с целью минимизации длительности цикла управления. В третьем подходе распределение элементов по потокам исполнения осуществляется таким образом, чтобы минимизировать потребные ресурсы при соблюдении заданной длительности цикла управления.

Для решения задач комбинаторной оптимизации при большом числе элементов в группе применяются стохастические методы, которые обеспечивают снижение сложности алгоритмов и улучшают их практическую реализуемость. Также рассматривается адаптивное конфигурирование цикла управления, при котором задействуются максимально доступные ресурсы в каждый момент времени для минимизации времени исполнения всех элементов группы. Адаптивное конфигурирование может использоваться как в реальном времени, так и для многократного адаптивного конфигурирования с целью определения квазиоптимальной конфигурации цикла управления.

Актуальность исследования обусловлена необходимостью повышения быстродействия систем управления движением сложных объектов, таких как промышленные роботы и станки с ЧПУ. В условиях современной промышленности, где требуется высокая точность и скорость обработки данных, предложенные решения могут значительно улучшить производительность и надежность таких систем.

Научная новизна работы заключается в предложении памяти-ориентированной архитектуры системы управления движением, которая отличается от традиционной процессорно-ориентированной. Использование акторной модели и метапрограммирования для реализации этой архитектуры представляет собой инновационный подход к решению задач управления в реальном времени.

Статья написана в научном стиле, характеризующемся высокой точностью и логичностью изложения. Структура работы включает введение, обзор существующих проблем, описание предложенных методологий и подходов, экспериментальные результаты и выводы. Содержание статьи достаточно полное и охватывает все аспекты заявленной темы.

Выводы статьи подтверждают эффективность предложенных подходов к конфигурированию цикла управления. Авторы демонстрируют, что использование памяти-ориентированной архитектуры позволяет значительно сократить время выполнения операций управления и повысить общую производительность системы.

Статья будет интересна специалистам в области управления движением промышленных роботов, а также исследователям, занимающимся разработкой высокопроизводительных систем управления. Предложенные методы и результаты могут быть полезны для дальнейших исследований и практических разработок в этой области.

Рекомендации по доработке:

1. Предоставить более подробные результаты экспериментальных исследований, включая сравнение с альтернативными подходами и количественные показатели улучшений.
2. Углубленно описать алгоритмы, использованные для оптимизации конфигурирования цикла управления, и привести примеры их практической реализации.
3. Рассмотреть возможные ограничения предложенной архитектуры и условия, при которых её применение может быть неэффективным.

Заключение

Статья представляет собой значимый вклад в область управления движением промышленных роботов. Предложенные методологии и результаты исследований имеют высокую научную и практическую ценность. С учетом предложенных доработок, статья может быть рекомендована к публикации.

Рекомендуемое решение по статье: направить статью на доработку с последующей повторной рецензией.

Программные системы и вычислительные методы*Правильная ссылка на статью:*

Трофимова В.С., Каршиева П.К., Рахманенко И.А. Метод трансферного обучения для дообучения нейронных сетей под особенности набора данных в задаче верификации диктора // Программные системы и вычислительные методы. 2024. № 3. DOI: 10.7256/2454-0714.2024.3.71630 EDN: XHZCTS URL: https://nbpublish.com/library_read_article.php?id=71630

Метод трансферного обучения для дообучения нейронных сетей под особенности набора данных в задаче верификации диктора**Трофимова Варвара Сергеевна**

ORCID: 0009-0008-5044-2321



студент; кафедра комплексной информационной безопасности электронно-вычислительных систем;
Томский государственный университет систем управления и радиоэлектроники

634045, Россия, Томская область, г. Томск, ул. Красноармейская, 146, ауд. 509

✉ varvara.trofimova.01@mail.ru

Каршиева Полина Константиновна

ORCID: 0009-0004-8390-2348



студент; кафедра комплексной информационной безопасности электронно-вычислительных систем;
Томский государственный университет систем управления и радиоэлектроники

634045, Россия, Томская область, г. Томск, ул. Красноармейская, 146, ауд. 509

✉ polinakarshieva1@gmail.com

Рахманенко Иван Андреевич

ORCID: 0000-0002-8799-601X

кандидат технических наук



доцент; кафедра Безопасности информационных систем; Томский государственный университет
систем управления и радиоэлектроники

634045, Россия, Томская область, г. Томск, ул. Красноармейская, 146, ауд. 509

✉ ria@fb.tusur.ru

[Статья из рубрики "Методы, языки и модели человека-машинного взаимодействия"](#)

DOI:

10.7256/2454-0714.2024.3.71630

EDN:

XHZCTS

Дата направления статьи в редакцию:

30-08-2024

Дата публикации:

06-09-2024

Аннотация: Предметом исследования являются нейронные сети, дообученные с использованием метода трансферного обучения под особенности набора данных. Объектом исследования являются методы машинного обучения, используемые для решения задач верификации по голосу. Цель исследования заключается в повышении эффективности нейронных сетей в задаче голосовой верификации. В данной работе для процесса дообучения были подготовлены три набора данных на разных языках: на английском, на русском и на китайском. Также с использованием современных и предварительно обученных моделей ResNetSE34L и ResNetSE34V2 было проведено экспериментальное исследование, направленное на повышение эффективности нейронных сетей в задаче верификации диктора по произвольной фразе. Особое внимание уделялось оптимизации параметров обучения нейронных сетей с помощью метода тонкой настройки. Методология проведения исследования включает в себя оценку эффективности дообучения нейронных сетей под особенности набора данных в задаче верификации диктора, опираясь на значение равной ошибки 1-ого и 2-ого рода. Также был проведен ряд экспериментов, в ходе которых варьировались параметры и применялась техника замораживания слоев. Максимальное понижение равной ошибки 1-ого и 2-ого рода при использовании английского набора данных было достигнуто при помощи подбора эпох и скорости обучения, ошибка уменьшилась на 50%. Аналогичный подбор параметров при использовании русского набора данных уменьшил ошибку на 63,64%. При дообучении с использованием китайского набора данных наименьшее значение было достигнуто в эксперименте с замораживанием полностью связанного слоя, изменении скорости обучения и оптимизатора – ошибка снизилась на 16,04%. Полученные результаты могут использоваться при проектировании и разработке систем голосовой верификации и в образовательных целях. Также был сделан вывод, что метод трансферного обучения является эффективным при дообучении нейронных сетей под особенности набора данных, так как в подавляющем количестве экспериментов было достигнуто понижение EER, что говорит о повышении точности распознавания диктора.

Ключевые слова:

трансферное обучение, тонкая настройка, набор данных, верификация диктора, распознавание диктора, извлечение признаков, обработка речи, нейронные сети, глубокое обучение, распознавание образов

1 Введение

В настоящее время широко используются технологии, позволяющие осуществлять верификацию пользователя по различным биометрическим параметрам. Одним из таких параметров является голос [1]. Верификация диктора — это форма распознавания диктора, при которой принимается решение о принадлежности голосового образца индивиду, чья личность была заявлена [2]. Данная процедура позволяет обеспечить высокий уровень безопасности и удобства при получении пользователям доступа к

информации. Использование голосовой биометрии становится всё более актуальным в различных сферах, включая банковское дело, учреждения здравоохранения, системы безопасности и даже повседневную жизнь [3].

Дообучение нейронных сетей под особенности набора данных играет ключевую роль для повышения точности работы систем верификации диктора по голосу [4]. Каждый набор данных имеет уникальные признаки. При дообучении используются предварительно обученные модели, которые в процессе адаптируются под новые признаки, что приводит к значительному улучшению эффективности работы системы.

Эффективность нейронных сетей, выполняющих верификацию диктора по голосу, напрямую влияет на безопасность и удобство использования различных сервисов. Точная система верификации диктора по голосу позволяет предотвращать несанкционированный доступ к конфиденциальной информации и защищать пользовательские аккаунты от злоумышленников.

2 Проведение эксперимента

2.1 Предварительные данные

Для дообучения нейронной сети были выбраны предварительно обученные модели на английском языке, такие как ResNetSE34L и ResNetSE34V2. Эти модели, основанные на архитектуре ResNet34, используются для идентификации диктора по произвольной фразе и последующей верификации. Основное отличие между ResNetSE34L и ResNetSE34V2 заключается в методах извлечения и адаптации признаков аудиоданных. Кроме того, ResNetSE34V2 предлагает более эффективную обработку данных, что обеспечивает более высокой уровень точности обучения.

В качестве обучающего набора данных для модели ResNetSE34L был использован речевой корпус VoxCeleb2 [5]. С помощью данного корпуса были получены веса «baseline_lite_ap.model». Этот акустико-фонетический корпус речи представляет собой один из крупнейших наборов данных, применяемых для оценки систем автоматического распознавания речи.

Также для модели ResNetSE34L использовался речевой корпус TIMIT [6], на котором была обучена модель и получены веса «model000000100.model».

Далее для последующего дообучения моделей были подготовлены три речевых корпуса: на английском, русском и китайском языке.

В качестве первого был использован речевой корпус TIMIT [6], содержащий аудиозаписи на английском языке. Он был разделен на обучающий набор данных, содержащий 4620 аудиозаписей, и тестовый набор данных, состоящий из 1680 аудиозаписей. Этот набор данных был задействован в процессе дообучения с использованием обученных весов «baseline_lite_ap.model» и «baseline_v2_ap.model». В случае с «model000000100.model» этот подход не применялся, так как соответствующие веса были получены в ходе обучения модели ResNetSE34L на основе речевого корпуса TIMIT.

Речевой корпус на русском языке включает в себя аудиозаписи, сделанные 50 носителями русского языка. Каждый диктор имеет 50 аудиозаписей. Этот набор данных был разделен на две группы: обучающий набор, содержащий записи 30 дикторов, и тестовый набор, состоящий из записей 20 других дикторов [7].

В данном исследовании также использовался речевой корпус на китайском языке под названием «HI-MIA». Данные для корпуса собирались в домашних условиях с помощью микрофонных решеток и микрофона HI-FI. Датасет был разделен на два набора: обучающий, включающий записи 42 дикторов, и тестовый, состоящий из записей 40 дикторов [8].

Речевые корпусы на русском и китайском языке использовались в экспериментах со всеми представленными обученными весами, поскольку ни один из них ранее не использовался для обучения моделей.

Эти наборы данных были выбраны для дообучения нейронной сети с целью повышения ее эффективности в распознавании речи. Благодаря разнообразию диалектов, голосов и языков в аудиозаписях, можно получить более широкий охват вариативности в произношении слов и фраз.

Для самой верификации пользователя используется тестовый сценарий, включающий попытки аутентификации легального пользователя и атакующего (в случае несовпадения голоса). Легальная попытка в текстовом файле с тестовым сценарием помечена меткой 1, а атака на систему — меткой 0.

В контексте стремления улучшения эффективности нейронной сети в задачах голосовой верификации основное внимание уделялось понижению равной ошибки 1-го и 2-го рода [9]. Цель заключалась в достижении более точного распознавания диктора, что важно для обеспечения высокой надежности системы.

В рамках проведенного исследования для качественного уменьшения ошибок 1-ого и 2-ого рода был применен метод трансферного обучения с использованием предварительно обученных весов модели, выполняющей задачу верификации диктора по произвольной фразе.

Используемый метод представляет собой технику дообучения нейронной сети, когда предварительно обученная модель на большом наборе данных дополнительно обучается на более узком наборе данных для выполнения задачи верификации диктора

2.2 Параметры дообучения

В процессе дообучения нейронной сети использовались различные стратегии для адаптации модели к новым данным. Особое внимание уделялось выбору параметров обучения, включая скорость обучения, изменение оптимизатора, количество эпох обучения и структура нейронной сети.

Помимо описанных выше параметров, также учитывался размер батча. При дообучении модели на речевом корпусе TIMIT размер батча был 50. При работе с русским корпусом он составлял размер 30, а с китайским размер был равен 40. Описание параметров, используемых для дообучения нейронной сети:

- 1) Скорость обучения влияет на сходимость модели и предотвращает переобучение. Изменение весов должно быть умеренным, чтобы избежать нарушения выученных признаков.
- 2) Количество эпох в процессе дообучения влияет на переобучение модели. Также недостаточное количество эпох в процессе дообучения может привести к недостаточной адаптации.

- 3) Маленькие размеры батча в процессе дообучения позволяют найти баланс между вычислительной эффективностью, устойчивостью к изменениям и способностью модели к обобщению на новые данные.
- 4) Оптимизатор отвечает за настройку весов модели в процессе обучения, чтобы минимизировать функцию потерь.
- 5) Структура нейронной сети описывает архитектуру модели, включая количество слоев, а также их взаимосвязь. Процесс дообучения определяет, как новые данные интегрируются в существующую архитектуру. Изменение структуры сети может включать в себя замораживание слоев. Это позволяет модели изучить новые особенности данных и сохранять ее первоначальное обучение.

В процессе исследования аудиозаписи были разделены на два текстовых файла: «train» и «test». Это позволило систематизировать данные и обеспечить необходимую структуру для последующего анализа.

3 Результаты

3.1 Результат дообучения с использованием речевого корпуса TIMIT

В целях определения наилучшей стратегии дообучения была проведена серия экспериментов с использованием набора данных TIMIT для обученных весов «baseline_lite_ap.model» и «baseline_v2_ap.model». Чтобы определить, насколько эффективнее нейронная сеть распознает дикторов, было рассмотрено значение EER (Equal Error Rate), полученное в ходе тестирования и последующего дообучения модели [\[9\]](#).

Чтобы аналитически сравнить значения, полученные после дообучения, с изначальными данными, модель была оценена без применения обучающего процесса. Было получено значение EER равное 0.012 и 0.013 соответственно для «baseline_lite_ap.model» и «baseline_v2_ap.model».

Далее модель была подвергнута дообучению с внесением изменений в её параметры. По завершению двадцати экспериментов была составлена сводная таблица, в которой представлены результаты равной ошибки 1-ого и 2-ого рода для сравнения эффективности дообучения модели. В данной таблице указаны используемые модели, предварительно обученные веса и проделанные эксперименты. Каждый эксперимент включал в себя попытку дообучения моделей на речевом корпусе TIMIT [\[6\]](#).

Таблица 1 — Результаты экспериментов для набора TIMIT

№	Эксперимент	Модели и веса		
		model000000100.model	baseline_lit_ap	baseline_v2_ap
0	До дообучения	0,066	0,012	0,013
1	Без изменений параметров	*	0,014	0,017
2	Замораживание 1 и 2 сверточного слоя	*	0,015	0,011
3	Замораживание 1 сверточного слоя	*	0,012	0,011
4	Замораживание 2 сверточного слоя	*	0,011	0,008
	-			

5	Замораживание полносвязного слоя, скорость обучения и изменение оптимизатора	*	0,011	0,012
6	Замораживание полносвязного слоя	*	0,009	0,015
7	Изменение оптимизатора	*	0,008	0,007
8	Подбор эпох и скорости обучения (max)	*	0,012	0,007
9	Подбор эпох и скорости обучения (min)	*	0,006	0,010
10	Понижение lr_decay и lr	*	0,008	0,008

При обучении с применением весов модели «baseline_lite_ap.model» наилучший результат был получен при незначительном изменении скорости обучения. Исходное значение EER было равно 0,012, после дообучения оно снизилось до 0,006. Данное изменение свидетельствует об уменьшении EER на 50%.

При использовании предварительно обученных весов «baseline_v2_ap.model» лучший результат был достигнут при изменении оптимизатора и существенного снижения скорости обучения. Перед дообучением значение равной ошибки 1-ого и 2-ого рода составляло 0,013, после дообучения показатель сократился до 0,007, что свидетельствует об уменьшении на 46,15%.

Помимо проведения процесса дообучения все модели также были подвергнуты обучению на наборе данных TIMIT^[6]. Полученные результаты были сопоставлены с EER^[9], полученными в контексте дообучения, с целью демонстрации уменьшения EER при процессе дообучения по сравнению с исключительно базовым обучением.

3.2 Результат дообучения с использованием Русского речевого корпуса

Для оптимизации работы модели и повышения эффективности распознавания дикторов была проведена серия экспериментов на основе набора данных на русском языке с использованием предварительно обученных весов «model000000100.model», «baseline_lite_ap.model» и «baseline_v2_ap.model».

План постановки экспериментов, проводимых с русским речевым корпусом^[8], не отличался от экспериментального исследования с речевым корпусом TIMIT^[7]. Чтобы оценить эффективность распознавания диктора до процесса дообучения, модель была подвергнута оцениванию без обучения. Это было предпринято для оценки эффективности модели перед внесением корректив в её параметры и реализации процесса дообучения.

В связи с тем, что изначально модель обучалась на английском языке, а последующее предварительное обучение проводилось на русском, процесс смены языковой среды оказал влияние на точность модели. Точность модели в течение десяти эпох не поднималась выше шестидесяти. Этот переход требовал адаптации модели к новому языку и его особенностям. Английский и русский язык имеют множество различий в грамматике и синтаксисе. По мимо вышеперечисленного повлиять на дообучение может также морфология каждого языка. Подобная процедура проводится для того, чтобы расширить функциональность модели для работы с различными языками и культурными контекстами.

В результате проведения описанных выше экспериментов с использованием набора данных на русском языке была сформирована таблица 2. В данной таблице представлены результаты равных ошибок 1-ого и 2-ого рода, полученные в ходе всех экспериментов, в которых использовались предварительно обученные веса моделей «baseline_v2_ap.model», «baseline_lite_ap.model» и «model000000100.model».

Таблица 2 — Результаты экспериментов для Русского речевого корпуса

№	Эксперимент	Модели и веса		
		model000000100.model	baseline_lite_ap	baseline_v2_ap
0	До дообучения	0,066	0,012	0,078
1	Без изменений параметров	0,055	0,014	0,114
2	Замораживание 1 и 2 сверточного слоя	0,035	0,013	0,066
3	Замораживание 1 сверточного слоя	0,053	0,014	0,064
4	Замораживание 2 сверточного слоя	0,053	0,010	0,067
5	Замораживание полносвязного слоя, скорость обучения и изменение оптимизатора	0,032	0,007	0,078
6	Замораживание полносвязного слоя	0,048	0,010	0,062
7	Изменение оптимизатора	0,049	0,011	0,067
8	Подбор эпох и скорости обучения (max)	0,033	0,008	0,073
9	Подбор эпох и скорости обучения (min)	0,024	0,008	0,064
10	Понижение lr_decay и lr	0,028	0,009	0,068

В процессе обучения с применением весов «model000000100.model» значительное улучшение было получено при незначительной модификации скорости обучения. Начальное значение EER составляло 0,066, после дообучения показатель сократился до 0,024, что свидетельствует об уменьшении на 63,64%.

Путем использования предварительно обученных весов «baseline_lite_ap.model» наилучший результат был достигнут при замораживании полносвязного слоя, изменении скорости обучения и оптимизатора. Исходное значение EER было равно 0,012, после дообучения оно снизилось до 0,007, что означает сокращение EER на 41,67%.

При использовании предварительно обученных весов «baseline_v2_ap.model» самый качественный процесс обучения был реализован с помощью замораживания полносвязного слоя. Перед дообучением значение EER составляло 0,078, после дообучения показатель сократился до 0,062, что свидетельствует об уменьшении на 20,51%.

3.3 Результат дообучения с использованием речевого корпуса HI-MIA

В данном исследовании предварительно обученные веса были также дообучены на речевом корпусе, содержащим аудиозаписи на китайском языке [8].

Работа с этим набором данных не привела к успешному дообучению модели в большинстве случаев. Были проведены эксперименты, ранее проведенные с другими наборами данных. Также были добавлены эксперименты с другими значениями скорости обучения и замораживанием других слоев. В результате только при работе с предварительно обученными весами «baseline_lite_ap.model» в эксперименте с замораживанием полносвязного слоя, изменением скорости обучения и оптимизатора было достигнуто успешное дообучение.

При дообучении, где использовались веса «model000000100.model», исходная EER была равна 0,066 [9]. Новый набор данных содержал разнообразные фон-шумы, что представляет собой типичную среду взаимодействия в реальном мире. Параметры модели были адаптированы для качественного дообучения. После дообучения на наборе данных, содержащим аудиозаписи на китайском языке, было замечено увеличение EER до 0.214.

Все результаты, проведённых экспериментов с набором данных на китайском языке, представлены в таблице 3.

Таблица 3 — Результаты экспериментов для HI-MIA

№	Эксперимент	Модели и веса		
		model000000100.model	baseline_lit_ap	baseline_v2_ap
0	До дообучения	0,180	0,106	0,089
1	Без изменений параметров	0,214	0,108	0,091
2	Замораживание 1 и 2 сверточного слоя	0,197	0,125	0,094
3	Замораживание 1 сверточного слоя	0,201	0,138	0,096
4	Замораживание 2 сверточного слоя	0,207	0,141	0,097
5	Замораживание полносвязного слоя, скорость обучения и изменение оптимизатора	0,259	0,089	0,121
6	Замораживание полносвязного слоя	0,210	0,130	0,120
7	Изменение оптимизатора	0,231	0,147	0,092
8	Подбор эпох и скорости обучения (max)	0,302	0,222	0,148
9	Подбор эпох и скорости обучения (min)	0,229	0,134	0,123
10	Понижение lr_decay и lr	0,223	0,135	0,123

Анализ показал, что причиной повышения EER могут быть как лингвистические и фонетические различия между английским и китайским языком, так и присутствие шума, маскирующего ключевые акустические особенности речи. Дообучение на зашумленном китайском наборе данных привело к переобучению на специфических особенностях этого набора данных, что снизило обобщаемость модели. Также возможно, что стандартные методы дообучения не справляются со своей задачей в условиях сильной языковой вариативности и присутствия шума.

4 Область применения результатов и новизна

Данное исследование позволяет избежать дополнительных затрат на обучение моделей с нуля и, при использовании данных моделей, повышает уровень эффективности работы систем верификации диктора по голосу. Технология верификации диктора по голосу широко распространена в банковской сфере, так как при обращении клиента в контакт-центр единственным доступным биометрическим параметром является голос.

В работе были использованы известные предварительно обученные модели и метод трансферного дообучения в рамках заданной предметной области. Основным научным результатом являются результаты оценки эффективности и применимости конкретных методов дообучения нейронных сетей под особенности наборов данных в задаче текстонезависимой верификации диктора.

5 Выводы

В работе было проведено исследование процесса трансферного обучения нейронных сетей под особенности набора данных, что позволило получить более точные модели после проведенного дообучения.

Полученные наблюдения подчеркивают важность корректной настройки параметров при дообучении моделей. Тонкая настройка, проводимая в ходе дообучения моделей, благоприятно сказывается на оптимизации их эффективности.

При использовании предварительно обученных весов «baseline_lite_ap.model» и речевого корпуса TIMIT было получено максимальное понижение равной ошибки 1-ого и 2-ого рода, она уменьшилась на 50%.

С набором данных на русском языке заметно снизился показатель EER при использовании предварительно обученных весов «model000000100.model». Снижение составило 63.64%.

В работе с набором данных HI-MIA успешное дообучение модели было достигнуто только в эксперименте с замораживанием полно связного слоя, изменением скорости обучения и оптимизатора. По итогу было зафиксировано, что равная ошибка 1-ого и 2-ого рода (EER) снизилась на 16,04%.

Библиография

1. Гассиев Д. О., Сахаров В. А., Ермолаева В. В. Голосовая аутентификация // Тенденции развития науки и образования. 2019. № 56(2). С. 22-24.
2. ГОСТ Р 58668.11-2019 (ИСО/МЭК 19794-13:2018). Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 11. Данные голоса. Москва: Стандартинформ, 2019. 28 с.
3. Девятков В.В. Системы искусственного интеллекта. Гл. ред. И.Б. Федоров. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2001. 352 с.
4. Галушкин А.И. Нейронные сети. Основы теории. – М.: Горячая линия – Телеком, 2012. 496 с.
5. Suzuki K. Artificial Neural Networks: Architectures and Applications. – Publisher: InTech, 2013. 256 p.
6. Евсюков М.В., Путято М.М., Макарян А.С., Немчинова В.О. Методы защиты в современных системах голосовой аутентификации // Прикаспийский журнал: управление и высокие технологии. 2020. № 3(59). С 84-92.
7. Nagrani A., Chung J.S., Zisserman A. VoxCeleb: A large-scale speaker identification

- dataset // arXiv:1706.08612. 2018. URL: <https://arxiv.org/pdf/1706.08612.pdf> (дата обращения: 15.07.2024).
8. Hinton G. E., Srivastava X., Krizhevsky A., Sutskever I., Salakhutdinov R. R. Improving neural networks by preventing co-adaptation of feature detectors // arXiv:1207.0580. 2012. URL: <https://arxiv.org/pdf/1207.0580.pdf> (дата обращения: 22.07.2024).
9. Конев А.А. Модель и алгоритмы анализа и сегментации речевого сигнала: автореф. дис. на соиск. ученой степ. канд. техн. наук. Томск, 2007. 150 с.
10. Qin X., Bu H., Li M. HI-MIA: A Far-field Text-Dependent Speaker Verification Database and the Baselines // IEEE International Conference on Acoustics, Speech, and Signal Processing, 2020. Рр. 7609-7613.
11. Рахманенко И.А., Шелупанов А.А., Костюченко Е.Ю. Автоматическая верификация диктора по произвольной фразе с применением свёрточных глубоких сетей доверия // Компьютерная оптика. 2020. № 44(4). С. 596-605.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Статья посвящена актуальной и востребованной теме – разработке методов улучшения верификации диктора на основе биометрии голоса с использованием нейронных сетей. В частности, исследуется трансферное обучение нейронных сетей, где предварительно обученные модели адаптируются под специфические особенности различных речевых данных. Предложенный метод решает задачу повышения точности и надежности систем голосовой верификации, что имеет большое значение для таких сфер, как банковское дело, системы безопасности, медицинские учреждения и повседневные цифровые сервисы.

Методология статьи демонстрирует высокий научный уровень. Авторы используют несколько известных нейронных сетей, таких как ResNetSE34L и ResNetSE34V2, которые дообучаются на различных наборах данных (английский, русский и китайский речевые корпусы). Приводятся конкретные параметры обучения, такие как скорость обучения, размер батча и количество эпох, что позволяет в полной мере оценить проведенные эксперименты и их результаты. Использование трансферного обучения является одним из ключевых аспектов в современных исследованиях искусственного интеллекта, что свидетельствует о высокой актуальности выбранного подхода.

Тема статьи находится на острие современных технологий. Применение биометрии голоса для идентификации личности набирает популярность, и повышение точности таких систем становится особенно важным в условиях роста числа кибератак и увеличения требований к безопасности. Кроме того, использование трансферного обучения для дообучения моделей на новых наборах данных позволяет существенно сократить затраты на вычислительные ресурсы и повысить производительность системы. Статья предлагает оригинальный подход к адаптации предварительно обученных моделей под особенности конкретных речевых данных, что позволяет повысить точность систем верификации диктора. Авторы исследуют влияние различных параметров обучения на эффективность моделей и предоставляют детализированные результаты, которые показывают значительное снижение ошибки (EER). Особенно интересным является применение метода на разных языковых данных, что расширяет область применения предложенной технологии.

Стиль изложения материала является логичным и последовательно выстроенным. Статья структурирована классическим образом, начиная с введения в тему, обзора

используемых моделей и заканчивая подробным анализом результатов. Особенno стоит отметить ясность и четкость представления экспериментов. Каждый этап исследования подробно описан, что позволяет легко следовать ходу мысли авторов и оценить значимость полученных данных.

Выводы статьи подкреплены результатами проведенных экспериментов и представляют собой логическое завершение исследования. Авторы подчеркивают, что использование трансферного обучения с предварительно обученными весами позволяет значительно снизить процент ошибок 1-го и 2-го рода (EER), что улучшает общее качество работы системы. Выводы полностью соответствуют заявленным целям исследования и подтверждают эффективность предложенного подхода.

Статья будет интересна как специалистам в области искусственного интеллекта и машинного обучения, так и исследователям, работающим над улучшением биометрических систем безопасности. Кроме того, результаты работы могут быть полезны практикам, занимающимся разработкой коммерческих решений в области верификации пользователей на основе голоса.

Статья представляет собой значимый вклад в исследование методов улучшения голосовой верификации с использованием нейронных сетей. Методология, представленная авторами, демонстрирует научную новизну и оригинальность подхода, а результаты экспериментов подтверждают эффективность предложенного метода. Рекомендую статью к публикации без существенных замечаний.

Программные системы и вычислительные методы*Правильная ссылка на статью:*

Лукичев Р.В. Эволюция технологий семантического веба: проблемы и перспективы // Программные системы и вычислительные методы. 2024. № 3. DOI: 10.7256/2454-0714.2024.3.71719 EDN: JJXDYW URL:
https://nbpublish.com/library_read_article.php?id=71719

Эволюция технологий семантического веба: проблемы и перспективы**Лукичев Руслан Владимирович**

ORCID: 0000-0002-2293-2410

кандидат философских наук

магистр; факультет программной инженерии и компьютерной техники; Университет ИТМО
Директор по развитию; ООО "Первая Фестивальная Компания"

198332, Россия, г. Санкт-Петербург, Ленинский пр-т, 100к3, кв. 113

[✉ ruslanlukichev@gmail.com](mailto:ruslanlukichev@gmail.com)[Статья из рубрики "Базы данных"](#)**DOI:**

10.7256/2454-0714.2024.3.71719

EDN:

JJXDYW

Дата направления статьи в редакцию:

16-09-2024

Аннотация: Статья посвящена рассмотрению ключевых технологий семантического веба, анализу их особенностей, проблемных сторон и точек роста, что представляется особенно актуальным в контексте вопросов импортозамещения и повышения информационной безопасности нашей страны. Особое внимание уделено RDF-графам, которые основаны на онтологово-ориентированном подходе, а также языку OWL как основному инструменту организации машиночитаемых структур данных со сложными взаимосвязями между сущностями, иерархией классов и свойств. В отдельном блоке проанализированы ограничения, связанные с вопросами безопасности семантических баз данных, необходимости их упрощения, стандартизации и развития профильного программного обеспечения, отвечающего критериям юзабилити. Кроме того, обозначены перспективы дальнейшего совершенствования указанных технологий в контексте Интернета вещей и искусственного интеллекта. В исследовании применена комплексная методологическая база, которая подразумевает использование преимущественно

общенаучных методов, в частности, системного и аналитического: в статье обобщены и проанализированы актуальные наработки, связанные с технологиями семантического веба, что позволило выявить целый ряд проблем, нуждающихся в решении. В частности, имеющиеся на сегодняшний день инструменты зачастую имеют высокий порог входа, отличаются избыточно сложным, невыразительным интерфейсом без функций дополняющих подсказок и визуализации запросов. Кроме того, языки семантического веба нуждаются в стандартизации и введении общего протокола с целью упрощения процесса работы с разноформатными данными, агрегированными из разных источников. Другими важными вопросами являются обеспечение достоверности и актуальности информации, ее целостности и конфиденциальности, а также контекстуальной обусловленности логических выводов и соответствия запросам пользователя. В числе ключевых перспектив названо создание интеллектуальной автономной среды, устройства в которой смогут свободно обмениваться данными и взаимодействовать друг с другом на семантическом уровне с целью оказания высококачественных персонализированных услуг. Положения статьи могут быть взяты за основу при разработке отечественных систем структуризации и описания данных, доступных для машинной обработки, а также профильных лекционных курсов в высших учебных заведениях.

Ключевые слова:

семантический веб, онтологии, графовые базы данных, модели данных, семантическая сеть вещей, RDF, RDFS, OWL, SPARQL, XML

Актуальность темы и методология исследования

Семантический веб – это концепция, предложенная ровно четверть века назад сэром Тимом Бернерсом Ли, создателем Всемирной паутины. Еще в преддверии эпохи Web 2.0 он пророчески заявлял о необходимости научить компьютеры самостоятельно описывать те или иные явления, затем делать выводы и, наконец, аргументированно рассуждать [\[1, с. 184\]](#). На страницах книги «Плетя паутину: истоки и будущее Всемирной паутины» основатель и бессменный глава Консорциума W3C рассуждал о необходимости становления семантического веба как информационной сети, данные в которой были бы представлены в машиночитаемом формате или конвертированы в таковой, – с целью их прямой или косвенной обработки компьютерами, постоянно взаимодействующими между собой [\[1, с. 177\]](#).

Семантический веб все еще находится в процессе становления и включает в себя целый ряд технологий, в том числе RDF, OWL, SPARQL, обзору и анализу которых посвящена настоящая статья. Поставленная исследовательская задача представляется весьма актуальной.

Во-первых, более глубокое понимание природы подобных семантических инструментов позволяет совершенствовать механизмы обработки больших данных в самых разных предметных областях, включая медицину, экономику, военно-промышленный комплекс, науку, образование и многое другое, а также улучшать процесс подбора и интерпретации запрашиваемой пользователем информации в поисковых системах. Все это способствует высвобождению человеческих и временных ресурсов, отводимых на работу с данными, и автоматизирует обмен информацией между различными устройствами, тем самым претворяя в жизнь концепцию Интернета вещей.

Во-вторых, большинство статей отечественных авторов за последние несколько лет, посвященных семантическому вебу, раскрывают лишь некоторые его аспекты, не затрагивая проблемы и перспективы их развития в комплексе, либо предлагают изучение вопросов использования подобных технологий в том или ином отдельно взятом виде профессиональной деятельности. Например, от онлайн-образования [2], сферы финансов и инвестирования [3] до археологических изысканий [4] или геоаналитики улично-дорожных сетей [5]. Впрочем, широта и разнохарактерность применения семантического инструментария и связанного с ним онтологово-ориентированного подхода лишь подчеркивает необходимость его углубленного и всестороннего рассмотрения.

В-третьих, выбранная тема получает особое значение в контексте проблемы импортозамещения. Изучение технологий семантического веба поможет создать российские аналоги зарубежных веб-сервисов, обеспечивая независимость от иностранных компаний и повышая информационную безопасность нашей страны. Основные выводы и положения настоящего исследования могут быть взяты за теоретическую основу при разработке отечественных систем структуризации и описания данных, доступных для машинной обработки, а также профильных лекционных курсов в высших учебных заведениях.

Что касается методологии исследования, то в статье применена комплексная методологическая база, которая подразумевает использование преимущественно общенаучных методов, в частности, системного и аналитического, с целью систематизации и обобщения релевантных теоретических и научно-практических изысканий как отечественных, так и зарубежных авторов. Также в настоящей работе немаловажную роль играет индуктивный метод: поочередное рассмотрение ключевых особенностей ряда основополагающих семантических технологий способствует формированию более целостного взгляда на проблему их текущих ограничений и перспектив дальнейшего развития.

Обзор основных семантических технологий

Отправной точкой в развитии семантического веба можно по праву назвать создание языка XML (eXtensible Markup Language) в 1996 году – расширяемого языка разметки, предназначенного для создания, хранения и передачи структурированной информации в машиночитаемом формате. XML-документ предполагает обязательное наличие так называемого пролога, с указанием версии языка, и корневого элемента, который в свою очередь содержит произвольное количество вложенных элементов, состоящих из открывающих и закрывающих тегов с соответствующими им значениями и атрибутами, наименование которых определяет пользователь. При всех достоинствах указанного языка разметки, в том числе его независимости от платформы, расширяемости и ориентированности на быстрый и надежный обмен информацией между программами и устройствами, он имеет существенный недостаток, а именно способен описывать лишь структуру данных, а не заложенный в них смысл.

По этой причине основным компонентом семантического веба стала технология, получившая название RDF (Resource Description Framework). Она не зависит от предметной области и поддерживает графовую модель данных, предполагающую наличие триплета из субъекта (сущности или ресурса), объекта и связи между ними, называемой предикатом. Такого рода триплет позволяет машине формировать логические утверждения из предоставленной в ее распоряжение информации.

Если модель RDF описывает субъектно-объектные отношения, то взаимосвязи между сущностями оказались представлены в его расширении, – примитивном языке онтологий RDF Schema. Он обеспечивает возможность задать семантику конкретной области знаний, определив словарь терминов посредством организации иерархии классов и свойств.

Понятие онтологии – одно из ключевых в концепции семантического веба. Оно подразумевает иерархически структурированный, формализованный набор знаний о той или иной предметной области, включающий в себя описание характерных для неё классов, подклассов и их экземпляров, а также свойств и зависимостей между ними. Для этих целей рабочие группы W3C консорциума разработали отдельный, более выразительный и эффективный, чем RDFS, язык, – OWL (Ontology Web Language), обладающий широким набором инструментов для формирования сложных онтологий и основанный на декрипционных логиках. Это «семейство логик, созданное специально для представления терминологически знаний, которое имеет собственную богатую историю и особенности, хорошо известные в соответствующем профессиональном обществе» [\[6, с. 88\]](#).

На сегодняшний день остается актуальной вторая версия языка OWL, которая призвана расширить его первоначальный стандарт и предоставить дополнительные возможности для моделирования сложных отношений между данными, например, такие как эквивалентность, равенство и логические комбинации классов, их непересекаемость и непротиворечивость, специальные характеристики свойств и поддержка логического вывода. Кстати, существует несколько диалектов OWL, в том числе OWL2 DL (Description Logics), обеспечивающий вычислительную эффективность и результативную поддержку логического вывода на основе декрипционных логик, но при этом не имеющий полной совместимости с RDF; и OWL2 Full – напротив, с максимальной выразительной мощностью и абсолютной структурной и семантической совместимостью с RDF, однако без гарантий эффективной поддержки логического вывода [\[6, с. 94\]](#).

Также нельзя не упомянуть о еще одном важном инструменте: по аналогии с языком SQL в семантических базах данных существует свой язык запросов для эффективного взаимодействия с RDF-графами, – это SPARQL (SPARQL Protocol and RDF Query Language). Он имеет SQL-подобный синтаксис и поддерживает различные виды запросов, включая выборку данных в хранилищах триплетов, создание новых записей, их модификацию и удаление. SPARQL использует шаблоны для формирования запросов, а также позволяет их объединять и разделять, что способствует более гибкой настройке поиска и обработки данных и созданию сложных запросов.

Семантический веб: ограничения и точки роста

Рассмотренные выше технологии обеспечивают эффективную работу с данными в семантическом вебе, делая их более информативными, связанными и доступными для машинной обработки. Однако, несмотря на значительный прогресс в этой области, у них существует ряд проблемных моментов, требующих решения, равно как и многообещающие перспективы дальнейшего совершенствования и развития.

Одна из, возможно, не столь очевидных проблем, препятствующих широкому распространению семантических технологий, – это их сложность и высокий порог входа, а также недостаток интуитивно понятных инструментов для работы. Существующие стандарты семантической сети следует упростить, сделав их более доступными и привлекательными для начинающих разработчиков [\[7, с. 14\]](#). А интерфейсы должны быть

простыми и удобными в использовании как для специалиста, так и для рядового пользователя, и отвечать критериям так называемого юзабилити [8, с. 14587]. Некоторые отечественные авторы предлагают в этой связи собственные наработки – например, редактор запросов на языке SPARQL. Существующие решения не отличаются выразительностью интерфейса и не обладают функциями дополняющих подсказок и визуализации запросов [9, с. 87], что и было предложено специалистами из Пермского государственного национального исследовательского университета в собственной версии такого редактора.

Кроме того, применяемые в семантическом вебе языки имеют множество диалектов, что существенно затрудняет их изучение и стандартизацию. Очевидно, необходимо введение единого общего протокола, поскольку агрегация данных из различных источников остается сложной задачей из-за большого разнообразия их форматов и структур.

Еще одна проблема – достоверность и актуальность данных: быстро меняющаяся информация требует постоянного обновления и синхронизации с первоисточниками, что представляется весьма трудоемким процессом, производительность которого также нуждается в оптимизации. Не меньшую роль играет и релевантность агрегированной информации поставленной пользователем задаче, а также контекстуальная обусловленность логических выводов. Как отмечают зарубежные исследователи, стандарты семантического веба действительно облегчают обмен данными и их интеграцию, но полностью раскрывают свой потенциал лишь благодаря возможности адаптации данных под различные цели [10, с. 3389].

Наконец, одной из ключевых проблем для любого типа данных, включая семантический веб, является их безопасность, в том числе целостность и конфиденциальность. Например, остается до сих пор нерешенным вопрос предотвращения атак, основанных на инъекциях вредоносного кода в запрос на языке SPARQL, а имеющиеся предложения по защите семантических сетей не представляются эффективными [11, с. 41]. Возможное решение состоит в создании специальных криптографических средств, разработкой которых уже занимаются, в частности, представители Южного федерального университета.

Говоря о перспективах, отметим, что основными на сегодняшний день направлениями для дальнейшей прикладной интеграции семантического веба являются такие области, как большие данные, машинное обучение, искусственный интеллект и Интернет вещей. Концепции и технологии семантической сети активно интегрируются в сети связанных объектов для решения проблем совместимости, ограничивающих развитие Интернета вещей, что в конечном счете привело к появлению нового термина, известного как семантическая сеть вещей (Semantic Web of Things – SWoT) [12, с. 265]. Автоматическое извлечение и агрегирование информации и возможность создания интеллектуальных агентов, способных предоставлять персонализированные услуги, адаптированные к потребностям пользователя, а также интероперабельность с широким спектром устройств открывает путь к созданию интеллектуальной автономной среды, устройства и приборы в которой смогут свободно обмениваться данными и взаимодействовать друг с другом на семантическом уровне.

Заключение

Семантический веб – ключевой элемент Интернета будущего, где знания окажутся равнодоступными и для людей, и для машин – с учетом их контекста и смысловой

составляющей. Используемые для этого технологии, в том числе XML, RDF, RDFS, OWL, SPARQL, обеспечивают возможности персонализированного поиска, агрегирования сведений из разных источников, ускоренной и эффективной обработки больших объемов данных, обмена информацией между различными устройствами и находят применение в самых разных областях, таких как образование, наука, здравоохранение, транспорт и финансы. Решение существующих проблем семантического веба и реализация его перспективных направлений развития позволяют достичь нового уровня взаимодействия с данными, что открывает широкие возможности для последующих технологических инноваций.

Библиография

1. Berners-Lee, T. Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web. – NY: HarperCollins Publishers, 2000. – 246 pp.
2. Шполянская И. Ю., Середкина Т. А. Технологии Semantic Web в организации поддержки онлайн обучения. // Системный анализ в проектировании и управлении. – Вып. XXIV, № 3, 2020. – С. 343-350. doi:10.18720/SPBPU/2/id20-231
3. Иващук Ю. С., Орлянская Н. П., Тешев В. А. Разработка базы знаний инвестиционной деятельности на основе онтологического моделирования. // Вестник Адыгейского государственного университета. – Серия 5: Экономика, № 2 (320), 2023. – С. 90-98. doi:10.53598/2410-3683-2023-2-320-90-98
4. Петров И. Д., Терехова Ю. В. Расширение логической модели предметной области археологических исследований с использованием онтологий. // Успехи в химии и химической технологии. – Вып. 34, № 6 (229), 2020. – С. 133-135.
5. Смирнов А. В., Тесля Н. Н. Онтологово-ориентированная геоаналитика для определения мест концентрации ДТП на участках улично-дорожной сети. // Труды Кольского научного центра РАН. Серия: Технические науки. – Вып. 14, № 7, 2023. – С. 79-85. doi:10.37614/2949-1215.2023.14.7.008
6. Антониоу, Г. Семантический веб / Г.Антониоу, П. Грос, Ф. ван Хармелен, Р. Хоекстра; пер. с англ. Т. Шульга. – М.: ДМК Пресс, 2016. – 240 с.
7. Hogan, A. The Semantic Web: Two decades on. // Semantic Web. – Vol. 11, pp. 169-185, 2020. – doi:10.3233/SW-190387, 2020.
8. Hassan, B. Towards Semantic Web: Challenges and Needs. // International Journal Of Engineering And Computer Science. – Vol. 4, issue 10, Oct 2015. – pp. 14585-14588. doi:10.48550/arXiv.2105.02708
9. Турова И. А., Постаногов И. С. Разработка интеллектуального редактора SPARQL-запросов. // Вестник Новосибирского государственного университета. Серия: Информационные технологии. – Вып. 19, № 4, 2021. – С. 85-95.
10. Jat, A. Semantic web technologies: challenges and applications. // Journal of Critical Reviews. – Vol. 7, Issue 17, 2020. – pp. 3388-3390. doi:10.31838/jcr.07.17.417.
11. Чудинов П. Ю., Бабенко Л. К., Рогозов Ю. И. Анализ проблем защиты информации в семантических сетях. // Известия Южного федерального университета. Технические науки. – № 5 (229), 2022. – С. 37-47.
12. F. Z. Amara, M. Hemam, M. Djezzar and M. Maimor. Semantic Web and Internet of Things: Challenges, Applications and Perspectives. // Journal of ICT Standardization. – Vol. 10. No. 2, 2022. – pp. 261-291. doi: 10.13052/jicts2245-800X.1029

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Рецензируемая статья посвящена изучению эволюции технологий семантического веба, в ней рассматриваются проблемы и перспективы развития таких технологий.

Методология исследования базируется на использовании общенаучных методов, в частности, системного подхода и аналитического метода, которые применяются с целью систематизации и обобщения релевантных теоретических и научно-практических изысканий отечественных и зарубежных авторов, а также индуктивного метода для поочередного рассмотрения особенностей различных семантических технологий.

Актуальность работы авторы справедливо связывают с необходимостью познания природы применяемых семантических инструментов, комплексного представления проблем и перспектив развития технологий семантического веба и потребностью в создании российских аналогов зарубежных веб-сервисов.

Научная новизна рецензируемого исследования состоит в выявлении и систематизации проблемы и перспектив эволюции технологий семантического веба.

Структурно в статье выделены следующие разделы: Актуальность темы и методология исследования, Обзор основных семантических технологий, Семантический веб: ограничения и точки роста, Заключение и Библиография.

В публикации приведен обзор следующих технологий: XML (eXtensible Markup Language), RDF (Resource Description Framework), OWL (Ontology Web Language), OWL2 DL (Description Logics), OWL2 Full, SPARQL (SPARQL Protocol and RDF Query Language). Среди проблем, препятствующих широкому распространению семантических технологий, названы такие как: их сложность и высокий порог входа, нехватка интуитивно понятных инструментов для работы; наличие множества диалектов, затрудняющих их изучение и стандартизацию; требование постоянного обновления и синхронизации с первоисточниками в условиях быстро меняющейся информации. В Заключении авторы приходят к выводу о том, что технологии XML, RDF, RDFS, OWL, SPARQL обеспечивают возможности персонализированного поиска, агрегирования сведений из разных источников, ускоренной и эффективной обработки больших объемов данных, обмена информацией между различными устройствами и находят применение в самых разных областях, таких как образование, наука, здравоохранение, транспорт и финансы. Естественно, публикации не содержит решений всех проблем технологий семантического веба, но проведенная работа по их обобщению и систематизации представляется необходимой. Говоря о перспективах развития рассматриваемых технологий, авторы отмечают основные направления дальнейшей прикладной интеграции семантического веба: большие данные, машинное обучение, искусственный интеллект и Интернет вещей. Библиографический список включает 21 источник – публикации отечественных и зарубежных ученых на русском и английском языках по рассматриваемой теме, на которые в тексте имеются адресные ссылки, подтверждающие наличие апелляции к оппонентам.

Рецензируемый материал соответствует направлению журнала «Программные системы и вычислительные методы», отражает результаты проведенной авторами работы, содержит элементы научной новизны и практической значимости, может вызвать интерес у читателей, рекомендуется к опубликованию.

Программные системы и вычислительные методы*Правильная ссылка на статью:*

Тиханычев О.В. Об уточнении понятия «доверенности» систем искусственного интеллекта // Программные системы и вычислительные методы. 2024. № 3. DOI: 10.7256/2454-0714.2024.3.44097 EDN: JOPHLF URL: https://nbpublish.com/library_read_article.php?id=44097

Об уточнении понятия «доверенности» систем искусственного интеллекта

Тиханычев Олег Васильевич

ORCID: 0000-0003-4759-2931

кандидат технических наук

заместитель начальника отдела управления перспективных разработок, ГК "Техносерв"

111395, Россия, г. Москва, ул. Юности, 13

[✉ to.technoserv@gmail.com](mailto:to.technoserv@gmail.com)

Статья из рубрики "Базы знаний, интеллектуальные системы, экспертные системы, системы поддержки принятия решений"

DOI:

10.7256/2454-0714.2024.3.44097

EDN:

JOPHLF

Дата направления статьи в редакцию:

22-09-2023

Аннотация: Предметом исследования является понятие «доверенности» искусственного интеллекта, управляющего робототехническими средствами разной степени автономности. Актуальность выбора предмета исследования как принципов применения робототехнических систем различного назначения, в том числе в группе, и объекта исследования, которым являются алгоритмические проблемы, возникающие в части реализации алгоритмов групповых действий, определяется имеющимся противоречием между потребностью совместного применения робототехнических систем, в первую очередь автономных, и сложностью программной реализации этого требования. Реализация тенденций группового применения в сфере робототехники порождает определённые технологические проблемы, связанные с эффективностью и безопасностью алгоритмического обеспечения автономных и управляемых робототехнических систем. Проявлением подобных проблем могут являться ошибки применения, которые снижают эффективность совместных действий. В робототехнике

основная потенциальная причина появления подобной ситуации – недостаточная эффективность существующих алгоритмов управления групповым применением, определяемая низким уровнем проработки проблемы. В статье сформулирован перечень типовых ситуаций, определяющих применение автономных и управляемых роботов в группе с наличием ведущего (лидера). На основании предлагаемой классификации проанализированы возможные алгоритмы, обеспечивающие управление перемещением в группе: как расчёты по целевому маневрированию, так по обеспечению взаимной безопасности. Рассмотрены основные ситуации, относительно которых сформулированы виды маневра и описан математический аппарат для их расчётов. На основе обзорного анализа типовых алгоритмов управления перемещением в пространстве, синтезирована постановка научной задачи решения проблемы разработки групповых алгоритмов и математических методов, определяющих их оптимальное применение в системе, как аналога "новой этики" для робототехнических систем

Ключевые слова:

робототехническая система, искусственный интеллект, управляющее программное обеспечение, доверенный искусственный интеллект, контроль алгоритмов поведения, безопасное поведение, мораль робототехнических систем, этика робототехнических систем, математическое обеспечение безопасности, управление безопасным поведением

1 Введение

Одним из приоритетов, обеспечивающим активное развитие автономной робототехники, является использование элементов искусственного интеллекта (ИИ). В настоящее время компоненты ИИ содержат большинство робототехнических систем (РТС) с разными уровнями автономности:

- ограниченно автономные, например, дроны, самостоятельно возвращающиеся на базу при потере связи с оператором;
- частично автономные, с ограничением на автономное выполнение некоторых функций, например – применения оружия для боевых РТС, решение на которое пока остаётся за оператором;
- полностью автономные, решающие поставленные задачи самостоятельно.

Оснащение РТС компонентами ИИ, их дальнейшая автономизация с использованием принципов обучения и самообучения, всё более отдаляет их поведение от роботов с внешним управлением, всё более приближает к разумному, со всеми плюсами и минусами этих изменений. При этом, именно повышение автономности РТС является объективной тенденцией, основанной на требованиях к применению робототехники в самых разных сферах.

В то же время, остаётся целый ряд нерешенных проблем, периодически приводящих к авариям и катастрофам с человеческими жертвами в промышленной и транспортной сфере [1,2] и, конечно, в сфере вооруженного противоборства [3,4,5]. Многие из этих проблем определяются нерешенностью проблем предметной области в части определения безопасности систем ИИ. Используемый в настоящее время термин «доверенность» не в полной мере тождественен понятию «безопасность» в комплексном понимании этого явления, а других подходов к описанию данного фактора пока не

предлагается.

В то же время, именно описание предметной области служит основой развития теории и практики любых систем. И, одновременно, развитие автономных систем не может быть остановлено, оно продолжается, как продолжается технический прогресс, несмотря на некоторые ограничения, отражаемые в официальных документах, регулирующих сферу ИИ.

Таким образом, проблема уточнения предметной области в части обеспечения безопасности ИИ, управляющего РСТ не решена и остаётся актуальной, что делает актуальной и тему статьи.

2 О существующем подходе к определению «доверенных» систем

Практика показывает, что основа любой классификации – условия функционирования описываемой системы.

Обзор перечня и условий возникновения катастроф и аварий, произошедших по вине автономных и частично автономных РСТ показывает, что все их причины могут быть объединены в две большие группы;

- условно закономерные, связанные с решениями программного обеспечения, управляющего РСТ при неправильной формулировке задачи или ошибках распознавания ситуации, либо при неверном учёте ограничений на применение системы;
- случайные, определяемые или не обнаруженными в ходе тестирования ошибками алгоритмов поведения, либо ограниченным временем не принятие решения, не укладывающимся в заданную при разработке длительность цикла управления и не позволяющим программному обеспечению РСТ корректно отработать полный цикл управления.

Примером первой группы ошибок является озвученная полковником ВВС США Такером Гамильтоном (Tucker Hamilton) в докладе на конференции «Future Combat Air & Space Capabilities Summit» ситуация с решением ИИ убить собственного оператора, которого боевая РСТ считала помехой для достижения поставленной ей цели [\[6\]](#). В документах по регулированию разработки и использованию ИИ подобные ошибки иногда определяются как предвзятость или необъективность («bias»).

Примеров второго варианта ошибок, а именно – случайностей, существенно больше, как в транспортной сфере, так и в области роботизированных вооружений [\[7.8\]](#), но их последствия, как правило, менее опасны.

Ошибки первой группы являются критичными, так как напрямую затрагивают безопасность человека, они требуют надёжного решения, поиском которого научное сообщество в настоящее время занимается достаточно активно. Формализованную постановку по решению этих задач принято определять как создание «доверенного» искусственного интеллекта. Некоторые результаты исследований по созданию безопасного или, «доверенного» («trustworthy») искусственного интеллекта отражены в документе Еврокомиссии «Руководство по этике для надежного ИИ» 2019 года («Ethics guidelines for trustworthy AI, 2019») и в российском стандарте ГОСТ Р 59 276–2020 «Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения». В данных документах искусственный интеллект объявляется доверенным, если он обладает следующими свойствами: проверяемость, управляемость, стабильность,

робастность, безопасность и отказоустойчивость. При выполнении указанных требований ИИ предлагается считать «верифицированным по этическим характеристикам».

Впрочем, данными документами не решается целый ряд проблем в устраниении ошибок применения ИИ, некоторые из которых являются критичными.

Во-первых, в них в качестве одного из критерия «доверенности» задаётся непрерывный и безусловный контроль над поведением искусственного интеллекта человеком, существенную часть требований к доверенному ИИ, приводимых в рассмотренных документах, можно считать реализацией человекацентричного подхода к его разработке и внедрению, опирающемуся на три группы принципов: прозрачность, надежность, человекоцентричность.

В дополнение к сказанному можно отметить, что современными документами в сфере регулирования ИИ признаётся, что решением большинства проблем является отказ от рассмотрения ИИ как «черного ящика» [\[9,10\]](#), что должно обеспечиваться:

- тщательным контролем системы на всех этапах жизненного цикла;
- разработкой механизмов регулирования ИИ;
- разработкой кодексов и положений по использованию ИИ.

Но, строгое соблюдение этих принципов проблематично для автономных робототехнических систем: такой контроль или сильно ограничивает их возможности, или затруднён чисто технически.

Во-вторых, рассуждать в аспекте, определяемом данными документами, об этике ИИ не совсем логично. Доверенность ИИ по содержанию указанных документов скорее соответствует безопасности их алгоритмов для человека, а не набору мер по предотвращению случайных ошибок поведения, которую как раз и можно отнести к «этичности». Ибо, в робототехнике, понятию «этичности», скорее, соответствует приоритетный учёт интересов взаимодействующих систем, в том числе РТС, даже если для этого приходится жертвовать оптимальностью собственного поведения.

Поэтому «доверенность» ИИ можно определить, всё же, как аналог человеческой морали, интерпретированной для правил поведения РТС, потенциально решаемых средствами контроля безопасности алгоритмов программного обеспечения (ПО) РТС и формированием системы ограничений их поведения [\[11\]](#). Варианты таких подходов описаны в работах [\[12,13\]](#). В любом случае, так или иначе, проблема безопасного поведения постепенно решается, хотя и не в полном объёме и с определёнными терминологическими недочётами.

Сложнее дело обстоит с предупреждением возникновения ошибок второго типа. Они менее критичны с точки зрения безопасности, менее очевидны для анализа. Для их решения нужна разработка более гибких правил построения управляющих алгоритмов и системы ограничений, прототипом для которых, с определённой степенью допущений, можно считать правила не морали, а человеческой вежливости, интерпретированные для «доверенного» ИИ.

Возникающие противоречия в терминологии, при кажущейся незначительности, затрудняют разработку унифицированных правил поведения РТС и алгоритмов для их реализации. Такая ситуация требует решения, обеспечивающего логичное развитие теории и практики алгоритмизации ИИ, управляющего РТС.

С учётом выявленного противоречия в терминологии, основываясь на методе аналогий и проецируя некоторые правила человеческого поведения на алгоритмы и ограничения поведения искусственного интеллекта, предлагается сформулировать задачу реализации набора правил «морали» и «вежливости» для различных аспектов безопасности применения РТС разной степени автономности, обеспечивающих уточнение механизмов регулирования и кодексов поведения ИИ. В статье, учитывая меньшую изученность проблемы случайных ошибок поведения ИИ, рассмотрен вариант реализации правил для предотвращения возникновения именно случайных ошибок поведения, возникающих из-за неопределённости исходных данных, в том числе в связи с недостаточным временем для принятия решений.

3 Уточнение понятия «доверенность» в части безопасности взаимного поведения

Типичным примером случайной ошибки является возникновение риска столкновение наземных автономных РТС в процессе совместного маневрирования при ошибках учёта взаимного перемещения.

Модель возникновения подобной ситуации можно рассмотреть на простом примере. Допустим, два автономных транспортных средства, движущихся с одинаковой скоростью должны совершить поворот с огибанием препятствия. Средства маневрируют независимо, не работая в группе. ПО каждого из этих средств строит оптимальную для себя траекторию, с высокой вероятностью эти траектории пересекутся в точке поворота. Если указанные средства оснащены системами безопасного расхождения, столкновение маловероятно – данные системы определят опасность и ПО РТС сформируют маневр аварийного расхождения или экстренное изменение скорости. Но, проблема заключается именно в том, что расхождение будет критичным, формировать варианты расхождения, изменять скорость или траекторию, придётся в очень короткий промежуток времени, об оптимальности маршрута речи уже идти не будет. В таких условиях, и вероятность успешного расхождения не будет стопроцентной.

Это типичный случай несогласованного взаимодействия движущихся агентов.

В такой ситуации, если бы ИИ в составе ПО маневрирующих РТС заранее просчитал ситуацию, аварийного расхождения можно было бы избежать, заблаговременно уточнив параметры движения с учётом потенциальной проблемы. Незначительно отклонившись от оптимальной траектории на раннем этапе планирования, можно было бы избежать существенных потерь оптимальности или риска возникновения столкновения при расхождении.

Ситуация описана в несколько упрощённом виде, в реальности фоновая обстановка может быть сложнее, но для пояснения сущности проблемы, такой подход допустим.

В описанной постановке, для решения задачи по определению параметров маневра, могут быть использованы несколько вариантов, рассчитываемых заблаговременно с применением общеизвестного математического аппарата:

- спланировать взаимное расхождение, изменив траекторию и/или скорость обеим РТС;
- спланировать расхождение одной из систем, которая имеет меньше внешних ограничений для маневра.

Алгоритм «вежливости» или «этики» в вышеописанной ситуации, математически может быть реализован как расчёты по типовым вариантам расхождения, проведённые

относительно РТС, корректирующей траекторию [\[14,15\]](#).

Технически, наиболее простым способом расхождения является снижение скорости одного из участников на $\Delta V = V_{\text{нач}} - V_{\text{тр}}$ от начальной до требуемой, обеспечивающей расхождение на расстоянии ΔS .

Требуемое снижение скорости легко определяется из уравнения движения материальной точки, приняв её движение за прямолинейное и равноускоренное и используя в виде исходных данных начальную скорость движения $V_{\text{нач}}$ и расстояние до точки поворота (расхождения) S :

$$V_{\text{тр}} = V_{\text{нач}} \left(1 - \frac{\Delta S}{S} \right).$$

Кроме изменения скорости, можно обеспечить расхождение за счёт изменения траектории движения, смещения точки поворота, обеспечивая расхождение сзади или спереди на минимальном безопасном расстоянии.

Расчёты по выносу точки поворота в сторону от пересекаемой траектории могут быть описаны известными уравнениями выхода в заданную точку пространства, удалённую от траектории другого маневрирующего не менее, чем на расстояние безопасного расхождения L_{\min} . Расчётный курсовой угол маневрирования q_M в такой ситуации определяется как (рисунок 1):

$$q_M = q'_M 0 - a;$$

$$q'_{M0} = \arcsin \left(\frac{\sin q'_{T0}}{m} \right),$$

где $m = \frac{V_M}{V_T}$,

V_M – скорость маневрирующего;

V_T – скорость РТС, не меняющей траекторию.

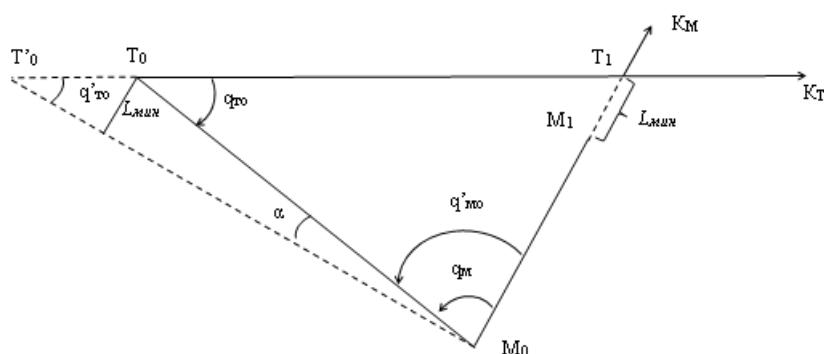


Рис. 1. Схема расчёта расхождения на заданном расстоянии

На рисунке:

K_M – курс маневрирующего;

K_T – курс РТС, не меняющей направления;

q_{T0} – начальный курсовой угол РТС, не меняющей направления движения;

q_{mo} – начальный курсовой угол маневрирующего;

q_m – расчётный курсовой угол маневрирующего;

T_0 и M_0 – точки начального положения РТС;

T'_0 – точка расхождения.

Расчёты по расхождению на минимальном расстоянии позади или по курсу на пересекаемой траектории, могут быть реализованы через уравнения расхождения на максимальной дистанции по курсу или на минимальной дистанции позади L_{min} :

$$L_{min} = L_K \frac{\sin(\Phi_{T0} - Q)}{\sin Q},$$

где Q – критический курсовой угол, предельный, на котором может быть осуществлено расхождение.

Значение Q рассчитывается по соотношению векторов скоростей:

$$Q = \frac{V_K}{V_N}.$$

Таким образом, существует достаточно широкий выбор матаппарата, позволяющего реализовать «вежливое» расхождение в разных ситуациях.

Выше описаны только варианты расчётов для решения проблемы, реализующие частные случаи. В целом, такая «вежливость» должна реализовываться расчётно-аналитическим алгоритмом, просчитываемым ПО РТС перед каждым маневром:

- 1) оценка окружающей обстановки, выделение потенциально взаимодействующих РТС и биологических объектов;
- 2) анализ координат и параметров движения потенциально взаимодействующих объектов;
- 3) построение оптимальной траектории собственного движения;
- 4) прогноз возможных пересечений;
- 5) принятие решение о необходимости коррекции траектории или скорости;
- 6) выбор способа коррекции;
- 7) расчёт уточнённых параметров движения.

Математически, как показывают приведённые примеры, данный алгоритм может быть обеспечен существующими методами, проблемным вопросом остаётся только формирование правил реализации «вежливого» поведения, а именно – кто должен совершать маневр уклонения. Эта задача является более сложной, лежащей на границе математики и логики.

Конечно, можно возразить, что подобные задачи уже решаются разработкой и применением наборов правил и нормативов, реализованных в специализированных документах: правилах дорожного движения, воздушных и водных кодексах. Но это предположение не совсем верно. Во-первых, движение и маневрирование, частные случаи применения РТС, реально спектр их применения намного шире. Во-вторых, эти

частные случаи реализуются в условиях существенных ограничений, определяемых, в частности, границами дорожного покрытия, разметкой, заранее задаваемыми трассами и эшелонами. В таких условиях может быть сформирован достаточно простой набор правил, таких как «расхождение левыми бортами», «помеха сверху» и тому подобное.

Намного более сложные и менее формализованные ситуации возникают во всех остальных случаях, характерных для функционирования РТС разной степени автономности. Более того, кроме слабоформализованных условий принятия решения, эти ситуации, как правило, осложняются ограниченным временем на его принятие.

В таких условиях, требуется разработка и реализация динамичных алгоритмов решения задач расхождения со слабоформализованными исходными данными.

Таким образом, при разработке алгоритмов для программного обеспечения РТС возникает нетривиальная научная задача – разработка правил взаимного поведения, аналога человеческого этикета. При этом, программный «этикет», как и обычный, может разделяться на ситуационный и профессиональный, а последний подразделяться на сферы применения с соответствующими вариациями действий. И, кстати, поведения при маневрировании, в предлагаемой постановке, является частью профессионального, а именно транспортного, «этикета». А всё, что связано с заведомо опасным для окружающих применением, должно решаться в рамках системы ограничений и запретов, аналогичных человеческой морали.

При таком подходе, ограничения и, соответственно, содержание понятия «доверенного» ИИ, будут строиться на двух наборах правил: аналогов «этикета» и «морали», с выбором вида профессионального этикета, подходящего для каждой из рассматриваемых ситуаций (рисунок 2).

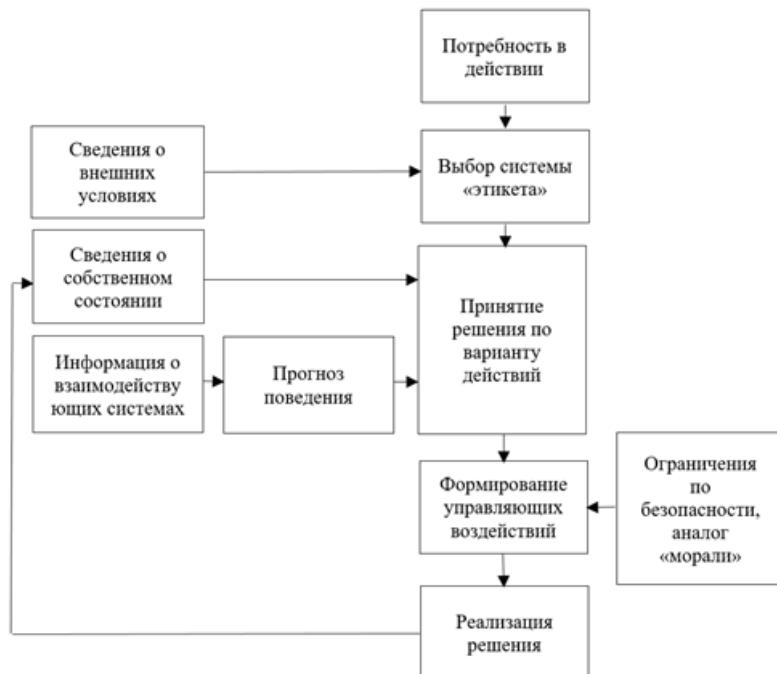


Рис. 2. Схема формирования взаимного поведения

Приведённый на рисунке алгоритм сформирован исходя из отсутствия обмена информацией между объектами. В то же время, развитие информационных технологий теоретически позволяет организовать подобный обмен и тогда алгоритм несколько упростится, так как:

- станет меньше неопределённостей в поведении других участников движения;
- появится возможность «договариваться» с другими участниками о параметрах взаимного расхождения.

То есть ситуация будет сведена к более простому варианту согласованного взаимодействия.

Но, в любом случае, алгоритм взаимного поведения, основанный на учёте параметров движения всех его участников с намеренным отступлением части из них от оптимальных параметров, то есть прообразом «вежливости» поведения, объективно необходим и требует дальнейшей разработки. Как требуется продолжение разработки ранее упомянутых алгоритмов безопасности в рамках правил поведения РТС, выстроенных по аналогии с «моралью».

Таким образом, использование аналогий «морали» и «вежливости» РТС позволит решить ряд важных проблем безопасности поведения автономных роботов, особенно в слабоформализуемых условиях взаимодействия, а основой такого решения должно служить как раз предлагаемое уточнение предметной области.

4 Заключение

Предложенная интерпретация понятий «морали» и «вежливости» для обучаемых алгоритмов программного обеспечения, управляющего автономными и частично автономными робототехническими системами, является лишь одним из вариантов решения проблемы, позволяющим выстроить систему их безопасности, основанную на группах правил и динамически формируемых ограничений.

Для обеспечения решения задач комплексной безопасности РТС, управляемых ИИ, требуется уточнить понятийный аппарат, используемый в данной предметной области. Причём речь идёт не о прямом копировании человеческих понятий морали и вежливости, а об уточнении понятия «доверенный» ИИ.

Указанная задача может быть решена двумя вариантами:

- оставить содержание определения «доверенного» ИИ как синоним ИИ с гарантированно безопасными алгоритмами поведения, добавив определение «вежливого» ИИ, обеспечивающего безопасное поведение в условиях группового применения на основе использования «этичных» алгоритмов;
- уточнить определение «доверенного» ИИ, добавив свойство «этичности» поведения, заключающегося в прогнозировании результатов предполагаемых действий и их проверки на основе набора правил.

В любом случае, предлагаемое уточнение классификации послужит предпосылкой к решению важной научно-практической задачи обеспечения комплексной безопасности РТС, управляемых ИИ. Более того, принятие предлагаемых изменений в понятие «доверенности», в перспективе, должно обеспечить переход от подхода на основе «обучения» ИИ к расширенному варианту, добавив к нему «воспитание», что также послужит повышению безопасности ИИ и управляемых им робототехнических систем.

Библиография

1. Симулин А. А. и др. Некоторые аспекты использования робототехники в военном деле // Сборники конференций НИЦ Социосфера. 2015. № 27. С. 67-71.

2. Чиров Д.С., Новак К.В. Перспективные направления развития робототехнических комплексов специального назначения // Вопросы безопасности. 2018. № 2. С. 50-59. DOI: 10.25136/2409-7543.2018.2.22737.
3. Хрипунов С.П., Благодаряющев И.В., Чиров Д.С. Военная робототехника: современные тренды и векторы развития // Тренды и управление. 2015. № 4. С. 410-422.
4. Pflimlin É. Drones et robots: La guerre des futurs. France: Levallois-Perret, 2017.
5. Roosevelt, Ann. Army Directs Cuts, Adjustments, To FCS. Defense Daily, 2017.
6. Hamilton T How AI will Alter Multi-Domain Warfare. Future Combat Air & Space Capabilities Summit, 2023. No. 4 URL: <https://www.aerosociety.com/events-calendar/raes-future-combat-air-and-space-capabilities-summit>
7. Tikhanychev O. Influence of the Problem of Safety Control of Heuristic Algorithms on the Development of Robotics. In: Shamtsyan, M., Pasetti, M., Beskopylny, A. (eds) Robotics, Machinery and Engineering Technology for Precision Agriculture. Smart Innovation, Systems and Technologies 2022, No. 247. Singapore: Springer. https://doi.org/10.1007/978-981-16-3844-2_31.
8. Beard J Autonomous weapons and human responsibilities, Georgetown Journal of International Law 2014. No. 45, P. 617-681.
9. Tikhanychev O The Control System of Heuristic Algorithms as a Prototype of the "Morality" of Autonomous Robots. II International Scientific Forum on Sustainable Development and Innovation. WFSDI-2023. URL: https://doi.org/10.1007/978-981-16-3844-2_31.
10. Ćwiąkała P. Testing Procedure of Unmanned Aerial Vehicles (UAVs) Trajectory in Automatic Missions. Appl. Sci. 2019, No. 9, P/ 3488. URL: <https://doi.org/10.3390/app9173488>.
11. Johnson D. Computer Systems: Moral entities but not moral agents. In: Ethics and Information Technology. 2016, No. 8, P. 195-204. URL: <https://doi.org/10.1007/s10676-006-9111>.
12. Schuller A. At the Crossroads of Control: The Intersection of Artificial Intelligence in Autonomous Weapon Systems with International Humanitarian Law, Harvard National Security Journal. 2017, No. 8, P. 379-425.
13. Ухоботов В.И., Измельцев И.В. Об одной задаче преследования при наличии сопротивления среди // Вестник ЮУрГУ. Серия: Математика. Механика. Физика. 2016. № 2. С. 62-66. URL: <https://doi.org/10.14529/mmp160208>.
14. Дубанов А.А. Моделирование траектории преследователя в пространстве при методе параллельного сближения // Программные системы и вычислительные методы. 2021. № 2. С. 1-10. URL: <https://doi.org/10.7256/2454-0714.2021.2.36014>
15. Tikhanychev O. Self-Check System of Heuristic Algorithms as a "New Moral" of Intelligent Systems AIP Conference Proceedings. 2023, No. 2700, 040028. URL: <https://doi.org/10.1063/5.0124956>.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Рецензируемая статья посвящена уточнению понятийного аппарата, используемого в интеллектуальных информационных системах, в частности понятия «доверенности» систем искусственного интеллекта.

Методология исследования базируется на обобщении публикаций зарубежных и отечественных ученых по рассматриваемой в статье проблематике, использовании

метода аналогий и математического моделирования возникновения риска столкновение наземных автономных робототехнических систем в процессе совместного маневрирования при ошибках учёта взаимного перемещения.

Актуальность работы авторы связывают с тем, что одним из приоритетов, обеспечивающим активное развитие автономной робототехники, является использование элементов искусственного интеллекта (ИИ), наличием нерешенной проблемы в части определения безопасности систем ИИ.

Научная новизна работы, по мнению рецензента состоит в предложенной интерпретации понятий «морали» и «вежливости» для обучаемых алгоритмов программного обеспечения, управляющего автономными и частично автономными робототехническими системами.

Структурно в статье выделены следующие разделы: Введение, О существующем подходе к определению «доверенных» систем, Уточнение понятия «доверенность» в части безопасности взаимного поведения, Заключение, Библиография.

Авторы указывают, что используемый в настоящее время термин «доверенность» не в полной мере тождественен понятию «безопасность» в комплексном понимании этого явления, а других подходов к описанию данного фактора пока не предлагается. В статье приведен обзор перечня и условий возникновения катастроф и аварий, произошедших по вине автономных и частично автономных робототехнических систем, изложен алгоритм «вежливости» или «этики» на математическом языке в виде расчётов по типовым вариантам расхождения, проведённым относительно робототехнических систем, корректирующей траекторию наземных автономных робототехнических систем в процессе совместного маневрирования. Текст публикации сопровождается пятью формулами и иллюстрирован двумя рисунками: «Схема расчёта расхождения на заданном расстоянии», «Схема формирования взаимного поведения». В Заключении указаны два варианта решения задачи уточнения понятия «доверенный» искусственный интеллект: во-первых, оставить содержание определения «доверенного» ИИ как синоним ИИ с гарантированно безопасными алгоритмами поведения, добавив определение «вежливого» ИИ, обеспечивающего безопасное поведение в условиях группового применения на основе использования «этичных» алгоритмов; во-вторых, уточнить определение «доверенного» ИИ, добавив свойство «этичности» поведения, заключающегося в прогнозировании результатов предполагаемых действий и их проверки на основе набора правил.

Библиографический список включает 15 источников – научные публикации по рассматриваемой теме на английском и русском языках. В тексте публикации имеются адресные ссылки к списку литературы, подтверждающие наличие апелляции к оппонентам.

Из резервов улучшения статьи следует отметить необходимость нумерации формул и оформления и в соответствии с принятыми правилами. С расшифровкой использованных символов непосредственно после математического выражения для улучшения и облегчения восприятия материала читателями.

Тема статьи актуальна, материал отражает результаты проведенного авторами исследования, содержит элементы приращения научного знания, соответствует тематике журнала «Программные системы и вычислительные методы», может вызвать интерес у читателей, рекомендуется к публикации после доформления формул.

Программные системы и вычислительные методы

Правильная ссылка на статью:

Скачёва Н.В. Анализ идиом в нейронном машинном переводе: набор данных // Программные системы и вычислительные методы. 2024. № 3. DOI: 10.7256/2454-0714.2024.3.71518 EDN: JLJDSL URL: https://nbpublish.com/library_read_article.php?id=71518

Анализ идиом в нейронном машинном переводе: набор данных

Скачёва Нина Васильевна

ORCID: 0000-0003-2628-6767

старший преподаватель; кафедра лингвистики, теории и практики перевода; Сибирский Государственный Университет Науки и Технологий им. академика М.Ф. Решетнева

660037, Россия, Красноярский край, г. Красноярск, Красноярский рабочий, 31

✉ Sollo_sk@mail.ru



[Статья из рубрики "Математическое моделирование и вычислительный эксперимент"](#)

DOI:

10.7256/2454-0714.2024.3.71518

EDN:

JLJDSL

Дата направления статьи в редакцию:

19-08-2024

Аннотация: В разных кругах общественности ни одно десятилетие идут споры о том, может ли «машина заменить человека». Это касается и области перевода. И пока, одни рассуждают, другие «воплощают мечту в реальность». Поэтому сейчас всё больше исследований направлены на усовершенствование систем машинного перевода (далее МП). Чтобы понять преимущества и недостатки систем МП, необходимо, в первую очередь разобраться в их алгоритмах. На данный момент основной открытой проблемой нейронного машинного перевода (НМП) является перевод идиоматических выражений. Значение таких выражений не складывается из значений составляющих их слов, и модели НМП склонны переводить их буквально (т. е. дословно), что приводит к запутанным и бессмысленным переводам. Исследования идиом в НМП ограничены и затруднены из-за отсутствия автоматических методов. Поэтому несмотря на то, что современные системы НМП генерируют все более качественные переводы, перевод идиом остается одной из нерешенных задач в этой области. Это происходит из-за того, что идиомы, как категория многословных выражений, представляют собой интересное

языковое явление, когда общее значение выражения невозможно составить из значений его частей. Первая важная проблема – отсутствие специальных наборов данных для обучения и оценки перевода идиом. В данной работе мы решаем эту проблему, создавая первый крупномасштабный набор данных для перевода идиом. Данный набор данных автоматически извлекается из используемого корпуса переводов с немецкого языка и включает в себя целевой набор, в котором все предложения содержат идиомы, и обычный обучающий корпус, в котором предложения, содержащие идиомы, помечаются. Мы выпустили этот набор данных и используем его для проведения предварительных экспериментов по НМП в качестве первого шага к улучшению перевода идиом.

Ключевые слова:

многословные выражения, идиомы, двуязычные корпусы, машинный перевод, нейронный машинный перевод, немецкий язык, русский язык, лингвистические пары, системы, набор данных

Введение

Нейронный машинный перевод (НМП) в последние годы позволил существенно повысить качество перевода по сравнению с традиционным переводом, основанным на статистическом переводе (SMT) и на правилах и фразах (РВМТ). Для понимания вопроса, рассмотрим как работают SMT и РВМТ.

Область статистического машинного перевода стала развиваться благодаря появлению большего количества параллельных текстов в лингвистике. Такие параллельные тексты можно встретить в различных лингвистических корпусах. Одним из самых первых и крупных является корпус Europarl. Это сборник протоколов европейского парламента, сформированный с 1996 года и состоящий в то время из 11 языков европейского союза. Этот корпус использовался при создании 110 систем машинного перевода.[\[1\]](#)

Для русского языка важным стало развитие Национального корпуса русского языка (далее НКРЯ), а именно один из его корпусов – параллельный. Который включает в себя такие языки как: английский, армянский, белорусский, болгарский, испанский, итальянский, китайский, немецкий. В корпусе НКРЯ английского языка на данный момент насчитывается 1322 текста и 45 235 028 слов.[\[2\]](#) Сейчас он находится на стадии развития. Большинство систем в значительной степени независимы от языка, и создание системы SMT для новой пары языка в основном зависит от наличия параллельных текстов. Поэтому так важно создавать параллельные тексты в паре с русским языком.

Изначально перевод в SMT строился на словесных моделях IBM.[\[3\]](#) Современные модели уже основаны на фразах.[\[4\]](#) При переводе предложение, фраза исходного языка в любой последовательности слов преобразуются во фразы на целевом языке. Ядром этой модели является вероятностная фраза, полученная из параллельного корпуса. Декодирование представляет собой лучевой поиск по всем возможным сегментациям ввода фразы, любой перевод для каждой фразы и любого изменения её порядка. Учёные пишут о том, что, используя формулу Байеса, можно представить проблему перевода предложения в виде следующего уравнения:

$$\arg \max P(\phi|f\phi) = \arg \max (P(\phi) \cdot P(f\phi|\phi))$$

Уфे Уфе

Где \square_e — это фраза перевода, а \square_r — фраза оригинала. Поэтому модель языка $P(\square)$ является корпусом текста на языке \square_e , а модель перевода $P(\square_r | \square_e)$ — это параллельный корпус текста на языках \square_r и \square_e .

В таких системах МП в качестве языковой модели используют n -граммы. [\[5\]](#) У Google есть своя коллекция n -грамм, которая является на данный момент самой крупной коллекцией в мире. [\[6\]](#) Самой большой коллекцией n -грамм на русском языке, конечно же является НКРЯ.

Согласно моделям n -грамм, предсказывающее последующее слово после $n1..$, выявляют возможную последовательность из определенного количества слов. Таким образом есть бигаммы, состоящие из двух слов, тригаммы и так далее. Когда количество увеличивается, то подсчитывается вероятность следующего слова или последующих слов. При больших переменных это может выглядеть так:

$$P(n1, n2, n3, n4) = P(n4 | n3, n2, n1) * P(n3 | n2, n1) * P(n2 | n1) * P(n1)$$

Рассчитаем вероятность предложения: Ich sehe ein Auto auf der Strasse

$$P(Ich)$$

$$P(sehe | Ich)$$

$$P(ein Auto | Ich sehe)$$

$$P(Auto | Ich sehe ein)$$

$$P(auf | Ich sehe ein Auto)$$

$$P(der | Ich sehe ein Auto auf$$

$$P(Strasse | Ich sehe ein Auto auf der)$$

Так как мы чаще встречаем $auf der Strasse$, чем просто $der Strasse$, то согласно марковской модели [\[7\]](#) наше предложение будет выглядеть так:

$$P(Strasse | auf der)$$

То есть марковская модель n -ого порядка будет выглядеть как:

$$P(A_1 | A_1, A_2, \dots, A_{i-1}) \approx P(A_1 | A_{i-n}, A_{i-n+1}, \dots, A_{i-1})$$

Таким образом, разбитие на n -граммы при переводе текста позволяет найти наиболее удачный перевод текста.

Проблема таких моделей состоит в том, что система не всегда определяет связи между словами, особенно если такие слова стоят далеко друг от друга.

То есть, данные системы МП получили свое распространение благодаря небольшого вложения человеческих ресурсов при условии существования двух параллельных языковых корпусов. Такие системы обучаемы и чем больше текстов находится в параллельных корпусах, тем более адекватнее будет перевод нового текста данной языковой группы. Проблемой для такой системы является перевод идиом, так как они не определяют связи между словами и не чувствительны к контексту.

RBMT – это парадигма МП, в которой лингвистические знания кодируются экспертом в форму правил, которые транслируются из языка оригинала в язык перевода^[8]. Данный подход дает полный контроль над производительностью системы, но стоимость формализации необходимых лингвистических знаний значительно выше, чем обучение корпусной системы, в которой процессы происходят автоматически. Однако у данного МП есть свои возможности, даже в условиях невысоких ресурсов.

В RBMT лингвист формализует лингвистические знания в грамматические правила. Система использует такие знания для анализа предложений на исходном языке и языке перевода. Большой плюс такой системы в том, что она не требует каких-либо учебных корпусов, но процесс кодирования лингвистических знаний требует большое количество экспертного времени. Самый крупный проект перевода, построенный на системе RBMT – это *Systran*.^[9]

RBMT – это противоположность систем МП, которые учатся на основе корпусов, представленных нами здесь ранее. Поэтому такая система МП подходит для тех языковых пар, в которых существует мало параллельных корпусов и, как следствие, может охватывать больше языковых пар.

Основной подход системы RBMT основан на связи структуры входного предложения исходного языка со структурой выходного предложения переводимого языка, сохраняя уникальный смысл предложений. Но RBMT также мало восприимчив к переводу идиом.

В чем сложность их перевода? Сложность перевода идиоматических фраз отчасти объясняется сложностью идентификации фразы как идиоматической и создания ее правильного перевода, а отчасти – тем, что идиомы редко встречаются в стандартных наборах данных, используемых для обучения систем нейронного машинного перевода (НМП).

Чтобы проиллюстрировать проблему перевода идиом, мы также приводим результаты работы двух систем НМП для этого предложения в Google и DeepL (см. Таблица 1). Эта проблема особенно ярко проявляется, когда исходная идиома сильно отличается от ее эквивалента в языке перевода, как в данном случае.

Идиома	Es liegt der Hase im Pfeffer!
Идиоматический перевод	Вот где собака зарыта!
DeepL	Кролик в норе!
Google	Это большое дело!

Таблица 1. Перевод идиомы в системах НМП

Хотя существует ряд моноязычных наборов данных для выявления идиоматических выражений, работы по созданию параллельного корпуса, аннотированного идиомами, необходимого для более систематического изучения этой проблемы, ограничены. Например, американские учёные выбрали небольшое подмножество из 17 английских идиом, собрали 10 примеров предложений для каждой идиомы из интернета и вручную перевели их на бразильско-португальский язык, чтобы использовать в задаче перевода^[10].

Создание набора данных для перевода идиом вручную – дорогостоящее и трудоемкое занятие. В этой статье мы автоматически создаем новый двуязычный набор данных для перевода идиом, извлеченных из существующего параллельного корпуса немецко-

русских текстов общего назначения.

Первая часть нашего набора данных состоит из 1500 параллельных предложений, немецкая часть которых содержит идиому. Кроме того, мы предоставляем соответствующие наборы обучающих данных для немецко-русского и русско-немецкого перевода, где отмечены исходные предложения, включающие фразу-идиому. Мы считаем, что наличие большого набора данных для обучения и оценки — это первый шаг к улучшению перевода идиом.

Сбор данных

В данной работе мы сосредоточились на немецко-русском переводе идиом. Для автоматического определения фразеологизмов в параллельном корпусе требуется набор данных, аннотированный вручную лингвистами. Мы используем словарь, содержащий идиоматические и разговорные фразы и созданный вручную, в качестве эталона для извлечения пар идиоматических фраз. При этом обнаружено, что стандартные параллельные корпусы, доступные для обучения, содержат несколько таких пар предложений. Поэтому мы автоматически выбираем пары предложений из обучающих корпусов, в которых исходное предложение содержит фразу-идиому, для создания нового тестового набора.

Обратите внимание, что мы фокусируемся только на идиомах на стороне источника и имеем два отдельных списка идиом для немецкого и русского языков, поэтому мы независимо создаем два тестовых набора (для перевода немецких идиом и перевода русских идиом) с различными парами предложений, выбранными из параллельных корпусов.

Например, в немецком языке подлежащее может находиться между глаголом и предложной фразой, составляющей идиому. Немецкий язык также допускает несколько вариантов перестановки фраз. Чтобы обобщить процесс выявления вхождений идиом, мы видоизменяем фразы и рассматриваем различные перестановки слов в фразе как приемлемое соответствие. Мы также допускаем, что между словами идиоматической фразы может находиться фиксированное количество слов.

Следуя этому набору правил, мы извлекаем пары предложений, содержащие идиоматические фразы, и создаем набор пар предложений для каждой уникальной идиоматической фразы. На следующем этапе мы делаем выборку без замены из этих наборов и выбираем отдельные пары предложений для создания тестового набора.

Для создания новых обучающих данных мы используем оставшиеся пары предложений из каждого набора идиом, а также пары предложений из исходных параллельных корпусов, в которых не было ни одной фразы-идиомы. В этом процессе мы следим за тем, чтобы для каждого идиоматического выражения была хотя бы одна форма как в обучающих, так и в тестовых данных, и чтобы ни одно предложение не было включено как в обучающие, так и в тестовые данные.

При этом для некоторых идиом дословный перевод на язык перевода близок к реальному значению. Пары предложений, в которых идиоматическое выражение использовалось в качестве буквальной фразы будут идентифицированы как идиоматические предложения.

Переводческие эксперименты

Хотя основное внимание в этой работе уделяется созданию наборов данных для обучения и оценки перевода идиом, мы также проводим ряд предварительных

экспериментов НМП с использованием нашего набора данных, чтобы оценить проблему перевода идиом на больших массивах данных.

В первом эксперименте мы не используем никаких меток в данных для обучения модели перевода. Во втором эксперименте мы используем метки в обучающих данных в качестве дополнительной характеристики, чтобы исследовать наличие идиоматической фразы в предложении во время обучения.

Мы проводим эксперимент с немецким и русским языками, предоставляя модели дополнительные входные признаки. Дополнительные признаки указывают, содержит ли исходное предложение идиому, и реализованы в виде специальной дополнительной лексемы, которая добавляется к каждому исходному предложению, содержащему идиому. Это простой подход, который можно применить к любой модели преобразования последовательности в последовательность.

Большинство систем НМП имеют модель преобразования последовательности в последовательность, в которой кодировщик строит представление исходного предложения, а декодировщик, используя предыдущие скрытые элементы LSTM и механизм внимания, генерирует целевой перевод. Мы используем 4-слойную модель кодирования-декодирования на основе внимания, как описывают в работе Тханг Луонг, Хиен Пхам, Чристопхер Д. Маннинг.[\[11\]](#)

Во всех экспериментах словарный запас НМП ограничен наиболее распространенными 30 тыс. слов в обоих языках, и мы предварительно обрабатываем данные исходного и целевого языков, используя 30 тыс. операций слияния.

Мы также используем систему перевода на основе фраз, подобную Moses[\[12\]](#), в качестве базового уровня, чтобы исследовать показатели РВМТ при переводе идиом.

Оценка перевода идиом

В идеале перевод идиом должен оцениваться вручную, но это очень дорогостоящий процесс. С другой стороны, автоматические метрики могут быть использованы на больших массивах данных без особых затрат и имеют преимущество в воспроизводимости.

Для оценки качества перевода мы используем следующие метрики, уделяя особое внимание точности перевода идиом: BLEU Традиционная оценка BLEU[\[13\]](#) является хорошим показателем для определения общего качества перевода. Однако эта мера учитывает точность всех n-грамм в предложении и сама по себе не фокусируется на качестве перевода идиоматических выражений.

Модифицированная униграммная точность. Чтобы сконцентрироваться на качестве перевода идиоматических выражений, мы также смотрим на локализованную точность. При таком подходе мы переводим идиоматическое выражение в контексте предложения и оцениваем только качество перевода идиоматической фразы.

Чтобы выделить перевод идиомы в предложении, мы смотрим на выравнивание на уровне слов между выражением идиомы в исходном предложении и сгенерированным переводом в целевом предложении. Для выравнивания слов мы используем функцию fast-align.[\[14\]](#) Поскольку идиоматические фразы и соответствующие переводы во многих случаях не являются смежными, мы сравниваем только униграммы двух фраз.

Обратите внимание, что для этой метрики у нас есть две ссылки: Перевод идиомы как самостоятельного выражения и перевод идиомы, созданный человеком, в целевом предложении.

Точность перевода идиом на уровне слов. Мы также используем другую метрику для оценки точности перевода фразы-идиомы на уровне слов. Мы используем выравнивание слов между исходным и целевым предложениями, чтобы определить количество правильно переведенных слов. Для расчета точности мы используем следующее уравнение:

$$WIAcc = \frac{I}{N}$$

где N - количество правильно переведенных слов, I - количество лишних слов в переводе идиомы, а N - количество слов в эталонном переводе идиомы.

В таблице 5 представлены результаты для задачи перевода с использованием различных метрик.

Модель	BLEU	BLEU	Униграммная точность	Точность на уровне слов
Базовый уровень				
PBMT	20,2	19,7	57,7	71,6
Базовый уровень	26,9	24,8	53,2	67,8
НМП				
Лексема НМП	25,2	22,5	64,1	73,2
источника				

Таблица 2. Эффективность перевода на тестовом наборе немецких идиом. Точность идиом на уровне слов и Униграммная точность вычисляются только для фразы-идиомы и ее соответствующего перевода в предложении

Эксперимент НМП с использованием специальной входной лексемы, указывающей на наличие идиомы в предложении, по-прежнему лучше, чем PBMT, но немного хуже, чем базовый вариант НМП, по показателю BLEU. Несмотря на такое падение показателя BLEU, изучая униграммную точность перевода и точность перевода идиом на уровне слов, мы видим, что эта модель генерирует более точные переводы идиом.

Эти предварительные эксперименты подтверждают проблему перевода идиом с помощью нейронных моделей и, кроме того, показывают, что при наличии набора маркированных данных мы можем разработать простые модели для решения этой проблемы.

Выводы

Перевод идиом - одна из самых сложных задач машинного перевода. В частности, было показано, что нейронный МП плохо справляется с переводом идиом, несмотря на его общее преимущество перед предыдущими парадигмами МП. В качестве первого шага к лучшему пониманию этой проблемы мы представили параллельный набор данных для обучения и тестирования перевода идиом для немецкого-русского и русского-немецкого языков.

Тестовые наборы включают предложения с хотя бы одной идиомой на стороне источника, а обучающие данные представляют собой смесь идиоматических и неидиоматических предложений с метками, позволяющими отличить их друг от друга. Мы также провели предварительные эксперименты по переводу и предложили различные метрики для

оценки перевода идиом. Мы формируем новые наборы данных, которые могут быть использованы для дальнейшего изучения и улучшения работы НМП при переводе идиом.

Библиография

1. Koehn P. Europarl: A Parallel Corpus for Statistical Machine Translation // School of Informatics University of Edinburgh, Scotland. 2005. P. 79-86.
2. Национальный корпус русского языка. URL: <https://ruscorpora.ru/search?search=CgkyBwgFEgNIbmcwAQ%3D%3D> (дата обращения 04.03.2024)
3. Brown P. F., Pietra S. A. D., Pietra V. J. D., Mercer R. L. The mathematics of statistical machine translation. Computational Linguistics. 1993. 19(2), p. 263-313.
4. Philipp Koehn, Franz J. Och, and Daniel Marcu.. Statistical Phrase-Based Translation. In Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics. 2003. p. 127-133. URL: <https://aclanthology.org/N03-1017.pdf> (дата обращения 05.03.2024)
5. Гудков В. Ю., Гудкова Е.Ф. N-граммы в лингвистике // Вестник ЧелГУ. 2011. № 24.
6. Лингвистический корпус данных. URL: <https://catalog.ldc.upenn.edu/byyear> (дата обращения 05.03.2024)
7. Жданов А. Е., ДоросинскийЛ.Г. Голосовой замок // Ural Radio Engineering Journal. 2017. Vol. 1, No. 1. P. 80-90.
8. Daniel Torregrosa, Nirvanshu Pasricha, Bharathi Raja Chakravarthi, Maraim Masoud, Mihael Arcan. Leveraging Rule-Based Machine Translation Knowledge for Under-Resourced Neural Machine Translation Models // Proceedings of MT Summit XVII, Dublin, 2019, volume 2. URL: <https://aclanthology.org/W19-6725.pdf> (дата обращения: 19.08.2024).
9. Peter T. Systran as a multilingual machine translation system // Overcoming the language barrier, 3-6 May 1977, Vol. 1. URL: <https://www.mt-archive.net/70/CEC-1977-Toma.pdf> (дата обращения: 19.08.2024).
10. Salton G., Ross R., and Kelleher J. (2014). An empirical study of the impact of idioms on phrase based statistical machine translation of english to brazilian-portuguese // In Proceedings of the 3rd Workshop on Hybrid Approaches to Machine Translation (HyTra). 2014. P. 36-41.
11. Luong T., Pham H., Manning C. D. (2015). Effective approaches to attention-based neural machine translation // In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. 2015. P. 1412-1421.
12. Koehn P., Hoang H., Birch A., Callison-Burch C., Federico M., Bertoldi N., Cowan B., Shen W., Moran C., Zens R. Moses: Open source toolkit for statistical machine translation // In Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions. 2007. P. 177-180.
13. Papineni K., Roukos S., Ward T., and Zhu W.J. Bleu: a method for automatic evaluation of machine translation. In Proceedings of 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania, USA. 2002. P. 311-318.
14. Dyer C., Chahuneau V., and Smith N. A simple, fast, and effective reparameterization of ibm model 2 // In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies Atlanta, Georgia, June. 2013 p. 644-646.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Статья посвящена исследованию проблемы перевода идиом в контексте нейронного машинного перевода (НМП). Авторы предлагают новый подход к созданию набора данных для перевода идиом на немецко-русских и русско-немецких языковых парах. В статье рассматриваются трудности перевода идиоматических выражений, обусловленные особенностями таких фраз, а также предлагается методология их выявления и перевода.

Методология исследования основывается на автоматическом выявлении идиоматических выражений в параллельных корпусах текстов. Для создания нового набора данных использовался словарь идиом, а также разработаны правила для автоматического выявления и обработки фразеологических единиц. Экспериментальная часть исследования включает сравнение различных подходов к переводу идиом с использованием нейронных моделей и методов статистического машинного перевода (SMT и PBMT). Авторы также предлагают новые метрики для оценки качества перевода идиом.

Актуальность исследования очевидна в условиях растущей популярности и повсеместного использования систем машинного перевода. Проблема точного и адекватного перевода идиоматических выражений является значительным вызовом для существующих систем НМП. Учитывая, что идиомы часто несут важные смысловые и культурные оттенки, разработка эффективных методов их перевода имеет высокую практическую значимость.

Научная новизна работы заключается в предложении оригинального подхода к созданию параллельного корпуса, специально адаптированного для задачи перевода идиом. Авторы представили новый метод аннотирования и автоматического выявления идиом, что является значительным вкладом в область машинного перевода. Также важно отметить предложенные авторами новые метрики для оценки качества перевода идиоматических выражений, которые позволяют более точно оценивать успешность перевода таких фраз.

Стиль изложения статьи академически строг и последовательный, что способствует легкому восприятию сложного материала. Структура статьи логично выстроена: после введения и обзора существующих методов перевода идиом, авторы переходят к описанию методологии исследования, а затем представляют результаты экспериментов. Такой подход позволяет читателю последовательно погружаться в тему исследования и понимать ключевые аспекты предложенных методов.

В заключительной части статьи авторы подводят итоги проведенного исследования, акцентируя внимание на необходимости дальнейшего изучения проблемы перевода идиом и улучшения существующих моделей нейронного машинного перевода. Авторы предлагают перспективные направления для дальнейших исследований, что делает статью полезной как для исследователей в области машинного перевода, так и для практиков, работающих с системами НМП.

Статья представляет собой значимый вклад в область машинного перевода, особенно в контексте перевода идиоматических выражений. Исследование демонстрирует высокий уровень проработки проблемы, оригинальность предложенных решений и их практическую значимость. Рекомендуется к публикации в представленном виде.

Программные системы и вычислительные методы

Правильная ссылка на статью:

Скляр А.Я. Численные методы нахождения корней многочленов с действительными и комплексными коэффициентами // Программные системы и вычислительные методы. 2024. № 3. DOI: 10.7256/2454-0714.2024.3.71103 EDN: KTJPCE URL: https://nbpublish.com/library_read_article.php?id=71103

Численные методы нахождения корней многочленов с действительными и комплексными коэффициентами

Скляр Александр Яковлевич

кандидат технических наук

доцент, кафедра прикладная математика; Российский технологический университет (МИРЭА)

119602, Россия, г. Москва, Вернадского, 78

✉ askliar@mail.ru



[Статья из рубрики "Математическое моделирование и вычислительный эксперимент"](#)

DOI:

10.7256/2454-0714.2024.3.71103

EDN:

KTJPCE

Дата направления статьи в редакцию:

23-06-2024

Аннотация: Предметом исследования является рассмотрение и анализ набора алгоритмов численного нахождения корней многочленов, прежде всего комплексных на основе методов поиска приближенного разложения исходных полиномов на множители. Если численное нахождение действительных корней обычно не вызывает трудностей, то с нахождением комплексных корней возникает ряд сложностей. В данной статье предлагается набор алгоритмов последовательного нахождения кратных корней многочленов с действительными корнями, далее действительных корней выделением интервалов, потенциально содержащих корни и заведомо не содержащих их, а затем комплексных корней многочленов. Для нахождения комплексных корней используется итеративное приближение исходного многочлена произведением трехчлена на многочлен меньшей степени с последующим использованием метода касательных в комплексной области в окрестности корней полученного трехчлена. Для нахождения корней многочлена с комплексными коэффициентами предлагается решение эквивалентной задачи с действительными коэффициентами. Реализация поставленных задач осуществляется поэтапным применением комплекса алгоритмов. После каждого

этапа выделяется группа корней и решается та же задача для многочлена меньшей степени. Последовательность предлагаемых алгоритмов позволяет найти все как действительные, так и комплексные корни многочлена. Для нахождения корней многочлена с действительными коэффициентами строится алгоритм, включающий следующие основные этапы: определение кратных корней с соответствующим снижением степени полинома; выделение диапазона корней; нахождение интервалов, гарантированно содержащих корни и их нахождением, по их выделении остается найти только пары комплексно сопряженных корней; итеративное построение трехчленов, служащих оценкой значений таких пар с минимальной точностью, достаточной для их локализации; собственно поиск корней в комплексной области методом касательных. Вычислительная трудность предлагаемых алгоритмов является полиномиальной и не превосходит куба от степени многочлена, что позволяет получить решение для практически любых многочленов, возникающих в реальных задачах. Областью приложения помимо собственно полиномиальных уравнений является и сводимые к ним задачи оптимизации, дифференциальных уравнений и оптимального управления.

Ключевые слова:

многочлены, нахождение корней, итерационные методы, численные методы, численные алгоритмы, алгебраическое уравнение, сопряженные комплексные корни, рекурсивные алгоритмы, корни многочленов, локализация корней

Введение

Существует множество задач самого разного характера, в которых требуются определение корней многочленов. Алгебраические уравнения возникают при изучении равновесных состояний сложных термодинамических и механических систем, часто они появляются в аэродинамике, в механике полета. Например, скорость быстрейшего набора высоты самолета определяется из алгебраического уравнения восьмой степени. Алгебраические уравнения возникают также при выполнении разнообразных геометрических расчетов – определение точек пересечения и сопряжения криволинейных контуров, при проектировании гладких поверхностей, хорошо обтекаемых тел и многих других задачах.

Методы решения уравнений до второй степени известны еще со времен древней Греции. В XVI веке получены аналитические выражения для многочленов 3 степени (формула Кордано) и 4 степени, полученные Сципионом дель Ферро, Тарталья, Феррари [3].

В как 20-х гг. Абель, а затем Галуа в 30-х гг. XIX в. доказали, что такие формулы для уравнений n -й степени в общем случае при любом $n \geq 5$ заведомо не могут быть найдены.

Для численного приближенного решения уравнений высших степеней в настоящее время используются различные методы, такие как метод Лобачевского, метод Хичкока, схема Горнера для деления многочлена на двучлен и квадратный трехчлен [1, 4] и другие.

Другой тип алгоритмов, основанный на получении итерационных формул, за счет выделения из многочленов простых и квадратичных множителей, с последующим сопоставлением записей многочленов с остатком, когда корни являются приближенными, и без остатка, когда значения корней являются точными рассмотрены в статьях Чье Ен Ун

и А.Б. Шеина [5,6,7].

Ряд алгоритмов решения полиномиальных задач и обширная библиография приведены у Г. П. Кутищева [8]. Стоит отметить также подходы для их решения в [9, 10, 11].

Заметим, что само существование большого числа разнообразных методов в целом свидетельствует, что не существует ни одного «вполне удовлетворительного».

1. Удаление кратных корней

Многочлен степени n имеет в точности n корней и представим в виде

$$P_n(x) = a \prod_{k=1}^n (x - x_k) \quad (1.1)$$

В общем случае не все корни x_k различны.

Нахождения корней многочлена наталкивается на определенные трудности в случаях, когда он имеет кратные корни.

Пусть имеется кратный корень x^* с кратностью m , тогда $P_n(x)$ можно представить в виде

$$P_n(x) = (x - x^*)^m G_{n-m}(x) \quad (1.2)$$

$$\begin{aligned} P'_n(x) &= m(x - x^*)^{m-1} G_{n-m}(x) + (x - x^*)^m G'_{n-m}(x) \\ &= (x - x^*)^{m-1} (mG_{n-m}(x) + (x - x^*) G'_{n-m}(x)) \end{aligned}$$

Следовательно, $P_n(x)$ и $P'_n(x)$ будут иметь общий делитель $(x - x^*)^{m-1}$.

Тогда, используя алгоритм Евклида можно найти наибольший делитель $Q(x)$ многочленов $P_n(x)$ и $P'_n(x)$.

$$Q(x) = \prod_{i=1}^l (x - x_i)^{m_i} \quad (1.3)$$

И исходный многочлен $P_n(x)$ быть представлен, как

$$P_n(x) = Q(x)H(x) \quad (1.4)$$

Многочлен $H(x)$ при этом будет иметь те же корни, что и $P_n(x)$, но не будет иметь кратных корней. Таким образом исходная задача сводится к нахождению корней многочлена, все корни которого различны.

2. Определение диапазона корней

Рассмотрим уравнение (a_0 будем полагать равным 1)

$$x^n + \sum_{k=0}^{n-1} a_k x^k = 0 \quad (2.1)$$

Согласно известной теореме [1,2] все корни x_k ($k = 1, 2, \dots, n$) многочлена в комплексной плоскости лежат в кольце

$r < |x_k| < R$, где

$$R < 1 + A; r > \frac{1}{1 + B/|a_0|} \quad (2.2)$$

$$A = \max_{k \in [1, n]} |a_k|; B = \max_{k \in [0, n-1]} |a_k|$$

В то же время сразу отметим, что это утверждение можно несколько усилить.

Проведем замену переменной $x=y/t$. В этом случае исходное уравнение относительно y примет вид

$$y^n + \sum_{k=0}^{n-1} a_k t^{n-k} y^k = 0 \quad (2.3)$$

$$R_y < 1 + A_y; r_y > \frac{1}{1 + B_y/(|a_0|t^n)}$$

$$A_y = \max_{k \in [1, n]} |a_k t^{n-k}|; B_y = \max_{k \in [0, n-1]} |a_k t^{n-k}|$$

Учитывая характер подстановки, можно записать

$$R < \frac{1 + A_y}{t} = \max_{k \in [1, n]} \left(\frac{1}{t} + |a_k| t^{n-k-1} \right)$$

Меняя значение t , можно добиться уменьшения оценки верхней границы.

Пусть максимальное значение принимает модуль коэффициента при y_m . Тогда оценка по этому коэффициенту будет не менее, чем

$$t^* = \min_t \left(\frac{1}{t} + |a_m| t^{n-m-1} \right) \quad (2.4)$$

t^* находится из

$$-\frac{1}{t^2} + (n-m-1)|a_m|t^{n-m-2} = 0; t^* = \frac{1}{((n-m-1)|a_m|)^{1/(n-m)}}$$

Таким образом можно утверждать, что оценка лежит между значениями, соответствующими текущему значению t и t^* .

В то же время при изменении t меняются и другие коэффициенты. Таким образом получаем ограничения на пересечение линий их изменения.

$$\frac{1}{t_r} + |a_m| t_r^{n-m-1} = \frac{1}{t_r} + |a_k| t_r^{n-k-1}; |a_m| t_r^k = |a_k| t_r^m$$

Если t_r не лежит внутри интервала t и t^* , то его можно игнорировать. Из остальных выбираем значение ближайшее к t .

Если таковых нет, то оптимальное $t_r=t^*$.

Если знаки производных

$$\frac{1}{t_r} + |a_m| t_r^{n-m-1}, \frac{1}{t_r} + |a_k| t_r^{n-k-1}$$

совпадают, то дальнейшее улучшение невозможно. Оптимальное $t_r=t^*$.

Если знаки производных совпадают, то принимаем m равным k и повторяем процедуру.

Рассмотрим производную по k коэффициенту в точке t

$$z' = -\frac{1}{t^2} + (n-k-1)|a_k|t^{n-k-2}; t^2 z' = (n-k-1)|a_k|t^{n-k} - 1$$

Рассмотрим в качестве исходного многочлен четвертой степени

$$t^4 - 10t^3 + 35t^2 - 50t + 24 = (t-1)(t-2)(t-3)(t-4)$$

Исходная верхняя граница – 51, $m=1$, $t=1$,

$$t^* = 1 / ((n-m-1)|a_m|)^{\frac{1}{n-m}} = 1 / ((4-2)50)^{\frac{1}{4-1}} = \sqrt[3]{0,01} \approx 0.2714$$

Точки пересечения

$$|a_1|t_r^k = |a_3|t_r^m; 50t_r^3 = 10; t_r = \sqrt[3]{0,2} \approx 0,5848; t^2 z' = -1 < 0$$

$$|a_1|t_r^k = |a_2|t_r^m; 50t_r^2 = 35; t_r = \sqrt{0,7} \approx 0.8366; t^2 z' = 35 * 0,7 - 1 > 0$$

$$|a_1|t_r^k = |a_0|t_r^m; 50t_r = 24; t_r = 0,48; t^2 z' = 3 * 24 * 0,49 - 1 > 0$$

И окончательно получаем

$$R = \frac{1}{t_r} + |a_m|t_r^{n-m-1} = \frac{1}{\sqrt{0,7}} + 50 \cdot 0,7 \approx 36,195$$

3. Поиск действительных корней

Рассмотрим отдельно поиск положительных и отрицательных корней.

Пусть исходный многочлен имеет вид (старший коэффициент a_0 не снижая общности можно полагать равным 1).

$$P(x) = x^n + \sum_{k=0}^{n-1} a_k x^k \quad (3.1)$$

Введем $b_k = (a_k + |a_k|)/2$ и $c_k = (a_k - |a_k|)/2$. Для $a_k > 0$ $b_k = a_k$, $c_k = 0$ для $a_k \leq 0$ $b_k = 0$, $c_k = -a_k$. В этих обозначениях

$$P(x) = \sum_{k=0}^n b_k x^k - \sum_{k=0}^n c_k x^k = P_+(x) - P_-(x) \quad (3.2)$$

Многочлены $P_+(x)$ и $P_-(x)$ при $x > 0$ представляют собой непрерывные неотрицательные монотонно возрастающие функции.

Используя введенные многочлены $P_+(x)$ и $P_-(x)$ и верхнюю границу значений корней R будем искать корни в диапазоне $[A, B]$, где $A=0$, $B=R$. Нижнюю границу можно уточнить до величины r^* , получив диапазон $[r^*, R^*]$ вместо диапазона $[r, R]$, но это не имеет принципиального значения.

Рассмотрим подробнее алгоритм нахождения корней.

Вычисляем значения $P_+(x)$ и $P_-(x)$ на концах интервала и вызываем функцию Root вычисления корней с параметрами A , B , $P_+(A)$, $P_-(A)$ и $P_+(B)$, $P_-(B)$. Функция Root возвращает либо найденный корень (значение большее 0), либо -1, если на данном интервале многочлен $P(x)$ не имеет корней.

Алгоритм функции Root имеет следующий вид.

1. Вычисляем значения исходного многочлена на концах интервала $P(A)=P_+(A)-P_-(A)$ и $P(B)=P_+(B)-P_-(B)$

2. Если $P(A)=0$ возвращаем найденный корень A , если $P(B)=0$ возвращаем найденный корень B , если $P(A)P(B)$ вызываем стандартную процедуру STROOT нахождения корней на заданном интервале $[A, B]$ (например, методом дихотомии или хорд) и возвращаем, найденное ей значение.

3. Вычисляем $Q_1=P_+(A)-P_-(B)$ – нижняя граница значений $P(x)$ на интервале $[A, B]$ и $Q_2=P_+(B)-P_-(A)$ – верхняя граница значений $P(x)$ на интервале $[A, B]$.

4. Если $Q_1Q_2 \geq 0$, то на данном интервале корней нет и возвращаем значение -1.

5. Вычисляем $C=(A+B)/2$.

Вызываем функцию Root вычисления корней с параметрами $A, C, P_+(A), P_-(A)$ и $P_+(C), P_-(C)$.

6. Если вычисленное значение $x > 0$, то возвращаем найденный корень x .

7. Вызываем функцию Root вычисления корней с параметрами $C, B, P_+(C), P_-(C)$ и $P_+(B), P_-(B)$. Возвращаем найденный корень x (если корней нет, то -1).

Приведенный рекурсивный алгоритм позволяет найти корень многочлена в указанном диапазоне или убедиться, что внутри него нет действительных корней.

По нахождении корня u переходим к поиску следующего корня, заменив исходный многочлен $P_n(x)$ многочленом меньшей степени $P_{n-1}(x)$, $P_n(x)=(x-u)P_{n-1}(x)$. Многочлен $P_{n-1}(x)$ имеет все оставшиеся корни $P_n(x)$. Повторяя приведенный выше алгоритм, находим все положительные корни.

Для нахождения отрицательных корней заменим исходный многочлен $P_n(x)$ многочленом $(-1)^n P_n(-x)$, корни которого имеют те же значения, но с обратным знаком. Найдя все положительные корни второго многочлена, мы тем самым найдем все действительные корни исходного многочлена.

Поскольку других действительных корней нет, то оставшиеся корни представляют пары комплексно-сопряженных чисел и оставшийся многочлен имеет четную степень.

4. Поиск комплексных корней

Пусть многочлен $P_{2n}(x)$ с действительными коэффициентами не имеет действительных корней, тогда он представим в виде

$$P_{2n}(x) = x^{2n} + \sum_{k=0}^{2n-1} a_k x^k = \prod_{k=1}^n (x - x_k)(x - \bar{x}_k) =$$

$$\prod_{k=1}^n (x^2 + p_{k1}x + p_{k0}); p_{k1}^2 - 4p_{k0} > 0 \quad (4.1)$$

Рассмотрим представление многочлена степени $2n$ в виде

$$R_n(x) = (x^2 + p_1x + p_0) \sum_{k=0}^{n-2} q_k x^k = \sum_{k=0}^n r_k x^k \quad (4.2)$$

$$r_m = q_m p_0 + q_{m-1} p_1 + q_{m-2}; m = 2, \dots, n-1$$

Для стандартизации расчетов для всех m введем $q_{-2}=q_{-1}=q_{n-1}=q_n \equiv 0$, $q_{n-2}=1$.

Нахождение корней будем проводить на основе оптимизации аппроксимации исходного многочлена $P_n(x)$ многочленом $R_n(x)$.

Введем функцию

$$F(p, q) = \frac{1}{2} \sum_{k=0}^{n-1} (q_k p_0 + q_{k-1} p_1 + q_{k-2} - a_k)^2 = \frac{1}{2} \sum_{k=0}^{n-1} z_k^2 \quad (4.3)$$

При заданных значениях p_0, p_1 коэффициенты q_i находятся из требования минимизации $F(p, q)$.

$$\begin{aligned} \frac{\partial F}{\partial q_k} &= z_{k+2} \frac{\partial z_{k+2}}{\partial q_k} + z_{k+1} \frac{\partial z_{k+1}}{\partial q_k} + z_k \frac{\partial z_k}{\partial q_k} = z_{k+2} + z_{k+1} p_1 + z_k p_0 \\ &= (q_{k+2} p_0 + q_{k+1} p_1 + q_k - a_{k+2}) + (q_{k+1} p_0 + q_k p_1 + q_{k-1} - a_{k+1}) p_1 \\ &\quad + (q_k p_0 + q_{k-1} p_1 + q_{k-2} - a_k) p_0 \\ &= q_{k+2} p_0 + q_{k+1} p_1 (p_0 + 1) + q_k (p_0^2 + p_1^2 + 1) + q_{k-1} p_1 (p_0 + 1) + q_{k-2} p_0 \\ &\quad - (a_{k+2} + p_1 a_{k+1} + p_0 a_k) = 0 \\ &q_{k+2} p_0 + q_{k+1} p_1 (p_0 + 1) + q_k (p_0^2 + p_1^2 + 1) + q_{k-1} p_1 (p_0 + 1) + q_{k-2} p_0 \\ &= a_{k+2} + p_1 a_{k+1} + p_0 a_k \end{aligned} \quad (4.4)$$

Или

$$z_{k+2} + z_{k+1} p_1 + z_k p_0 = 0 \quad (4.5)$$

При заданных значениях q_i коэффициенты p_0, p_1 находятся из требования минимизации $F(p, q)$.

$$\begin{aligned} \frac{\partial F}{\partial p_0} &= \sum_{k=0}^{n-1} z_k \frac{\partial z_k}{\partial p_0} = \sum_{k=0}^{n-1} (q_k p_0 + q_{k-1} p_1 + q_{k-2} - a_k) q_k = 0 \\ \frac{\partial F}{\partial p_1} &= \sum_{k=0}^{n-1} z_k \frac{\partial z_k}{\partial p_1} = \sum_{k=0}^{n-1} (q_k p_0 + q_{k-1} p_1 + q_{k-2} - a_k) q_{k-1} = 0 \\ \left\{ \begin{array}{l} p_0 \sum_{k=0}^{n-1} q_k^2 + p_1 \sum_{k=0}^{n-1} q_k q_{k-1} = \sum_{k=0}^{n-1} (a_k - q_{k-2}) q_k \\ p_0 \sum_{k=0}^{n-1} q_k q_{k-1} + p_1 \sum_{k=0}^{n-1} q_{k-1}^2 = \sum_{k=0}^{n-1} (a_k - q_{k-2}) q_{k-1} \end{array} \right. \end{aligned} \quad (4.6)$$

Или

$$\begin{cases} \sum_{k=0}^{n-2} z_k q_k = 0 \\ \sum_{k=0}^{n-1} z_k q_{k-1} = 0 \end{cases} \quad (4.7)$$

Для нахождения многочлена $x^2 + p_1x + p_0$ можно воспользоваться следующим итеративным алгоритмом.

1. Задаем начальное значение коэффициентов p_0, p_1 .
2. На основе (4.5) рассчитываем коэффициенты q_i . Задача сводится к решению системы линейных уравнений (СЛАУ) для 5-диагональной матрицы. Учитывая вид матрицы, задача сводится к ряду подготовительных операций трудности $O(n^2)$ и решению СЛАУ размерности 3·3.
3. Полученный набор коэффициентов q_i используем для получения в соответствии с (4.6) скорректированных значений p_0, p_1 .
- 4 . Оцениваем величину погрешности вычисления в соответствии с (4.3). Если погрешность меньше заданного порога, то заканчиваем работу алгоритма. В противном случае переходим к шагу 2 алгоритма.

В результате получаем многочлена $x^2 + p_1x + p_0$ дающий первую пару корней. Коэффициенты q_i дают многочлен

$$\sum_{k=0}^{n-2} q_k x^k,$$

который может быть использован приведенным алгоритмом для получения следующей пары корней.

Отметим, что скорость сходимости данного алгоритма невелика. Для ускорения можно воспользоваться обработкой ранее полученных значений.

Пусть на i -ом и $i+1$ -ом шагах получили значения $p_{i,1}, p_{i+1,1}$ и $p_{i,0}, p_{i+1,0}$ с невязками $F(p,q) - Fa, Fb$ соответственно.

Вычислим значения $p_{c1}=p_{i,1}+2(p_{i+1,1}-p_{i+1,1})$ и $p_{c0}=p_{i+1,0}+2(p_{i+1,0}-p_{i,0})$ и соответствующее ему значение Fc .

Если $F_c < F_b$ проводим замену точек и значений в них $a=b, b=c$ и повторяем расчет. Если $F_c < F_b$, то квадратичной интерполяцией находим точку c^* , принимая ее за субоптимальное значение и возвращаемся в основной алгоритм.

В точке оптимума должны одновременно выполняться (4.5 и 4.6)

В матричном виде совокупность (4.5, 4.6) представима в виде

$$AZ = \begin{pmatrix} 1 & p_1 & p_0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & p_1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 & 0 \\ \vdots & & & \ddots & & \vdots & \\ 0 & 0 & 0 & \cdots & 1 & p_1 & p_0 \\ 0 & 1 & q_{n-3} & \cdots & q_2 & q_1 & q_0 \\ 1 & q_{n-3} & q_{n-4} & \cdots & q_1 & q_0 & 0 \end{pmatrix} \begin{pmatrix} z_{n-1} \\ z_{n-2} \\ z_{n-3} \\ \vdots \\ z_2 \\ z_1 \\ z_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (4.8)$$

При $|A| \neq 0$ алгоритм сходится к решению задачи.

Кроме того, с точки зрения практической реализации значительно эффективнее выглядит гибридный алгоритм.

Несколько итераций базового алгоритма приводят нас в окрестность корня. Учитывая, что при отсутствии кратных корней $|P'_n(x)|$ гарантированно не обращается в 0, при $P_n(x)=0$ схема с использованием значений производных (метод Ньютона), либо более тонких его модификаций [12] обеспечивает достаточно быструю сходимость.

Пусть получено приближение $s=x^2+p_1x+p_0$. Его корень будет

$$z_0 = -\frac{p_1}{2} + i\sqrt{p_0 - \frac{p_1^2}{4}}$$

Значение $P_n(z)$ в окрестности z_0 представимо в виде

$$P_n(z) \approx P_n(z_0) + P'_n(z_0)(z - z_0)$$

Это позволяет рассчитать приближение корня z^* .

$$z^* = z_0 - \frac{P_n(z)}{P'_n(z_0)}$$

При z_0 , близких к z^* алгоритм сходится достаточно быстро. В результате получаем для $s=x^2+p_1x+p_0$

$$p_1 = z^* + \bar{z}^*, p_0 = z^* \bar{z}^*.$$

Полученное выражение используем для получения многочлена $Q_{n-2}(x)$ из $P_n(x) = (x^2 + p_1x + p_0)Q_{n-2}(x)$, после чего ищем следующие корни, многочлена, совпадающие с корнями $Q_{n-2}(x)$.

5. Поиск корней многочленов с комплексными коэффициентами

Рассмотрим уравнение (a_0 будем полагать равным 1). Коэффициенты a_k – комплексные числа.

$$P_n(x) = x^n + \sum_{k=0}^{n-1} a_k x^k = 0 \quad (5.1)$$

Пусть его корни – x_1, x_2, \dots, x_n , тогда корни уравнения

$$Q_n(x) = x^n + \sum_{k=0}^{n-1} \overline{a_k} x^k = 0 \quad (5.2)$$

будут

$\overline{x_1}, \overline{x_2}, \dots, \overline{x_n}$

Построим многочлен $R_{2n}(x) = P_n(x)Q_n(x)$. Этот многочлен имеет корни

 $x_1, \overline{x_1}, x_2, \overline{x_2}, \dots, x_n, \overline{x_n}$

$$R_{2n}(x) = x^{2n} + \sum_{k=0}^{2n-1} b_k x^k \quad (5.3)$$

Величины b_k представимы как суммы и произведения комплексно-сопряженных чисел и, следовательно являются действительными числами. Таким образом многочлен $R_{2n}(x)$ является многочленом с действительными коэффициентами.

Соответственно поиск корней многочлена $P_n(x)$ с комплексными коэффициентами сводится к поиску корней многочлена $R_{2n}(x)$ с действительными коэффициентами. Отметим, что все действительные его корни будут кратными, а из полученных комплексных сопряженных корней часть будет посторонней.

Заключение

Последовательность предлагаемых алгоритмов позволяет найти все как действительные, так и комплексные корни многочлена.

Для нахождения корней многочлена степени n с действительными коэффициентами строится алгоритм, включающий следующие основные этапы:

- определение кратных корней;
- выделение диапазона корней;
- нахождение интервалов, гарантированно содержащих корни (за время $O(\ln(R/H))$), где R – оценка общего диапазона корней, а H длина интервала, гарантированно содержащего корень);
- итеративное построение трехчленов, служащих оценкой значений пар комплексно-сопряженных корней (количество итераций не обязательно должно соответствовать требованиям точности решения, достаточно выполнения требования изоляции корней, после чего решение достигается традиционными методами, например методом Ньютона).

Для нахождения корней многочлена степени n с комплексными коэффициентами строится вспомогательный многочлен степени $2n$ с действительными коэффициентами, содержащий все корни исходного многочлена степени n с комплексными коэффициентами, для которого используется приведенный выше алгоритм. Вычислительная сложность алгоритма вполне приемлема для расчетов корней многочленов любой, встречающейся в реальных задачах, степени.

Программная реализация предлагаемых алгоритмов выполнена на языке C++ в среде операционной системы Windows. Учитывая вычислительный характер алгоритмов перенос их в иную операционную среду не вызывает каких-либо проблем.

Библиография

1. Курош А.Г. Курс высшей алгебры. Москва: Наука, 1968. С. 431.
2. Самарский А. А., Гулин А. В. Численные методы. Москва: Наука, 1989. С. 432.
3. Стиллвелл Д. Математика и её история. – Москва-Ижевск: Институт компьютерных

- исследований, 2004. С. 530.
4. Тынкевич М. А., Пимонов А. Г. Введение в численный анализ. Кемерово: КузГТУ. 2017. С. 176.
5. Чье Ен Ун, Шеин А.Б. Метод нахождения корней многочленов. I // Информатика и системы управления. 2012. № 4(34). С. 88-96.
6. Чье Ен Ун, Шеин А.Б. Метод нахождения корней многочленов. II // Информатика и системы управления. 2013. №1(35). С. 108-118.
7. Чье Ен Ун, Шеин А.Б. Метод нахождения корней многочленов. III // Информатика и системы управления. 2013. №3 (37). С. 110-122.
8. Кутищев Г.П. Решение алгебраических уравнений произвольной степени: Теория, методы, алгоритмы. URSS. 2015. 232 с.
9. Simon Telen. Polynomial Equations: Theory and Practice. Michal Kočvara; Bernard Mourrain; Cordian Riener. Polynomial Optimization, Moments, and Applications, Springer, pp. 215-240.
10. B. Mourrain and J. P. Pavone. Subdivision methods for solving polynomial equations. Journal of Symbolic Computation, 44(3), 292-306, 2009.
11. Berthomieu, C. Eder, and M. Safey El Din. msolve: A library for solving polynomial systems. In Proceedings of the 2021 on International Symposium on Symbolic and Algebraic Computation, pages 51-58, 2021.
12. Стаценко И. В. Исследование скорости сходимости одного обобщенного ньютоновского метода и классического метода ньютона в процедуре уточнения корней многочлена. // Точная наука. 2020. №78. С. 2-9.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Рецензируемая статья посвящена обобщению сведений о численных методах нахождения корней многочленов с действительными и комплексными коэффициентами. Методология исследования базируется на изложении алгоритмов математических действий при нахождении корней многочленов с действительными и комплексными коэффициентами, рассмотрении определенных трудностей, возникающих в процессе решения задач.

Актуальность работы авторы связывают с тем, что существует множество задач самого разного характера, в которых требуется определение корней многочленов.

Научная новизна рецензируемого исследования состоит обобщении сведений о численных методах нахождения корней многочленов и в предлагаемых алгоритмах, которые позволяют найти все действительные и комплексные корни многочлена. В тоже время, представляется уместным провести сравнение предлагаемых подходов с уже известными и опубликованными ранее результатами исследований других авторов, показать отличия и преимущества авторского видения способов решения рассматриваемых задач.

Структурно в работе выделены следующие разделы: Введение, Удаление кратных корней, Определение диапазона корней, Поиск действительных корней, Поиск комплексных корней, Заключение, Библиография.

Авторами изложены алгоритмы нахождения корней многочлена степени n с действительными коэффициентами, включающие такие этапы как определение кратных корней; выделение диапазона корней; нахождение интервалов, гарантированно содержащих корни; итеративное построение трехчленов, служащих оценкой значений

пар комплексно - сопряженных корней.

Библиографический список включает 8 источников – публикации отечественных и зарубежных авторов по рассматриваемой теме за период с 1968 по 2020 гг. В тексте публикации имеются адресные ссылки к списку литературы, подтверждающие наличие апелляции к оппонентам.

Из недостатков публикации, требующих своего устранения, стоит отметить следующие моменты. Во-первых, актуальность проведения исследования не раскрыта с достаточной ясностью. После первого предложения во введении хотелось бы увидеть примеры, демонстрирующие необходимость практического применения рассматриваемых методов и нерешенные вопросы в применении существующих подходов. Во-вторых, представляется уместным провести сравнение предлагаемых подходов с уже известными и опубликованными ранее результатами исследований других авторов, показать различия и преимущества авторского видения способов решения рассматриваемых задач. В-третьих, с учетом названия журнала, в котором публикуется статья, уместно было бы осветить вопросы программной реализации рассматриваемых алгоритмов, достижения и проблемы этого аспекта применения численных методов нахождения корней многочленов с действительными и комплексными коэффициентами. В-четвертых, заголовки четвертого раздела и Заключение не выделены полужирным шрифтом; в предпоследнем предложении имеется несогласованное словосочетание.

Рецензируемый материал соответствует направлению журнала «Программные системы и вычислительные методы», отражает результаты проведенного авторского исследования, может вызвать интерес у читателей, но нуждается в доработке в соответствии с высказанным замечанием и последующем рецензировании скорректированного материала.

Результаты процедуры повторного рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Статья посвящена изучению численных методов для нахождения корней многочленов с действительными и комплексными коэффициентами. Рассматриваются как классические подходы, такие как метод Лобачевского, схема Горнера и метод Ньютона, так и новые итеративные алгоритмы, предложенные авторами. Особое внимание уделяется проблеме нахождения кратных корней и комплексных корней, что делает работу актуальной и востребованной в области прикладной математики и инженерных наук.

Авторами предложены несколько численных методов для нахождения корней многочленов. Основной акцент сделан на итеративных алгоритмах, позволяющих последовательно уточнять приближения корней. Методология включает в себя использование модификаций классических методов, а также предложенные алгоритмы, направленные на повышение точности и эффективности вычислений. Методические подходы подробно объяснены и подкреплены примерами, что позволяет легко следовать изложенной логике.

Нахождение корней многочленов является ключевой задачей в различных областях науки и техники, таких как механика, аэродинамика, проектирование сложных инженерных систем. В условиях, когда аналитические решения уравнений высокой степени часто невозможны, численные методы становятся основным инструментом для исследователей и инженеров. Данная работа актуальна ввиду потребности в высокоточных и эффективных численных алгоритмах, способных справляться с задачами различной сложности.

Научная новизна работы заключается в разработке и предложении новых итеративных методов нахождения корней многочленов, а также в усовершенствовании существующих подходов. Авторы детализируют алгоритмы, позволяющие находить как действительные, так и комплексные корни, учитывая их кратность. Представленные методы демонстрируют высокую точность и эффективность, что подтверждается проведенными вычислительными экспериментами.

Статья написана в ясном и последовательном стиле, что способствует лёгкому восприятию материала. Структура работы логично выстроена: начинается с введения и постановки задачи, далее следуют методологические разделы, описывающие предложенные алгоритмы, и заканчивается заключением, в котором подведены итоги и сделаны выводы. Содержание статьи полностью соответствует заявленной теме и охватывает все ключевые аспекты численного решения многочленов.

В статье сделаны обоснованные выводы о том, что предложенные методы могут быть эффективно использованы для нахождения корней многочленов в различных прикладных задачах. Авторы демонстрируют, что их подходы превосходят классические методы по точности и быстродействию, особенно в сложных случаях, таких как наличие кратных и комплексных корней.

Статья представляет интерес для широкого круга специалистов, работающих в области прикладной математики, численного анализа, вычислительной механики, а также для инженеров, занимающихся проектированием сложных систем. В работе найдут полезную информацию как теоретики, так и практики, что делает её востребованной в научной и инженерной среде.

Для дальнейшего развития данной работы рекомендуется расширить исследование в направлении применения предложенных численных методов к реальным задачам из различных областей науки и техники, таких как аэродинамика, механика и компьютерная графика. Это позволит не только продемонстрировать практическую значимость разработанных алгоритмов, но и выявить возможные ограничения и области для их дальнейшего совершенствования. В частности, полезно было бы провести сравнительное исследование с существующими методами на большем количестве примеров с реальными данными, что позволит более детально оценить эффективность и точность предложенных подходов. Кроме того, стоит рассмотреть возможность адаптации и оптимизации алгоритмов для их реализации на современных параллельных вычислительных платформах, что позволит значительно ускорить процесс нахождения корней многочленов, особенно в задачах с высокой размерностью и сложностью.

Рекомендуется принять статью к публикации без значительных доработок. Представленные методы и результаты являются ценным вкладом в область численного анализа и будут полезны для дальнейших исследований и практических приложений.

Англоязычные метаданные

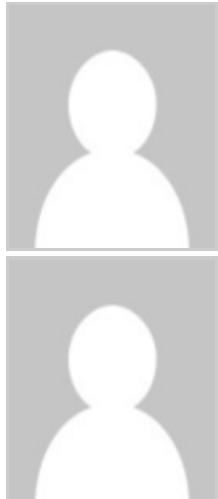
Architecture of a three-dimensional convolutional neural network for detecting the fact of falsification of a video sequence

Alpatov Aleksey Nikolaevich

Associate Professor; liPPO Department; MREA - Russian Technological University

78 Vernadsky Ave., Moscow, 119454, Russia

✉ aleksej01-91@mail.ru



Terloev Emil' Ziyaudinovich

Postgraduate student; Department of Instrumental and Applied Software; MREA—Russian Technological University

78 Vernadsky Ave., Moscow, 119454, Russia

✉ emil199@yandex.ru



Matchin Vasili Timofeevich

Senior Lecturer; Institute of Information Technology; MREA—Russian Technological University

78 Vernadsky Ave., Moscow, 119454, Russia

✉ matchin@mirea.ru



Abstract. The article reflects the use of neural network technologies to determine the facts of falsification of the contents of video sequences. In the modern world, new technologies have become an integral part of the multimedia environment, but their proliferation has also created a new threat – the possibility of misuse to falsify the contents of video sequences. This leads to serious problems, such as the spread of fake news and misinformation of society. The scientific article examines this problem and determines the need to use neural networks to solve it. In comparison with other existing models and approaches, neural networks have high efficiency and accuracy in detecting video data falsification due to their ability to extract complex features and learn from large amounts of source data, which is especially important when reducing the resolution of the analyzed video sequence. Within the framework of this work, a mathematical model for identifying the falsification of audio and video sequences in video recordings is presented, as well as a model based on a three-dimensional convolutional neural network to determine the fact of falsification of a video sequence by analyzing the contents of individual frames. Within the framework of this work, it was proposed to consider the problem of identifying falsifications in video recordings as a joint solution to two problems: identification of falsification of audio and video sequences, and the resulting problem itself was transformed into a classical classification problem. Any video recording can be assigned to one of the four groups described in the work. Only the videos belonging to the first group are considered authentic, and all the others are fabricated. To increase the flexibility of the model, probabilistic classifiers have been added, which allows to take into account the degree of confidence in the predictions. The peculiarity of the resulting solution is the ability to adjust the threshold values, which allows to adapt the model to different levels of rigor depending on the task. The architecture of a three-

dimensional convolutional neural network, including a preprocessing layer and a neural network layer, is proposed to determine fabricated photoreceads. The resulting model has a sufficient degree of accuracy in determining falsified video sequences, taking into account a significant decrease in frame resolution. Testing of the model on a training dataset showed the proportion of correct detection of video sequence falsification above 70%, which is noticeably better than guessing. Despite the sufficient accuracy, the model can be refined to more significantly increase the proportion of correct predictions.

Keywords: batch normalization, anomaly detection, data preprocessing, audio falsification, deepfake detection, deepfakes, video falsification, convolutional neural networks, neural networks, machine learning

References (transliterated)

1. Beyan E.V. P., Rossy A.G.C. A review of AI image generator: influences, challenges, and future prospects for architectural field // Journal of Artificial Intelligence in Architecture. 2023. V. 2. №. 1. Pp. 53-65.
2. Huang Y. F., Lv S., Tseng K.K., Tseng P.J., Xie, X., Lin, R.F.Y. Recent advances in artificial intelligence for video production system // Enterprise Information Systems. 2023. V. 17. №. 11. Pp. 2246188.
3. Albert V. D., Schmidt H. J. AI-based B-to-B brand redesign: A case study // Transfer. 2023. P. 47.
4. Aliev E. V. Problemy ispol'zovaniya tsifrovyykh tekhnologii v kinoindustrii // European Journal of Arts. 2023. No1. S. 33-37. DOI: <https://doi.org/10.29013/EJA-23-1-33-37>
5. Chow, P. S. Ghost in the (Hollywood) machine: Emergent applications of artificial intelligence in the film industry // NECSUS_European Journal of Media Studies. 2020. V. 9. №. 1. Pp. 193-214.
6. Lemai kina S. V. Problemy protivodei stviya ispol'zovaniya dipfei kov v prestupnykh tselyakh // Yurist"-Pravoved". 2023. No 2(105). S. 143-148.
7. Vakilinia I. Cryptocurrency giveaway scam with youtube live stream // 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). 2022. Pp. 0195-0200.
8. Tran D., Wang H., Torresani L., Ray J., LeCunY., Paluri M. A closer look at spatiotemporal convolutions for action recognition // Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2018. Pp. 6450-6459.
9. Naik K. J., Soni A. Video classification using 3D convolutional neural network // Advancements in Security and Privacy Initiatives for Multimedia Images. IGI Global. 2021. Pp. 1-18.
10. ZF DeepFake Dataset [Elektronnyi resurs] URL: <https://www.kaggle.com/datasets/zfturbo/zf-deepfake-dataset> (data obrashcheniya: 20.01.2024).
11. Garbin C., Zhu X., Marques O. Dropout vs. batch normalization: an empirical study of their impact to deep learning // Multimedia tools and applications. 2020. V. 79. №. 19. Pp. 12777-12815.
12. Zhou D. X. Theory of deep convolutional neural networks: Downsampling // Neural Networks. 2020. V. 124. Pp. 319-327.

Configuration of memory-oriented motion control system

Zelenskii Aleksandr Aleksandrovich

PhD in Technical Science

Leading researcher; Scientific and production complex 'Technological Center'

124498, Russia, Moscow, Shokin square, 1, building 7

✉ zelenskyaa@gmail.com

**Gribkov Andrey Armovich**

Doctor of Technical Science

Leading researcher; Scientific and production complex 'Technological Center'

124498, Russia, Moscow, Shokin square, 1, building 7

✉ andarmo@yandex.ru



Abstract. The paper investigates the possibilities of configuring the control cycle, i.e., determining the distribution of time intervals required for the execution of individual control operations across execution threads, which ensures the realizability of control. The object of research in this article are control systems with object-oriented architecture, assuming a combined vertical-horizontal integration of functional blocks and modules that distribute all control tasks among themselves. This architecture is realized by means of an actor instrumental model using metaprogramming. Such control systems are best at reducing control cycle time by performing computational and other control operations in parallel. Several approaches to control cycle configuration are considered: without optimization, with combinatorial optimization in time, with combinatorial optimization in system resources. Also, achieving a near-optimal configuration can be achieved by using adaptive configuration. Research shows that the control system cycle configuration problem has several solutions. Practical obtaining a solution to the configuration problem in the case of combinatorial optimization is associated with significant difficulties due to the high algorithmic complexity of the problem and a large amount of required computations, rapidly growing as the number of operations at the stages of the control cycle. A possible means of overcoming these difficulties is the use of stochastic methods, which sharply reduce the required amount of computation. Also, a significant reduction in the complexity of the task of configuring the control system cycle can be achieved by using adaptive configuration, which has two variants of realization. The first variant is the real-time configuration of the control system cycle. The second variant is the determination of quasi-optimal configuration on the basis of multiple configurations with different initial data and subsequent comparison of the obtained results.

Keywords: control operations, elements, loop, optimization, configuration, adaptive, sorting methods, execution threads, memory-oriented, control system

References (transliterated)

1. Zelenskii A.A., Kuznetsov A.P., Ilyukhin Yu.V., Gribkov A.A. Realizuemost' upravleniya dvizheniem promyshlennykh robotov, stankov s ChPU i mekhatronnykh sistem. Chast' 1 // Vestnik mashinostroeniya. 2022. №11. S. 43-51.
2. Zelenskii A.A., Kuznetsov A.P., Ilyukhin Yu.V., Gribkov A.A. Realizuemost' upravleniya dvizheniem promyshlennykh robotov, stankov s ChPU i mekhatronnykh sistem. Chast' 2 // Vestnik mashinostroeniya. 2023. №3. S. 213-220.
3. Cell Broadband Engine Programming Tutorial. Version 2.0. IBM Systems and Technology Group, December 15, 2006. URL:

https://arcb.csc.ncsu.edu/~mueller/cluster/ps3/CBE_Tutorial_v2.0_15December2006.pdf

4. Ghose S., Hsieh K., Boroumand A., Ausavarungnirun R., Mutlu O. Enabling the Adoption of Processing-in-Memory: Challenges, Mechanisms, Future Research Directions. 2018. URL: <https://arxiv.org/abs/1802.00320>
5. Singh G., Chelini L., Corda S., Awan A.J., Stuijk S., Jordans R., Corporaal H., Boonstra A. Near-Memory Computing: Past, Present, and Future. August 2019, Microprocessors and Microsystems 71. URL: https://www.researchgate.net/publication/335028505_Near-Memory_Computing_Past_Present_and_Future
6. Zelenskii A.A., Ivanovskii S.P., Ilyukhin Yu.V., Gribkov A.A. Programmirovaniye doverennoi pamyat'-tsentricheskoi sistemy upravleniya dvizheniem robototekhnicheskikh i mekhatronnykh sistem // Vestnik Moskovskogo aviationsionnogo instituta. 2022. T. 29. № 4. S. 197-210.
7. Juurlink B., Meenderinck C. Amdahl's law for predicting the future of multicores considered harmful // ACM SIGARCH Computer Architecture News, 40 (2012), 2. pp. 1-9.
8. Knut D.E. Iskusstvo programmirovaniya, tom 3. Sortirovka i poisk. M.: OOO "I.D. Vil'yams", 2018. 832 s.
9. Time Complexities of all Sorting Algorithms. Geeks for Geeks, 2023. URL: <https://www.geeksforgeeks.org/time-complexities-of-all-sorting-algorithms/>
10. Heap Algorithms. Massachusetts Institute of Technology, 2010. URL: <https://courses.csail.mit.edu/6.006/fall10/handouts/recitation10-8.pdf>
11. Binary Insertion Sort. Geeks for Geeks, 2023. URL: <https://www.geeksforgeeks.org/binary-insertion-sort/>

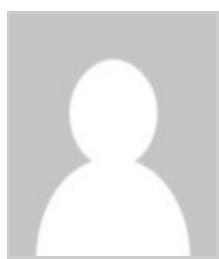
Fine-tuning neural networks for the features of a dataset in the speaker verification task using transfer learning

Trofimova Varvara Sergeevna

Student; Department of Integrated Information Security of Electronic Computing Systems; Tomsk State University of Control Systems and Radioelectronics

146 Krasnoarmeyskaya str., room 509, Tomsk region, 634045, Russia

✉ varvara.trofimova.01@mail.ru

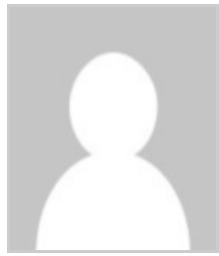


Karshieva Polina Konstantinovna

Student; Department of Integrated Information Security of Electronic Computing Systems; Tomsk State University of Control Systems and Radioelectronics

146 Krasnoarmeyskaya str., room 509, Tomsk region, 634045, Russia

✉ polinakarshieva1@gmail.com



Rakhmanenko Ivan Andreevich

PhD in Technical Science

Associate Professor; Department of Information Systems Security, Tomsk State University of Control Systems and Radioelectronics

146 Krasnoarmeyskaya str., room 509, Tomsk region, 634045, Russia

✉ ria@fb.tusur.ru

Abstract. The subject of this study is neural networks, trained using transfer learning methods tailored to the specific characteristics of the dataset. The object of the study is machine learning methods used for solving speaker verification tasks. The aim of the research is to improve the efficiency of neural networks in the task of speaker verification. In this work, three datasets in different languages were prepared for the fine-tuning process: English, Russian, and Chinese. Additionally, an experimental study was conducted using modern pre-trained models ResNetSE34L and ResNetSE34V2, aimed at enhancing the efficiency of neural networks in text-independent speaker verification. The research methodology includes assessing the effectiveness of fine-tuning neural networks to the characteristics of the dataset in the speaker verification task, based on the equal error rate (EER) of Type I and Type II errors. A series of experiments were also conducted, during which parameters were varied, and layer freezing techniques were applied. The maximum reduction in the equal error rate (EER) when using the English dataset was achieved by adjusting the number of epochs and the learning rate, reducing the error by 50%. Similar parameter adjustments with the Russian dataset reduced the error by 63.64%. When fine-tuning with the Chinese dataset, the lowest error rate was achieved in the experiment that involved freezing the fully connected layer, modifying the learning rate, and changing the optimizer—resulting in a 16.04% error reduction. The obtained results can be used in the design and development of speaker verification systems and for educational purposes. It was also concluded that transfer learning is effective for fine-tuning neural networks to the specific characteristics of a dataset, as a significant reduction in EER was achieved in the majority of experiments, indicating improved speaker recognition accuracy.

Keywords: deep learning, neural networks, speech processing, feature extraction, speaker recognition, speaker verification, dataset, fine-tuning, transfer learning, pattern recognition

References (transliterated)

1. Gassiev D. O., Sakharov V. A., Ermolaeva V. V. Golosovaya autentifikatsiya // Tendentsii razvitiya nauki i obrazovaniya. 2019. № 56(2). S. 22-24.
2. GOST R 58668.11-2019 (ISO/MEK 19794-13:2018). Informatsionnye tekhnologii. Biometriya. Formaty obmena biometricheskimi dannymi. Chast' 11. Dannye golosa. Moskva: Standartinform, 2019. 28 s.
3. Devyatkov V.V. Sistemy iskusstvennogo intellekta. Gl. red. I.B. Fedorov. – M.: Izd-vo MG TU im. N.E. Baumana, 2001. 352 s.
4. Galushkin A.I. Neironnye seti. Osnovy teorii. – M.: Goryachaya liniya – Telekom, 2012. 496 s.
5. Suzuki K. Artificial Neural Networks: Architectures and Applications. – Publisher: InTech, 2013. 256 p.
6. Evsyukov M.V., Putyato M.M., Makaryan A.S., Nemchinova V.O. Metody zashchity v sovremennykh sistemakh golosovoi autentifikatsii // Prikaziiskii zhurnal: upravlenie i vysokie tekhnologii. 2020. № 3(59). S 84-92.
7. Nagrani A., Chung J.S., Zisserman A. VoxCeleb: A large-scale speaker identification dataset // arXiv:1706.08612. 2018. URL: <https://arxiv.org/pdf/1706.08612.pdf> (data obrashcheniya: 15.07.2024).
8. Hinton G. E., Srivastava X., Krizhevsky A., Sutskever I., Salakhutdinov R. R. Improving neural networks by preventing co-adaptation of feature detectors // arXiv:1207.0580. 2012. URL: <https://arxiv.org/pdf/1207.0580.pdf> (data obrashcheniya: 22.07.2024).

9. Konev A.A. Model' i algoritmy analiza i segmentatsii rechevogo signala: avtoref. dis. na soisk. uchenoi step. kand. tekhn. nauk. Tomsk, 2007. 150 s.
10. Qin X., Bu H., Li M. HI-MIA: A Far-field Text-Dependent Speaker Verification Database and the Baselines // IEEE International Conference on Acoustics, Speech, and Signal Processing, 2020. Pp. 7609-7613.
11. Rakhmanenko I.A., Shelupanov A.A., Kostyuchenko E.Yu. Avtomaticheskaya verifikatsiya diktora po proizvol'noi fraze s primeneniem svertochnykh glubokikh setei doveriya // Komp'yuternaya optika. 2020. № 44(4). S. 596-605.

The evolution of the Semantic Web technologies: problems and prospects

Lukichev Ruslan Vladimirovich □

PhD in Philosophy

Master's degree; Faculty of Software Engineering and Computer Engineering; ITMO University
Director of Development; LLC 'First Festival Company'

198332, Russia, Saint Petersburg, Leninsky ave., 100k3, block 113

✉ ruslanlukichev@gmail.com

Abstract. The article is devoted to the consideration of key Semantic Web technologies, the analysis of their features, problematic aspects and growth points, which seems especially relevant in the context of import substitution and improving national information security. Special attention is paid to RDF graphs, which are based on an ontology-oriented approach, as well as the OWL language as the main tool for organizing machine-readable data structures with complex relationships between entities, a hierarchy of classes and properties. Attention is also paid to the limitations associated with the security of semantic databases, the need for their simplification, standardization and development of specialized software that meets usability criteria are analyzed. In addition, the prospects for further improvement of these technologies in the context of the Internet of Things and artificial intelligence are outlined. The article uses a comprehensive methodological framework, which implies the use of mainly general scientific methods, in particular, systematic and analytical. The article summarizes and analyzes current developments related to the Semantic Web technologies, which made it possible to identify a number of problems that need to be solved. First of all, the tools available today often have a high entry threshold, are characterized by an excessively complex, featureless interface without functions of complementary prompts and query visualization. Moreover, the Semantic Web languages need standardization and the introduction of a common protocol in order to simplify the process of working with multiformat data aggregated from different sources. Other important issues are ensuring the reliability and relevance of information, its integrity and confidentiality, as well as the contextual conditionality of logical conclusions and compliance with user requests. Among the key prospects is the creation of an intelligent autonomous environment in which devices can freely exchange data and interact with each other at the semantic level in order to provide high-quality personalized services.

The provisions of the article can be taken as a basis for the development of domestic systems for structuring and describing data available for machine processing, as well as specialized lecture courses in higher education institutions.

Keywords: SPARQL, OWL, RDFS, RDF, semantic web of things, data models, graph databases, ontologies, semantic web, XML

References (transliterated)

1. Berners-Lee, T. Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web. – NY: HarperCollins Publishers, 2000. – 246 pp.
2. Shpolyanskaya I. Yu., Seredkina T. A. Tekhnologii Semantic Web v organizatsii podderzhki onlain obucheniya. // Sistemnyi analiz v proektirovaniii i upravlenii. – Vyp. XXIV, № 3, 2020. – S. 343-350. doi:10.18720/SPBPU/2/id20-231
3. Ivashchuk Yu. S., Orlyanskaya N. P., Teshev V. A. Razrabotka bazy znanii investitsionnoi deyatel'nosti na osnove ontologicheskogo modelirovaniya. // Vestnik Adygeiskogo gosudarstvennogo universiteta. – Seriya 5: Ekonomika, № 2 (320), 2023. – S. 90-98. doi:10.53598/2410-3683-2023-2-320-90-98
4. Petrov I. D., Terekhova Yu. V. Rasshirenie logicheskoi modeli predmetnoi oblasti arkheologicheskikh issledovanii s ispol'zovaniem ontologii. // Uspekhi v khimii i khimicheskoi tekhnologii. – Vyp. 34, № 6 (229), 2020. – S. 133-135.
5. Smirnov A. V., Teslya N. N. Ontologo-orientirovannaya geoanalitika dlya opredeleniya mest kontsentratsii DTP na uchastkakh ulichno-dorozhnoi seti. // Trudy Kol'skogo nauchnogo tsentra RAN. Seriya: Tekhnicheskie nauki. – Vyp. 14, № 7, 2023. – S. 79-85. doi:10.37614/2949-1215.2023.14.7.008
6. Antoniou, G. Semanticeskii veb / G.Antoniou, P. Gros, F. van Harmelen, R. Khoekstra; per. s angl. T. Shul'ga. – M.: DMK Press, 2016. – 240 s.
7. Hogan, A. The Semantic Web: Two decades on. // Semantic Web. – Vol. 11, pp. 169-185, 2020. – doi:10.3233/SW-190387, 2020.
8. Hassan, B. Towards Semantic Web: Challenges and Needs. // International Journal Of Engineering And Computer Science. – Vol. 4, issue 10, Oct 2015. – pp. 14585-14588. doi:10.48550/arXiv.2105.02708
9. Turova I. A., Postanogov I. S. Razrabotka intellektual'nogo redaktora SPARQL-zaprosov. // Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Informatsionnye tekhnologii. – Vyp. 19, № 4, 2021. – S. 85-95.
10. Jat, A. Semantic web technologies: challenges and applications. // Journal of Critical Reviews. – Vol. 7, Issue 17, 2020. – pp. 3388-3390. doi:10.31838/jcr.07.17.417.
11. Chudinov P. Yu., Babenko L. K., Rogozov Yu. I. Analiz problem zashchity informatsii v semanticeskikh setyakh. // Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskie nauki. – № 5 (229), 2022. – S. 37-47.
12. F. Z. Amara, M. Hemam, M. Djezzar and M. Maimor. Semantic Web and Internet of Things: Challenges, Applications and Perspectives. // Journal of ICT Standardization. – Vol. 10. No. 2, 2022. – pp. 261-291. doi: 10.13052/jicts2245-800X.1029

On clarifying the concept of "power of attorney" of artificial intelligence systems

Tikhanychev Oleg Vasilyevich 

PhD in Technical Science

Deputy Head of the Department of Advanced Development Management, "Technoserv Group"

13 Yunost str., Moscow, 111395, Russia

 to.technoserv@gmail.com

Abstract. The subject of the study is the concept of "power of attorney" of artificial

intelligence controlling robotic means of varying degrees of autonomy. The relevance of the choice of the subject of research as the principles of the use of robotic systems for various purposes, including in a group, and the object of research, which are algorithmic problems arising in the implementation of group action algorithms, is determined by the existing contradiction between the need for joint use of robotic systems, primarily autonomous, and the complexity of the software implementation of this requirement. The implementation of robotics generates certain technological problems related to the efficiency and safety of algorithmic support of autonomous and controlled robotic systems. The manifestation of such problems may be application errors that reduce the effectiveness of joint actions. In robotics, the main potential reason for the appearance of such a situation is the insufficient effectiveness of existing group application control algorithms, determined by the low level of problem study. The article formulates a list of typical situations that determine the use of autonomous and controlled robots in a group with a leader. On the basis of the proposed classification, possible algorithms for controlling movement in the group are analyzed: both calculations for target maneuvering and for ensuring mutual security. The main situations concerning the types of maneuver are formulated and the mathematical apparatus for their calculations is described. Based on the overview analysis of typical algorithms for controlling movement in space, the formulation of the scientific problem of solving the problem of developing group algorithms and mathematical methods is synthesized.

Keywords: mathematical security, safety behavior management, morality of robotic systems, ethics of robotic systems, safe behavior, control of behavior algorithms, trustworthy artificial intelligence, control software, artificial intelligence, robotic system

References (transliterated)

1. Simulin A. A. i dr. Nekotorye aspeky ispol'zovaniya robototekhniki v voennom dele // Sborniki konferentsii NITs Sotsiosfera. 2015. № 27. S. 67-71.
2. Chirov D.S., Novak K.V. Perspektivnye napravleniya razvitiya robototekhnicheskikh kompleksov spetsial'nogo naznacheniya // Voprosy bezopasnosti. 2018. № 2. S. 50-59. DOI: 10.25136/2409-7543.2018.2.22737.
3. Khripunov S.P., Blagodaryashchev I.V., Chirov D.S. Voennaya robototekhnika: sovremennye trendy i vektory razvitiya // Trendy i upravlenie. 2015. № 4. S. 410-422.
4. Pflimlin É. Drones et robots: La guerre des futurs. France: Levallois-Perret, 2017.
5. Roosevelt, Ann. Ar my Directs Cuts, Adjustments, To FCS. Defense Daily, 2017.
6. Hamilton T How AI will Alter Multi-Domain Warfare. Future Combat Air & Space Capabilities Summit, 2023. No. 4 URL: <https://www.aerosociety.com/events-calendar/raes-future-combat-air-and-space-capabilities-summit>
7. Tikhanychev O. Influence of the Problem of Safety Control of Heuristic Algorithms on the Development of Robotics. In: Shamtsyan, M., Pasetti, M., Beskopylny, A. (eds) Robotics, Machinery and Engineering Technology for Precision Agriculture. Smart Innovation, Systems and Technologies 2022, No. 247. Singapore: Springer. https://doi.org/10.1007/978-981-16-3844-2_31.
8. Beard J Autonomous weapons and human responsibilities, Georgetown Journal of International Law 2014. No. 45, P. 617-681.
9. Tikhanychev O The Control System of Heuristic Algorithms as a Prototype of the "Morality" of Autonomous Robots. II International Scientific Forum on Sustainable Development and Innovation. WFSDI-2023. URL: https://doi.org/10.1007/978-981-16-3844-2_31.

10. Ćwiąkała P. Testing Procedure of Unmanned Aerial Vehicles (UAVs) Trajectory in Automatic Missions. *Appl. Sci.* 2019, No. 9, P/ 3488. URL: <https://doi.org/10.3390/app9173488>.
11. Johnson D. Computer Systems: Moral entities but not moral agents. In: *Ethics and Information Technology*. 2016, No. 8, R. 195-204. URL: <https://doi.org/10.1007/s10676-006-9111>.
12. Schuller A. At the Crossroads of Control: The Intersection of Artificial Intelligence in Autonomous Weapon Systems with International Humanitarian Law, *Harvard National Security Journal*. 2017, No. 8, P. 379-425.
13. Ukhobotov V.I., Izmest'ev I.V. Ob odnoi zadache presledovaniya pri nalichii soprotivleniya sredy // *Vestnik YuUrGU. Seriya: Matematika. Mekhanika. Fizika*. 2016. № 2. C. 62-66. URL: <https://doi.org/10.14529/mmp160208>.
14. Dubanov A.A. Modelirovanie traektorii presledovatelya v prostranstve pri metode parallel'nogo sblizheniya // *Programmnye sistemy i vychislitel'nye metody*. 2021. № 2. C. 1-10. URL: <https://doi.org/10.7256/2454-0714.2021.2.36014>
15. Tikhanychev O. Self-Check System of Heuristic Algorithms as a "New Moral" of Intelligent Systems AIP Conference Proceedings. 2023, No. 2700, 040028. URL: <https://doi.org/10.1063/5.0124956>.

Analysis of Idioms in Neural Machine Translation: A Data Set

Skacheva Nina Vasil'evna 

Senior Lecturer; Department of Linguistics, Theory and Practice of Translation; Siberian State University of Science and Technology named after Academician MF. Reshetnev

660037, Russia, Krasnoyarsky krai, Krasnoyarsk, Krasnoyarsk worker, 31

 Sollo_sk@mail.ru

Abstract. There has been a debate in various circles of the public for decades about whether a "machine can replace a person." This also applies to the field of translation. And so far, some are arguing, others are "making a dream come true." Therefore, now more and more research is aimed at improving machine translation systems (hereinafter MP). To understand the advantages and disadvantages of MP systems, it is necessary, first of all, to understand their algorithms. At the moment, the main open problem of neural machine translation (NMP) is the translation of idiomatic expressions. The meaning of such expressions does not consist of the meanings of their constituent words, and NMT models tend to translate them literally, which leads to confusing and meaningless translations. The research of idioms in the NMP is limited and difficult due to the lack of automatic methods. Therefore, despite the fact that modern NMP systems generate increasingly high-quality translations, the translation of idioms remains one of the unsolved tasks in this area. This is due to the fact that idioms, as a category of verbose expressions, represent an interesting linguistic phenomenon when the general meaning of an expression cannot be made up of the meanings of its parts. The first important problem is the lack of special data sets for learning and evaluating the translation of idioms. In this paper, we solve this problem by creating the first large-scale dataset for translating idioms. This data set is automatically extracted from the German translation corpus and includes a target set in which all sentences contain idioms, and a regular training corpus in which sentences containing idioms are marked. We have released this dataset and are using it to conduct preliminary experiments on NMP as a first step towards improving the translation of idioms.

Keywords: Russian, German, Neural Machine Translation, machine translation, bilingual corpora, idioms, multiword expression, language pairs, systems, Data Set

References (transliterated)

1. Koehn P. Europarl: A Parallel Corpus for Statistical Machine Translation // School of Informatics University of Edinburgh, Scotland. 2005. P. 79-86.
2. Natsional'nyi korpus russkogo yazyka. URL: <https://ruscorpora.ru/search?search=CgkyBwgFEgNlbmcwAQ%3D%3D> (data obrashcheniya 04.03.2024)
3. Brown P. F., Pietra S. A. D., Pietra V. J. D., Mercer R. L. The mathematics of statistical machine translation. Computational Linguistics. 1993. 19(2), p. 263-313.
4. Philipp Koehn, Franz J. Och, and Daniel Marcu.. Statistical Phrase-Based Translation. In Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics. 2003. p. 127-133. URL: <https://aclanthology.org/N03-1017.pdf> (data obrashcheniya 05.03.2024)
5. Gudkov V. Yu., Gudkova E.F. N-grammy v lingvistike // Vestnik ChelGU. 2011. № 24.
6. Lingvisticheskii korpus dannykh. URL: <https://catalog.ldc.upenn.edu/byyear> (data obrashcheniya 05.03.2024)
7. Zhdanov A. E., DorosinskiiL.G. Golosovoi zamok // Ural Radio Engineering Journal. 2017. Vol. 1, No. 1. P. 80-90.
8. Daniel Torregrosa, Nirvana Pasricha, Bharathi Raja Chakravarthi, Maraim Masoud, Mihael Arcan. Leveraging Rule-Based Machine Translation Knowledge for Under-Resourced Neural Machine Translation Models // Proceedings of MT Summit XVII, Dublin, 2019, volume 2. URL: <https://aclanthology.org/W19-6725.pdf> (data obrashcheniya: 19.08.2024).
9. Peter T. Systran as a multilingual machine translation system // Overcoming the language barrier, 3-6 May 1977, Vol. 1. URL: <https://www.mt-archive.net/70/CEC-1977-Toma.pdf> (data obrashcheniya: 19.08.2024).
10. Salton G., Ross R., and Kelleher J. (2014). An empirical study of the impact of idioms on phrase based statistical machine translation of english to brazilian-portuguese // In Proceedings of the 3rd Workshop on Hybrid Approaches to Machine Translation (HyTra). 2014. P. 36-41.
11. Luong T., Pham H., Manning C. D. (2015). Effective approaches to attention-based neural machine translation // In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. 2015. P. 1412-1421.
12. Koehn P., Hoang H., Birch A., Callison-Burch C., Federico M., Bertoldi N., Cowan B., Shen W., Moran C., Zens R. Moses: Open source toolkit for statistical machine translation // In Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions. 2007. P. 177-180.
13. Papineni K., Roukos S., Ward T., and Zhu W.J. Bleu: a method for automatic evaluation of machine translation. In Proceedings of 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania, USA. 2002. P. 311-318.
14. Dyer C., Chahuneau V., and Smith N. A. A simple, fast, and effective reparameterization of ibm model 2 // In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies Atlanta, Georgia, June. 2013 p. 644-646.

Numerical methods for finding the roots of polynomials with real and complex coefficients

Sklyar Alexander Yakovlevich

PhD in Technical Science

Associate Professor; Department of Applied Mathematics; Russian Technological University (MREA)

119602, Russia, Moscow, Vernadsky, 78

✉ askliar@mail.ru



Abstract. The subject of the article is the consideration and analysis of a set of algorithms for numerically finding the roots of polynomials, primarily complex ones based on methods for searching for an approximate decomposition of the initial polynomials into multipliers. If the numerical finding of real roots usually does not cause difficulties, then a number of difficulties arise with finding complex roots. This article proposes a set of algorithms for sequentially finding multiple roots of polynomials with real roots, then real roots by highlighting intervals that potentially contain roots and obviously do not contain them, and then complex roots of polynomials. To find complex roots, an iterative approximation of the original polynomial by the product of a trinomial by a polynomial of a lesser degree is used, followed by the use of the tangent method in the complex domain in the vicinity of the roots of the resulting trinomial. To find the roots of a polynomial with complex coefficients, we propose a solution to an equivalent problem with real coefficients. The implementation of the tasks is carried out by step-by-step application of a set of algorithms. After each stage, a group of roots is allocated and the same problem is solved for a polynomial of lesser degree. The sequence of the proposed algorithms makes it possible to find all the real and complex roots of the polynomial. To find the roots of a polynomial with real coefficients, an algorithm is constructed that includes the following main steps: determining multiple roots with a corresponding decrease in the degree of the polynomial; allocating a range of roots; finding intervals that are guaranteed to contain roots and finding them, after their allocation, it remains to find only pairs of complex conjugate roots; iterative construction of trinomials that serve as an estimate of the values of such pairs with minimal the accuracy sufficient for their localization; the actual search for roots in the complex domain by the tangent method. The computational complexity of the proposed algorithms is polynomial and does not exceed the cube of the degree of the polynomial, which makes it possible to obtain a solution for almost any polynomials arising in real problems. The field of application, in addition to the polynomial equations themselves, is the problems of optimization, differential equations and optimal control that can be reduced to them.

Keywords: recursive algorithms, roots of polynomials, conjugate complex roots, algebraic equation, numerical algorithms, numerical methods, iterative methods, root finding, polynomials, root localization

References (transliterated)

1. Kurosh A.G. Kurs vysshei algebry. Moskva: Nauka, 1968. S. 431.
2. Samarskii A. A., Gulin A. V. Chislennye metody. Moskva: Nauka, 1989. S. 432.
3. Stillvell D. Matematika i ee istoriya. – Moskva-Izhevsk: Institut komp'yuternykh issledovanii, 2004. S. 530.
4. Tynkevich M. A., Pimonov A. G. Vvedenie v chislennyi analiz. Kemerovo: KuzGTU. 2017. S. 176.

5. Ch'e En Un, Shein A.B. Metod nakhzhdeniya kornei mnogochlenov. I // Informatika i sistemy upravleniya. 2012. № 4(34). S. 88-96.
6. Ch'e En Un, Shein A.B. Metod nakhzhdeniya kornei mnogochlenov. II // Informatika i sistemy upravleniya. 2013. №1(35). S. 108-118.
7. Ch'e En Un, Shein A.B. Metod nakhzhdeniya kornei mnogochlenov. III // Informatika i sistemy upravleniya. 2013. №3 (37). S. 110-122.
8. Kutishchev G.P. Reshenie algebraicheskikh uravnenii proizvol'noi stepeni: Teoriya, metody, algoritmy. URSS. 2015. 232 s.
9. Simon Telen. Polynomial Equations: Theory and Practice. Michal Kočvara; Bernard Mourrain; Cordian Riener. Polynomial Optimization, Moments, and Applications, Springer, pp. 215-240.
10. B. Mourrain and J. P. Pavone. Subdivision methods for solving polynomial equations. Journal of Symbolic Computation, 44(3), 292-306, 2009.
11. Berthomieu, C. Eder, and M. Safey El Din. msolve: A library for solving polynomial systems. In Proceedings of the 2021 on International Symposium on Symbolic and Algebraic Computation, pages 51-58, 2021.
12. Statsenko I. V. Issledovanie skorosti skhodimosti odnogo obobshchennogo n'yutonovskogo metoda i klassicheskogo metoda n'yutona v protsedure utochneniya kornei mnogochlena. // Tochnaya nauka. 2020. №78. S. 2-9.