

Том **74**
Выпуск **3**



ISSN 2079-0279

2024

ТРУДЫ

**ИНСТИТУТА СИСТЕМНОГО АНАЛИЗА
РОССИЙСКОЙ АКАДЕМИИ НАУК**

<http://www.ISA.ru>

Методы и модели системного анализа

Динамические системы

Компьютерный анализ текстов

Информатика сообществ и формирование социальных сетей



Труды Института системного анализа Российской академии наук (ИСА РАН)

<http://www.isa.ru/proceedings/>

Журнал основан в 2002 году
Журнал выходит ежеквартально

Учредитель

Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление»
Российской академии наук» (ФИЦ ИУ РАН)
119333, г. Москва, ул. Вавилова, д. 44, корп. 2, тел.: 8 (499) 135-62-60, e-mail: frccsc@frccsc.ru, <http://www.frccsc.ru/>

Главный редактор

Попков Ю.С., академик РАН,
119312, г. Москва, проспект 60-летия Октября, д. 9
тел. 8 (499) 135-42-22, e-mail: popkov@isa.ru

Редакционный совет

Попков Юрий Соломонович, академик РАН
(председатель Редакционного совета), г. Москва, Россия
Виссен Леонард (Leonardus J.G. van Wissen), проф., г. Амстердам, Нидерланды
Константинов Игорь Сергеевич, д.т.н., г. Белгород, Россия
Соловьев Александр Владимирович, д.т.н.,
(заместитель главного редактора) г. Москва, Россия
Шмудьян Борис Лейбович, проф., г. Тель-Авив, Израиль
Щербов Сергей Яковлевич, проф., г. Вена, Австрия

Редакционная коллегия

Алескеров Фуад Тагиевич, д.т.н., г. Москва, Россия
Ашимов Абдыкаппар Ашимович, д.т.н., г. Алматы, Казахстан
Белов Сергей Павлович, д.т.н., г. Белгород, Россия
Булычев Александр Викторович, к.т.н. (ответственный секретарь), г. Москва, Россия
Бурков Владимир Николаевич, д.т.н., г. Москва, Россия
Войтишек Антон Вацлавович, д. ф.-м.н.,
г. Новосибирск, Россия
Геловани Виктор Арчилович, академик РАН,
г. Москва, Россия
Ильин Александр Владимирович, чл.-корр. РАН,
г. Москва, Россия

Крутько Вячеслав Николаевич, д.т.н., г. Москва, Россия
Лексин Владимир Николаевич, д.э.н., г. Москва, Россия
Лившиц Вениамин Наумович, д.э.н., г. Москва, Россия
Магницкий Николай Александрович, д.ф.-м.н., г. Москва, Россия
Мижидон Арсалан Дугарович, д.т.н., г. Улан-Удэ, Россия
Орлова Елена Роальдовна, д.э.н., г. Москва, Россия
Петровский Алексей Борисович, д.т.н., г. Москва, Россия
Смирнов Александр Викторович, д.т.н., г. Санкт-Петербург, Россия
Тищенко Виктор Иванович, к.филос.н., г. Москва, Россия
Швецов Александр Николаевич, д.э.н., г. Москва, Россия

Редакция

Шибаева Л.С. (заведующая редакцией)

Адрес редакции

117312, г. Москва, проспект 60-летия Октября, д. 9,
E-mail: tisa@isa.ru, тел. 8 (499) 135-24-33, 8 (926) 490-87-87

Научный журнал «Труды Института системного анализа Российской академии наук (ИСА РАН)» публикует материалы по широкому кругу фундаментальных проблем развития методологии системного анализа и ее приложений к решению разнообразных задач в сфере науки и практики. Журнал предназначен для научных сотрудников, инженеров и исследователей, работающих в рамках этой проблематики, а также для политиков, сотрудников государственных и муниципальных органов, специалистов предприятий и представителей общественных организаций. Правила подготовки статей для публикации в журнале, а также правила их рецензирования приведены на сайте журнала.

Журнал зарегистрирован в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации средства массовой информации ПИ № ФС77-65718 от 20 мая 2016 г.
Подписной индекс журнала – 36089 в электронном каталоге ГК «УРАЛ-ПРЕСС» (<http://ural-press.ru/>).
Журнал включен в перечень ведущих рецензируемых научных изданий, рекомендованных ВАК для публикации основных научных результатов диссертаций на соискание ученых степеней доктора и кандидата наук.

© Федеральное государственное учреждение «Федеральный исследовательский центр
«Информатика и управление» Российской академии наук»

Содержание

Методы и модели системного анализа	3
Моделирование оценок синергических эффектов организации деятельности в системе естественно-монопольных характеристик (Н.И. Белоусова, С.П. Бушанский).....	3
Действие закона диффузии внешней среды организаций как социально-экономических систем (И.В. Найденов).....	13
Задача генерации выходных форм документов информационных систем (А.В. Соловьев)	22
Динамические системы	29
Решение дифференциально-алгебраических систем уравнений с помощью аппроксимации Паде матричной экспоненты (Ю.А. Бурцев).....	29
Компьютерный анализ текстов.....	39
Контекстно-независимый метод быстрой детекции текста для распознавания номеров телефонов (А.В. Гайер).....	39
Информатика сообществ и формирование социальных сетей	48
Данные как вызов (В.И. Тищенко).....	48
Исследование распределения расстояний между графами с упорядоченными вершинами (А.А. Рогов, Н.Д. Москин, Р.В. Воронов, К.А. Кулаков).....	58
Распознавание образов.....	67
Оценка обобщающей способности алгоритма вложенных контуров при анализе маммограмм (И.А. Егошин)	67
Управление рисками и безопасностью	78
Enhancing Kubernetes Security: The Crucial Role of DevSecOps (Ghadeer Darwesh, Jaafar Hammoud, A.A. Vorobeva)	78
Поиск уязвимостей в смарт-контрактах на основе машинного обучения (В.С. Белоус, И.А. Тарханов).....	89

PROCEEDINGS OF THE INSTITUTE FOR SYSTEMS
ANALYSIS OF RUSSIAN ACADEMY OF SCIENCES

IN THIS ISSUE

Methods and Models of Systems Analysis	3
Modeling assessments of synergy effects under organizational activity in the system of natural monopoly characteristics (<i>N.I. Belousova, S.P. Bushansky</i>)	3
The action of the law of diffusion of the external environment of organizations as socio-economic systems (<i>I.V. Naydenov</i>)	13
The problem of generating output forms of information systems documents (<i>A.V. Solovyev</i>)	22
Dynamical Systems	29
Solving Differential-Algebraic Equation Systems with Pade Approximation of Matrix Exponent (<i>Y.A. Burtsev</i>)	29
Text Mining	39
Context-independent fast text detection method for recognizing phone numbers (<i>A.V. Gayer</i>)	39
Community Informatics and formation of social networks	48
Data as a challenge (<i>V. I. Tishchenko</i>)	48
Research of the distribution of distances between graphs with ordered vertices (<i>A.A. Rogov, N.D. Moskin, R.V. Voronov, K.A. Kulakov</i>)	58
Pattern Recognition	67
Evaluation of an generalization ability of the nested contours algorithm in the mammograms analysis (<i>I.A. Egoshin</i>)	67
Risk-Management and Security	78
Enhancing Kubernetes Security: The Crucial Role of DevSecOps (<i>Ghadeer Darwesh, Jaafar Hammoud, A.A. Vorobeva</i>)	78
Search for vulnerabilities in smart contracts based on machine learning (<i>V.S Belous, I.A. Tarkhanov</i>)	89

Методы и модели системного анализа

Моделирование оценок синергических эффектов организации деятельности в системе естественно-монопольных характеристик

Н.И. Белоусова^I, С.П. Бушанский^{II}

^I Федеральный исследовательский центр «Информатика и управление»
Российской академии наук», г. Москва, Россия

^{II} Федеральное государственное бюджетное учреждение науки
«Центральный экономико-математический институт Российской
академии наук», г. Москва, Россия

Аннотация. Исследуются теоретико-прикладные подходы к оценке синергических эффектов организации деятельности на основе моделей и методов современной теории естественной монополии для решения задач анализа и стратегического проектирования инфраструктурных подсистем. Проанализированы взаимосвязи адекватного теоретического обоснования способов эффективной организации отраслевых рынков и теоретико-прикладных подходов к оценке естественно-монопольной синергии. В рамках моделей нормативной идентификации сформированы подходы к оценке синергических эффектов с использованием ключевых технологических детерминант естественно-монопольных индикаторов деятельности. Определены направления анализа факторов, влияющих на синергические эффекты организации деятельности в рамках теоретических представлений поведенческой идентификации естественной монополии с расширением набора используемых естественно-монопольных характеристик. Выделяются меры государственного регулирования, связанные с формированием рыночного экономического поведения естественного монополиста, прежде всего, ценового поведения, определяющие возможности корректировки оценок синергических эффектов организации деятельности.

Ключевые слова: стратегическое проектирование, теория естественной монополии, эффективная рыночная организация, синергические эффекты организации деятельности, нормативная и поведенческая идентификация, расщепление/ объединение инфраструктурных подсистем, сетевая специфика, система естественно-монопольных характеристик, государственное регулирование, ценовое поведение.

DOI: 10.14357/20790279240301 **EDN:** DRBVPJ

Введение

Широкое распространение и использование понятия синергии в экономическом и социально-

экономическом анализе в связи с рассмотрением инвестиционных, организационных и других мероприятий в различных сферах деятельности (но при этом зачастую с нечеткой и неоднозначной

трактовкой) за последние годы стало достаточно общим местом.

Применительно к сферам естественных монополий, инфраструктурным подсистемам с естественно-монопольным компонентом, при проведении структурных преобразований в значительной мере целесообразна ориентация на модели и методы, формируемые в современной теории естественной монополии [1-6]. Использование их позволяет дать конструктивную постановку и обеспечить выявление синергических эффектов организации деятельности при решении задач структурной реорганизации, анализа и выбора эффективных организационных мероприятий, определении эффективной структуры отрасли, способов эффективной организации отраслевых рынков, эффективных поведенческих моделей. При этом существенным образом учесть специфику естественно-монопольных инфраструктурных подсистем (прежде всего, сетевую).

1. Базовые теоретико-прикладные подходы к оценке синергических эффектов организации деятельности на основе моделей и методов нормативной идентификации

Исходя из представлений теории естественной монополии возможно формирование основополагающих подходов к оценке синергических эффектов – применительно к задаче выбора в процессе структурных преобразований наиболее эффективного способа проведения организационных мероприятий по расщеплению/объединению инфраструктурных фирм/предприятий и их подразделений.

Эффект естественно-монопольной синергии как оценка абсолютного снижения совокупных издержек, инициированных тем или иным организационным мероприятием, может быть определен (согласно установленным в теории естественной монополии предпосылкам, требованиям и ограничениям) при сравнении варианта организации деятельности в виде одной единственной фирмы (собственно естественной монополии) и вариантов организации с полностью или частично специализированным/диверсифицированным характером производства продукции и услуг.

Соответственно, базовая теоретико-прикладная трактовка синергических эффектов организации деятельности непосредственно увязывается с оценкой субаддитивности отраслевой функции совокупных издержек, многопродуктовой, в силу сложного характера деятельности и направленности моделирования ситуаций по проверке множе-

ства вариантов организационного расщепления/объединения.

Согласно теории, определяются специальные требования к виду функций, по которым может быть дана оценка совокупных издержек (гибкость, априорная непредсказуемость результатов оценки и др.) [1, 7-10]. Сетевая специфика дополнительно определяет необходимость использования при формировании оценок субаддитивности оптимальных сетевых технологий развития и транспортировки [6]. При формировании функциональных форм допускается возможность учета спектра существенных факторов оценки, включая гедонические и сетевые [11]. Включение гедонических факторов в многопродуктовую отраслевую функцию совокупных издержек определяет ситуацию, когда идентичные физические объемы выпусков могут определяться вариацией качественных характеристик и особенностями деятельности на сетях по сегментам отраслевого рынка.

Ориентация на формирование оценок субаддитивности при выборе вариантов (при заданных объемах выпуска – выявленных объемах спроса) позволяет обеспечить получение потенциально для общества в целом дополнительную экономию совокупных издержек, отвечающую эффективно-му способу организации производства услуг.

Исходя из теоретических моделей оценки с использованием понятия субаддитивности, синергический эффект организации деятельности может быть определен как разница совокупных издержек между их величиной для полного объема инфраструктурных услуг (в ситуации целостной организации $C(\bar{y})$) и суммой аналогичных издержек при потенциальном гипотетическом расщеплении, отвечающем удовлетворению выявленного суммарного объема спроса

$$\Delta C = C(\bar{y}) - \sum_{i=1}^k C(\bar{y}^i)$$

для любых допустимых наборов выпусков $\bar{y}^1, \bar{y}^2, \dots, \bar{y}^k$, таких, что $\sum_{i=1}^k \bar{y}^i = \bar{y}$, $\bar{y}^i \geq 0$, при существовании, как минимум, двух положительных значений $\bar{y}^{i1}, \bar{y}^{i2}$.

Следует учитывать, что при стратегическом проектировании и включении в рассмотрение изменений объемов и структуры спроса на различные виды инфраструктурных услуг, возможности конструирования вариантов расщеплений могут быть ограничены в силу неопределенности прогнозов развития рассматриваемой системы, необходимости преодоления проблемы информационной недостаточности и трудностей устранения пробелов по включению в анализ всего спектра допустимых вариантов структурных изменений [12].

Оценка субаддитивности на базе многопродуктовой функции совокупных издержек и, соответственно, синергических эффектов организации деятельности, предполагает проведение инструментальных процедур нормативной идентификации и тестирования с использованием ключевых технологических детерминант/естественно-монопольных индикаторов деятельности [1-6,13], прежде всего, экономии от структуры (SC) и средних приростных издержек (AIC).

Эти естественно-монопольные характеристики могут рассматриваться, собственно, как индикаторы естественно-монопольной синергии. Их уровень и взаимосвязанная динамика (учитываемая в системе необходимых и достаточных условий существования естественной монополии [1]) свидетельствуют о наличии или отсутствии эффектов синергии, допускают важные дополнительные аспекты представления и интерпретации.

Так, индикатор SC дает сравнительную оценку по издержкам гипотетических вариантов организации производства инфраструктурных услуг через построение ортогональных расщеплений, а не произвольного их набора. В множество рассматриваемых конструируемых и вычислимых способов включаются как варианты совместного производства всех видов продукции/услуг, так и полностью специализированные по одному или нескольким видам деятельности (при условии обеспечения суммарного одинакового объема спроса):

$$SC(\bar{y}_M) = \frac{\sum_i C(\bar{y}_i) - C(\bar{y}_M)}{C(\bar{y}_M)}.$$

Экономия по совокупным издержкам определяется для любого разбиения P подмножества M выпускаемых видов продукции/услуг, $M \subseteq N$, таких что $P = \{T_1, \dots, T_m\}$, где $\cup T_i = M$, $T_i \cap T_j = \emptyset$ для $i \neq j$, $T_i \neq \emptyset$, $m > 1$.

Оценка AIC связана с началом/прекращением выпуска какого-то одного вида продукции/услуг. В расчеты эффективности включается прирост издержек и (или) их экономия на единицу объема выпуска этого вида продукции. Если $X, Y, Z, \dots$ – объемы выпусков различных видов продукции/услуг, тогда при дополнительном производстве единицы продукта X , который ранее не выпускался, при определенной комбинации всех других видов продукции в объемах Y, Z , оценка этого индикатора дается по формуле:

$$AIC_x = [C(X, Y, Z, \dots) - C(0, Y, Z, \dots)] / X.$$

Согласно теории [1], достаточные условия субаддитивности могут быть определены через монотонное (не строгое) снижение индикатора средних приростных издержек AIC по каждо-

му продукту, а также снижение положительных значений индикатора экономии от структуры SC . Положительные значения SC для многопродуктовой функции издержек являются необходимыми условиями субаддитивности. При этом для однопродуктового случая индикатор $SC > 0$ полностью идентифицирует наличие субаддитивности.

Отметим, что для многопродуктового случая такой широко известный показатель, как экономия от масштаба S (оценивается как обратная величина к затратной эластичности или к сумме затратных эластичностей по каждому продукту для многопродуктового случая) не входит в систему необходимых и достаточных условий при тестировании субаддитивности издержек и является ограниченным по мощности инструментом – и в смысле нормативной идентификации естественной монополии, и собственно оценки естественно-монопольной синергии.

Некоторые результаты экспериментальных оценок естественно-монопольной синергии с использованием ключевых технологических детерминант/естественно-монопольных индикаторов деятельности представлены в [12].

Анализ показывает, что российская практика структурных преобразований в сферах естественных монополий лишь в определенной мере ориентирована на адаптацию имеющихся теоретических обоснований принимаемых управленческих решений. Соответственно, должным образом при формировании текущих и стратегических оценок инфраструктурного развития не нашли отражения теоретические обоснования способов эффективной организации отраслевых рынков и теоретико-прикладные подходы к анализу и оценке естественно-монопольной синергии, обеспечению их взаимосвязей на базе потенциала современной теории естественной монополии [14].

Инициация самой предлагаемой постановки задачи анализа и стратегической оценки естественно-монопольной синергии непосредственно связана с системой представлений эффективной рыночной организации отрасли. Применительно к естественным монополиям используются возможности рыночной организации одной единственной фирмы, определяемые особенностями отраслевой технологии, в сравнении с вариантами отраслевой структуры с полностью или частично диверсифицированным характером организации деятельности (по видам производимой продукции/услуг).

Выявление ситуации естественной монополии характеризует отраслевую структуру, которая обеспечивает (при выявленных объемах спроса) минимизацию (по числу предприятий, фирм-производителей) совокупных издержек и, соответ-

ственно, получение (потенциально для общества в целом) дополнительной экономии, отвечающей эффективному способу организации производства услуг при оптимальной отраслевой технологии.

Формирование множества исходных вариантов организационного расщепления/ объединения входящих в отрасль предприятий может рассматриваться как основа для оценки естественно-монопольной синергии. При этом оптимальность отраслевой технологии, технологического способа производства, экономически выражаемого через совокупные издержки и их минимизацию, одна из ключевых гипотез, заложенных в процедуры выбора. На практике, как правило, имеет место лишь субоптимальность, некоторое приближение к оптимальной технологии, включающей и организационный компонент. Предполагается, что организационные резервы, связанные использованием принятого технологического способа производства в отрасли и влияющие на уровень издержек, наилучшим образом задействованы.

Теоретические модели стоимостной оценки естественно-монопольной синергии в виде абсолютного снижения совокупных издержек, согласующиеся с подходами теории отраслевых рынков и организации промышленности [15], непосредственно развиваемые и конкретизируемые в современной теории естественной монополии [1-6], могут быть сформированы в результате проведения нормативной (технологической) идентификации и проверки многопродуктовой отраслевой функции совокупных издержек на субаддитивность. Прежде всего, такие подходы ориентированы на исследование характеристик деятельности инфраструктурных предприятий с потенциальным естественно-монопольным компонентом.

Соответственно, для условий выявленного суммарного объема спроса даются оценки разниц совокупных издержек между их величинами в ситуации целостной, с объединением на данном отраслевом рынке полного объема производимых услуг/ видов деятельности в рамках одной организации, и оценками аналогичных издержек при всех потенциально возможных и допустимых вариантах расщепления организации.

2. Направления анализа факторов, влияющих на синергические эффекты организации деятельности в системе теоретических представлений поведенческой идентификации естественной монополии

В методологическом плане расширение возможностей анализа и корректировки оценок си-

нергических эффектов организации деятельности, сформированных на базе процедур нормативной идентификации и, соответственно, выявления «технологической» естественной монополии, связано с привлечением теоретических представлений и модельных подходов, исследуемых в рамках поведенческого аспекта идентификации, соответствующих естественно-монопольных характеристик [1,3].

Центральная характеристика при поведенческой идентификации, применительно к выявленной «технологической» естественной монополии и получившей соответствующий статус, оценка устойчивости/неустойчивости в ценовом отношении. Устойчивость в ценовом отношении, с одной стороны, означает нецелесообразность для естественной монополии (как единственной фирмы, работающей на естественно-монопольном рынке) ни понизить цену (при заданном/ выявленном спросе) так, чтобы не понести убытка, ни повысить, чтобы не привлечь на данный сегмент рынка потенциальных конкурентов. С другой стороны, одновременно – это невозможность для фирм-новичков войти на естественно-монопольный сегмент рынка и работать на нем в безубыточном для себя режиме.

Другими словами, при формализованном выражении естественная монополия устойчива в ценовом отношении если при заданном спросе Q и использовании оптимальной технологии, отвечающей функции совокупных издержек C , существует по крайней мере один вектор цен p^{nm} , такой, что: при этих p^{nm} ценах естественный монополист работает в безубыточном (но и в бесприбыльном) режиме. Потенциальные фирмы-новички, способные выпускать на естественно-монопольном рынке K видов продукции/ услуг в объеме y_K^e из полного их набора N в объеме Q , производимого естественным монополистом, вынуждены назначать на свою продукцию только такие цены p_K^e , при которых их производство становится убыточным, т.е.

$$p_K^e y_K^e - C(y_K^e) < 0, \quad K \subseteq N, \quad p_K^e \leq p_K^{nm}, \\ y_K^e \leq Q^K, \quad y_K^e \neq Q(p^{nm}).$$

Обеспечение ценовой устойчивости принимаемых решений отвечает ситуации, когда участникам рынка нецелесообразно выходить из установленных границ ценовых режимов, принятой траектории эффективного ценообразования, в широком контексте – нарушать способность динамической системы сохранять движение по намеченной траектории (поддерживать намеченный режим функционирования, согласно заданному критерию), несмотря на возмущающие на нее возмущения.

Отклонения от режимов устойчивых цен (как целевой теоретической модели оценки социальной ориентации деятельности) являются существенными факторами снижения потенциальных социально-экономических результатов, заложенных при принятии решений о реструктуризации отраслевой структуры на основе оценок эффектов естественно-монопольной синергии. И результаты нормативной идентификации, проверки субаддитивности и выполнения условия существования естественной монополии могут быть «обнулены», выявленные эффекты естественно-монопольной синергии организации деятельности при осуществлении структурных мероприятий сведены на «нет».

Соответственно, в рамках государственного регулирования естественных монополий должны быть предусмотрены эффективные управленческие механизмы и модели, определяющие ценовое поведение участников в условиях естественной монополии и условия их формирования.

Другая гипотетическая проблемная ситуация, когда «технологическая» естественная монополия определена в рамках нормативной идентификации, и требуется государственное вмешательство в части обеспечения недискриминационного доступа к инфраструктуре (сетей различной природы – транспортных, энергетических, телекоммуникационных) – естественный монополист имеет доступ к некоторому звену сетевой инфраструктуры, который для других участников естественно-монопольного рынка представляет собой «узкое место», «бутылочное горлышко», перемещение по которому предполагает некоторую плату за доступ [16–18].

Так как отделение инфраструктуры от операций – одно из ключевых направлений структурных преобразований естественных монополий (согласно и отечественному и зарубежному опыту), этим во многом определяется приоритетность ценового регулирования именно в части оценки цен доступа к инфраструктуре. Соответственно, особое значение имеет ориентация (в пределах, охватываемых государственным регулированием) на адекватные (с позиций теории и возможностей формирования обоснованных теоретико-прикладных оценок) модели цен доступа к инфраструктуре во взаимосвязи с рыночными конкурентными средами [19], имеющие некоторую специфику по выбору приемлемых способов оценки, что явно недостаточно принимается в расчет в современных российских условиях.

При формировании цен доступа к сетевым инфраструктурным объектам, деятельность которых полностью или частично включается в сферы государственного регулирования естественных монополий (будь-то железнодорожная инфраструктура или

местная телефонная сеть) существенным образом следует принимать в расчет специфику ситуации, когда доминирующая компания, контролируя доступ к сети для других перевозчиков/операторов (в том числе, в системе *interconnection* взаимного доступа операторов к сетям связи [20, 21]), в значительной мере может инициировать определение условий конкурентного доступа и их стоимостной оценки как промежуточной продукции деятельности инфраструктурной подсистемы. В качестве промежуточных затрат эти оценки включаются как компоненты при формировании конечных цен (тарифов) на перевозки грузов и пассажиров, на услуги телефонной связи, влияют на их уровень и динамику.

В процессе структурных преобразований, применительно к инфраструктурным подсистемам железнодорожного транспорта, при доставке грузов и пассажиров выстраиваются некоторые технологические цепочки. В них осуществляется взаимодействие участников рынка, субъектов хозяйствования и определенным образом формируется их иерархия – в значительной мере исходя из управленческих механизмов, согласно которым устанавливаются ценовые режимы, как важнейшие условия эффективного коммерческого взаимодействия, и создается основа для формирования ценового поведения участников. При этом один из определяющих факторов действенности таких технологических цепочек, в условиях развития рынков операторов-перевозчиков и сохранения в собственности/управлении естественным монополистом сетевой инфраструктуры, обеспечение приемлемых, не препятствующих эффективной конкуренции (с учетом ее специфики применительно к сферам естественных монополий), условий недискриминационного доступа к инфраструктуре независимых перевозчиков.

Система ценовых ограничений, связанная с выработкой мер государственного регулирования и установлением правил «игры» на естественно-монопольных рынках инфраструктурных услуг, исходя из теоретических представлений поведенческой идентификации, существенным образом определяется типом рыночной среды. Тип рынка, который, согласно теории, значительным образом отвечает естественно-монопольной специфике – теоретические модели специальных конкурентоспособных рынков типа *contestable* [1–4]. По условиям функционирования этих рынков ценовые параметры рыночного поведения задаются не априори, а определяются в процессе взаимодействия реальных и потенциальных участников рынка. Причем «тон» в характере этого организационного взаимодействия задает естественный монополист, обладая некото-

рой свободой в определении цен, как правило, в рамках установленных регулятором границ.

В условиях развития цифровой экономики расширяются возможности установления рыночных цен на электронных площадках на договорных основах в системе установленных правил, но дополнительно требуется формирование обоснованных граничных оценок, определяющих допустимые (с учетом социальных установок) значения.

В качестве значимых факторов, определяющих характер ценового поведения участников в рассматриваемых сферах, могут рассматриваться особенности структурных преобразований и уровень их обоснованности, в том числе с теоретических позиций. Одна из ключевых, как правило, декларируемых целей структурных реформ, последующее снижение уровня цен для потребителей услуг. Вместе с тем, «автоматического», следующего за структурными преобразованиями, снижения ценового уровня может и не происходить. Напротив, цены через некоторый временной промежуток, согласно и российскому и зарубежному опыту структурного реформирования в естественно-монопольных сегментах, как правило, возрастают. Т.е. требуются специальные механизмы корректировки последующего ценового поведения участников.

Так, в зарубежной практике, применительно к сфере телекоммуникационных информационных технологий, классическим примером формирования определенного типа ценового поведения (хотя и неявного и нецеленаправленного) можно считать опыт структурной реорганизации, разбиения еще в 80-х годах прошлого века американского транснационального телекоммуникационного конгломерата *American Telephone and Telegraph*. Идентификация (с оценкой субаддитивности многопродуктовой функции совокупных издержек и с принятием в расчет результатов оценки в качестве обоснований в процедурах слушаний и судебных разбирательств) была проведена, применительно к единому рынку телефонной связи, с включением дальней (междугородной и международной) и местной телефонной связи [1,22,23]. Результаты оценки определили выделение собственно естественно-монопольного сегмента отраслевого рынка – местной телефонной связи. При этом реструктуризация во многом способствовала созданию новых линий бизнеса в телекоммуникационной отрасли, развитию информационной инфраструктуры и т.п., что является одной из целевых установок при проведении структурных мероприятий. Вместе с тем в последующие несколько лет имела место значимая возрастающая динамика тарифов

на местных линиях телефонной связи (т.е. в сфере с естественно-монопольным компонентом), в отличие от падающей динамики тарифов на рынках дальней телефонной связи. Учитывая структуру потребительского спроса, применительно к рассматриваемому периоду структурного реформирования, можно считать, что выгоды от падения цен на рынках дальней связи в значительной мере приходились на крупные корпорации и потребителей с высоким уровнем дохода. Потери же преимущественно доставались семьям с низким уровнем дохода, которые в основном пользовались рынком местной телефонной связи.

Подобную ситуацию можно интерпретировать и как пример недостаточной идентификации естественной монополии, ограниченности сформированных оценок лишь нормативным аспектом идентификации, отсутствия должного анализа поведенческого аспекта идентификации, проведения оценок устойчивости (неустойчивости) естественных монополий в ценовом отношении, выработки оценок допустимых границ ценовой неустойчивости для всех участников рынка.

При этом в определенной мере опыт реструктуризации японской телекоммуникационной компании *Nippon Telegraph and Telephone*, относящийся примерно к тому же периоду и также инициированный результатами формальной проверки на основе моделей теории естественной монополии [1,24,25], показывает возможность влияния на последующее ценовое поведение и поддержание цен путем специальной законодательной регламентации, расширения возможностей доступа к сети мелких фирм и развития конкуренции, категоризации деятельности бизнеса и дифференциации монопольных прав.

В условиях цифровой экономики необходимо учитывать опыт «доцифрового» развития, позитивный и негативный, и возможности цифровизации для формирования проблемно ориентированных массивов данных и информационно-аналитических систем моделей ценового поведения применительно к условиям конкурентоспособных рынков, выработки эффективных управленческих механизмов государственного регулирования в рассматриваемых сферах и, соответственно, инициации определенных моделей ценового поведения участников.

Особого внимания требует обеспечение приемлемого уровня обоснованности ценовых решений, принимаемых в ситуации структурного реформирования на регулируемых сегментах рынков инфраструктурных подсистем, выработка адекватных ориентиров ценообразования и специальных

механизмов отслеживания и корректировки ценового поведения с учетом социальных приоритетов.

Заключение

Предлагаемые базовые модели оценки естественно-монопольной синергии и подходы к расширению направлений оценки естественно-монопольных характеристик в рамках поведенческой идентификации могут быть важным дополнением – в части организационной составляющей оценок – к системе показателей эффективности инвестиционных программ развития отраслевых и многоотраслевых структур. Использование системы естественно-монопольных характеристик с включением не только технологических детерминант/естественно-монопольных индикаторов деятельности, но и характеристик ценового поведения на естественно-монопольных рынках, позволяет сформировать комплексное представление о ситуации и направлениях корректировки рекомендаций по структурным преобразованиям, основанным на базовых моделях оценок синергических эффектов организации деятельности.

Литература

1. *Baumol W.J., Panzar J.C., Willig R.D.* Contestable Markets and the Theory of Industry Structure. N.Y.: HBJ. 1982. 497 p.
2. *Baumol W.J., Willig R.D.* Contestability: Development since the Book // Oxford Economic Papers, New Series. Supplement: Strategic Behavior and Industrial Competition. 1986. Vol.38 (Nov.). P. 9-36.
3. *Белоусова Н.И., Васильева Е.М.* Вопросы теории государственного регулирования и идентификации естественных монополий. М.: Ком-Книга. 2006. 320 с.
4. *Белоусова Н.И., Васильева Е.М., Лившиц В.Н.* Модели идентификации естественных монополий и государственного управления ими (возможности расширения классической теории) // Экономика и математические методы. 2012. Т.48. №3. С. 64-78.
5. *Белоусова Н.И., Васильева Е.М.* Диагностика свойств сетевых инфраструктурных технологий в реформируемой системе госрегулирования российских естественных монополий // Российский экономический журнал. 2019. №3. С.25-35. DOI: 10.33983/0130-9757-2019-3-25-35.
6. *Белоусова Н.И., Бушанский С.П., Васильева Е.М., Васильев В.Б.* Естественно-монопольные свойства транспортных сетей: многопродукто-вые модели диагностики// Аудит и финансовый анализ. 2018. Вып. 2. С.129–147.
7. *Box G.E.P., Cox D.R.* An Analysis of Transformations // Journal of Royal Statistical Society. 1964. Ser. B (Methodological). Vol. 26. Issue 2. P. 211-252.
8. *Pulley L.B., Braunstein Y.M.* A Composite Cost Function for Multiproduct Firms with an Application to Economics of Scope in Banking // The Review of Economics and Statistics. 1992. Vol. 74. Issue 2. P. 221-230.
9. *Roller L.H.* Proper Quadratic Cost Functions with an Applications to the Bell System // The Review of Economics and Statistics. 1990. Vol. 72. Issue 2. P. 202-210.
10. *Shephard R.W.* Cost and Production Functions. N.Y.: Princeton University Press. 1953. 104p.
11. *Белоусова Н.И., Васильева Е.М.* О подходах к оценке инфраструктурных технологий с включением гедонических факторов и условий рисков/ Системное моделирование социально-экономических процессов: Труды 44-й межд. научной школы-семинара им. акад. С.С. Шаталина (Воронеж, 4–9 октября 2021). Воронеж: Изд-во «Истоки», 2021. С.75-78. DOI: 10.5281/zenodo.5701385.
12. *Белоусова Н.И., Бушанский С.П., Васильева Е.М., Миронова И.А.* О моделировании оценок естественно-монопольной синергии // Аудит и финансовый анализ. 2023. №5. С.4-10.
13. *Белоусова Н.И., Васильева Е.М.* Естественно-монопольные индикаторы деятельности: теоретические и прикладные аспекты анализа. Труды ИСА РАН. 2018. Т. 68. Вып. 3. С. 69-82. DOI: 10.14357/20790279180307.
14. *Белоусова Н.И., Васильева Е.М.* Эффективная организация отраслевого рынка и оценка естественно-монопольной синергии/ Стратегическое планирование и развитие предприятий: материалы XXV Всероссийского симпозиума. Москва, 9–10 апреля 2024 г. / Под ред. чл.-корр. РАН Г.Б. Клейнера. Электрон. текст. дан. (8,0 Мб). М.: ЦЭМИ РАН. 2024. С. 35-38.
15. *Тироль Ж.* Рынки и рыночная власть: теория организации промышленности / Под ред. В.А. Гальперина, Н.А.Зенкевича. В 2-т. СПб: Экономическая школа. 2000. Т. 1. 328 с. Т. 2. 450 с. (Tirole J. The Theory of Industrial Organization. Massachusetts Institute of Technology. 1988. 1992. 1997).
16. *Baumol W.J., Sidak J.G.* The Pricing of Input Sold to Competitors // Yale Journal on Regulation. 1994. Vol.11. No.1. P.171-202.
17. *Baumol W.J., Ordover J.A., Willig R.D.* Parity Pricing and its Critics: a Necessary Condition for

- Efficiency in the Provision of Bottleneck Services to Competitors // *Yale Journal on Regulation*. 1997. Vol.14. No. 1. P. 145-163.
18. *Baumol W.J., Willig R.D.* Competitive Rail Regulation Rules: Should Price Ceiling Constrain Final Products or Inputs? // *Journal of Transport Economics and Policy*. 1999. Vol.33. Part 1. P. 43-53.
19. *Laffont J.-J., Tirole J.* Access Pricing and Competition // *European Economic Review*. 1994. Vol. 38. Issue 9. P. 1673–1710.
20. *Laffont J.-J., Tirole J.* Creating Competition through Interconnection: Theory and Practice // *Journal of Regulatory Economics*. 1996. Vol. 10. Issue 3. P. 227-256.
21. *Amstrong M.* Network Interconnection in Telecommunications // *Economic Journal*. 1998. Vol. 108. Issue 448. P. 545-564.
22. *Evans D., Heckman J.* A Test for Subadditivity of the Cost Function with an Application to the Bell System // *American Economic Review*. 1984. Vol. 74. Issue 4. P. 615-623.
23. *Evans D., Heckman J.* A Test for Subadditivity of the Cost Function with the Application to the Bell System: Erratum // *American Economic Review*. 1986. Vol. 76. Issue 4. P. 856-858.
24. *Sueyoshi T.* Divestiture of Nippon Telegraph and Telephone // *Management Science*. 1996. Vol. 42. Issue 9. P. 1326-1351.
25. *Sueyoshi T.* Measuring Effectiveness and Returns of Scale of Nippon Telegraph & Telephone in Production and Cost Analyses // *Management Science*. 1997. Vol. 43. Issue 6. P. 779-796.

Белоусова Наталия Ивановна. Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия. Ведущий научный сотрудник. Доктор экономических наук. Область научных интересов: естественные монополии, методология системной оценки. E-mail: natabel.52@mail.ru; belousova@isa.ru

Бушанский Сергей Петрович. Центральный экономико-математический институт Российской академии наук, г. Москва, Россия. Старший научный сотрудник. Кандидат экономических наук. Область научных интересов: разработка методов построения оптимальных транспортных сетей, исследование проблем обоснования крупных инфраструктурных проектов, моделирование процессов принятия решений в системе государственного проектирования. E-mail: dbd-s@yandex.ru

Modeling assessments of synergy effects under organizational activity in the system of natural monopoly characteristics

N.I. Belousova^I, S.P. Bushansky^{II}

^I Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

^{II} Central Economics and Mathematics Institute of Russian Academy of Sciences, Moscow, Russia

Abstract. Theoretical and applied approaches to assessing synergy effects under organizational activity are researched based on the models and techniques of the modern natural monopoly theory for solving problems of analysis and strategic design for infrastructure subsystems. The relationships between adequate theoretical foundation of effective market organization and the approaches to assessing natural monopoly synergy are analyzed. Within the normative identification, approaches to assessing the synergetic effects using key technological determinants/ natural monopoly activity indicators have been elaborated. The directions are determined for analyzing factors associated with assessing synergy effects within the framework of theoretical concepts of behavioral identification, and, accordingly, expanding the set of using characteristics. Some measures of state regulation related to the formation of market economic behavior of natural monopolist, primarily pricing behavior, are highlighted; these measures determine the possibility of adjusting estimates of synergy effects under organizational activity.

Keywords: *strategy design, natural monopoly theory, effective market organization, synergy effects under organizational activity, normative and behavioral identification, the splitting/combining of infrastructure subsystems, network specifics, the system of natural monopoly characteristics, state regulation, price behavior.*

DOI: 10.14357/20790279240301 **EDN:** DRBVPJ

References

1. Baumol W.J., Panzar J.C., Willig R.D. Contestable Markets and the Theory of Industry Structure. N.Y.: HBJ; 1982. 497 p.
2. Baumol W.J., Willig R.D. Contestability: development since the book. Oxford Economic Papers, New Series. Supplement: Strategic Behavior and Industrial Competition. 1986; 38 (Nov.): 9-36.
3. Belousova N.I., Vasil'eva E.M. Issues of the Theory of State Regulation and Identification of Natural Monopolies. M.: KomKniga; 2006. 320 p. (in Russ.).
4. Belousova N.I., Vasil'eva E.M., Livshitz V.N. Models of identification of natural monopolies and their state regulation (opportunities for expanding the classical theory). Ekonomika i matematicheskie metody. 2012; 48(3): 64-78 (in Russ.).
5. Belousova N.I., Vasil'eva E.M. Diagnostics of network infrastructure technologies properties under reforming the system of Russian natural monopolies state reeregulation. Rossijskij e'konomicheskij zhurnal. 2019; 3: 25-35. doi: 10.33983/0130-9757-2019-3-25-35 (In Russ.).
6. Belousova N.I., Bushanskij S.P., Vasil'eva E.M., Vasil'ev V.B. Natural monopoly properties of transport networks: multiproduct models of diagnostics. Audit i finansovy'j analiz. 2018; 2: 129-147 (In Russ.).
7. Box G.E.P., Cox D.R. An analysis of transformations. Journal of Royal Statistical Society. Ser. B (Methodological). 1964; 26(2): 211-252.
8. Pulley L.B., Braunstein Y.M. A composite cost function for multiproduct firms with an application to economies of scope in banking. The Review of Economics and Statistics. 1992; 74(2): 221-230.
9. Roller L.H. Proper quadratic cost functions with an applications to the Bell System. The Review of Economics and Statistics. 1990; 72(2): 202-210.
10. Shephard R.W. Cost and Production Functions. N.Y.: Princeton University Press; 1953. 104p.
11. Belousova N.I., Vasil'eva E.M. Approaches to the assessment of infrastructure technologies with the inclusion of hedonic factors and risk conditions. In: system modeling of socio-economic processes: Proceedings of the 44th International Scientific school-seminar named acad. S.S. Shatalin. 4-9 October 2021, Voronezh: Istoki Publishing House; 2021. P.75-78. doi: 10.5281/zenodo.5701385 (In Russ.).
12. Belousova N.I., Bushanskij S.P., Vasil'eva E.M., Mironova I.A. On modeling assessments of natural monopoly synergy. Audit i finansovy'j analiz. 2023; 5: 4-10 (In Russ.).
13. Belousova N.I., Vasil'eva E.M. Natural monopoly activity indicators: theoretic and applied aspects

- of analysis. *Trudy ISA RAN*. 2018; 68(3): 69-82. doi: 10.14357/20790279180307 (In Russ.).
14. *Belousova N.I., Vasil'eva E.M.* Effective organization of the industry market and assessment of natural monopoly synergy. In: Kleiner G.B. (ed.): strategic planning and enterprise development: Proceedings of the XXV All-Russian Symposium. 9-10 April 2024, Moscow: CEMI RAS; 2024. P.35-38 (In Russ.).
15. *Tirol' Zh.* Markets and Market Power: Theory of Industrial Organization / Ed. by V.A. Galperin, N.A.Zenkevich. Vol. 1-2. St.Petersburg: Economic School; 2000. Vol.1-328p. Vol.2- 450p. (In Russ.). (Tirole J. The Theory of Industrial Organization. Massachusetts Institute of Technology, 1988, 1992, 1997).
16. *Baumol W.J., Sidak J.G.* The pricing of input sold to competitors. *Yale Journal on Regulation*. 1994; 11(1):171-202.
17. *Baumol W.J., Ordover J.A., Willig R.D.* Parity pricing and its critics: a necessary condition for efficiency in the provision of bottleneck services to competitors. *Yale Journal on Regulation*. 1997; 14(1):145-163.
18. *Baumol W.J., Willig R.D.* Competitive rail regulation rules: should price ceiling constrain final products or inputs? *Journal of Transport Economics and Policy*. 1999; 33(1):43-53.
19. *Laffont J.-J., Tirole J.* Access pricing and competition. *European Economic Review*. 1994; 38(9): 1673–1710.
20. *Laffont J.-J., Tirole J.* Creating competition through interconnection: theory and practice. *Journal of Regulatory Economics*. 1996; 10(3): 227-256.
21. *Amstrong M.* Network interconnection in telecommunications. *Economic Journal*. 1998; 108 (448): 545-564.
22. *Evans D., Heckman J.* A Test for subadditivity of the cost function with an application to the Bell System. *American Economic Review*. 1984; 74(4): 615-623.
23. *Evans D., Heckman J.* A test for subadditivity of the cost function with the application to the Bell System: Erratum. *American Economic Review*. 1986; 76(4): 856-858.
24. *Sueyoshi T.* Divestiture of Nippon Telegraph and Telephone. *Management Science*. 1996; 42(9):1326-1351.
25. *Sueyoshi T.* Measuring effectiveness and returns of scale of Nippon Telegraph &Telephone in production and cost analyses. *Management Science*. 1997; 43(6): 779-796.

Belousova N. Leading Researcher, Doctor (Economy), Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia. E-mail: natabel.52@mail.ru; belousova@isa.ru

Bushansky S. Phd (Economics), Senior Researcher, Central Economics and Mathematics Institute of Russian Academy of Sciences, Moscow, Russia. E-mail: dbd-s@yandex.ru

Действие закона диффузии внешней среды организаций как социально-экономических систем

И.В. НАЙДЕНОВ

Западный филиал Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации,
г. Калининград, Россия

Аннотация. Каждую организацию можно представить системой, сложность которой зависит от количества элементов ее составляющих и метода систематизации. Совокупности организаций образуют производственные и отраслевые объединения на региональном и национальном уровнях. Общая теоретическая направленность на изучение организаций в отраслях экономики и их представление, как целостных социально-экономических образований, позволяет уточнить главную задачу любого исследования и практической деятельности предприятия. Статья посвящена исследованию специфического организационного закона – закона диффузии внешней среды. Рассмотрены методы исчисления соответствующих коэффициентов эластичности для прогнозирования объемов работы организаций и их объединений на любом уровне агрегирования. Эта категория занимает центральное место в теории данного закона, поскольку очень важна для понимания содержания и сути вещей, основанных на объективной интерпретации динамического изменения объединений организаций на различных уровнях управления.

Ключевые слова: законы теории организации, методы описания социально-экономических систем, прогнозирование экономических показателей, темпы прироста объема работы предприятий и организаций.

DOI: 10.14357/20790279240302 **EDN:** EDBQPO

Введение

Система, как категория, является ключевой для соответствующей теории и современного представления организаций социально-экономических систем [6–8] в том числе и на уровне региона. Вместе с тем управление, как процесс, всегда происходит в привязке к конкретному месту расположения организации и по отношению к данному объекту, например юридическому лицу. На теоретическом уровне управление происходит в технических (технологических), биологических и социальных системах. Управление в социально-экономических системах, как правило, рассматривается с позиций организационного подхода. Принципиальной особенностью такого подхода и одной из наиболее трудных задач современного представления организаций, как социально-экономических образований, является их соответствующее сопоставление с элементами внутренней и внешней среды. Этой теме посвящено большое количество исследований [11, 13, 14, 19].

Анализ и соответствующая детализация конкретных организационных структур, в том числе

регионального уровня, предполагает необходимость формирования теоретической основы системного анализа организаций. Современный инструментарий реализации системного подхода для анализа и прогнозирования предприятий и их объединений весьма велик. Но традиционные методы и практика его применения зачастую узконаправленны, что не всегда позволяют реализовать принципы системного подхода, так как подменяются традиционным финансовым или экономическим анализом. Подобное замещение понятий не дает возможность достичь планируемых целей и возможности рассматривать организацию как систему в завершённом виде.

Принципиальной категорией для современных организаций является их отнесение к открытым или закрытым системам и соответствующее определение их поведения по отношению к элементам внешней среды. Логическое построение анализа методов управления и дифференциация на этой основе из организационной среды основных признаков для построений и сопоставления

полученных результатов, позволяют сформировать задачу о необходимости выделения главных компонентов исследования: темпов роста объемов работы, динамики значений удельного веса объемов работы предприятий и их объединений во внешней среде и соотношения этих ключевых показателей, как инструментов оценки и прогнозирования результатов работы организаций.

1. Основные категории теории и методы описания социально-экономических систем

Определений категории *система* существует очень много, поскольку его содержание зависит от отраслевой принадлежности или области знаний, в которой работает специалист. Каждое из них выделяет основные свойства или стороны теории систем, которые представляются экспертам главными для разработки и решения соответствующей задачи. Одно из наиболее лаконичных определений: *система* – множество взаимосвязанных элементов, образующих целостность.

Людвиг фон Бергаланфи, основоположник теории систем, определяет ее как комплекс взаимодействующих элементов, находящихся в определенных отношениях друг с другом и со средой [2]. Другой известный автор теории систем А. Холл определяет систему как множество предметов вместе со связями между предметами и их признаками [20]. В.А. Острейковский [18] предложил формальное описание последовательности определений системы как множества $\{Do\}_n$, где $n=1,2,...,N$, в котором может учитываться такое количество элементов, связей и действий в реальной системе, которое необходимо для решаемой задачи, для достижения поставленной цели. В символьном виде некоторые из определений системы из коллекции В.А. Острейковского представляются следующей последовательностью.

D1. Система S есть нечто целое:

$$S=(I,0);$$

D2. Система S есть организованное множество:

$$S=(opg, M),$$

где opg – оператор организации, M – множество;

D3. Система S есть множество вещей, свойств и отношений:

$$S=(m, n, r),$$

где m – вещи, n – свойства, r – отношения.

Одно из определений, которое В.А. Острейковский приводит специально для организационных систем, выглядит следующим образом:

$$S=(PL, RO, RJ, EX, PR, DT, SV, RD, EF), \quad (1)$$

где PL – цели и планы, RO – внешние ресурсы, RJ – внутренние ресурсы, EX – исполнители, PR – процесс, DT – помехи, SV – контроль, RD – управление, EF – эффект.

Для социально-экономических систем данное видение организации представляется недостаточно полным и логичным, так как элементы такого определения составлены случайным образом. Поэтому определение организации с позиций системного подхода может быть сформулировано следующим образом: «Организация – социально-экономическая динамическая система, состоящая из множества взаимосвязанных элементов, которые образуют определенную целостность, действуют в соответствии с поставленной целью и объективными законами, стремясь максимально продлить свой жизненный цикл и сохранить ресурс»:

$$S=(X, Y, Z, H, RI, D, PS), \quad (2)$$

где X – входы, Y – выходы, Z – состояния, H – операторы переходов, RI – внутренние ресурсы, D – помехи, PS – цели.

В качестве основных понятий, характеризующих строение и функционирование систем, назовем следующие: *внешняя среда, внутренняя среда, элемент, структура, связь, состояние, поведение, модель, равновесие, устойчивость, цель*. Под *элементом* будем понимать неделимую и функционально-самостоятельную часть системы. Что является такой частью для социально-экономических систем? Ответ находится в интервале: {человек} – {функциональное звено управления, линейное звено управления}. Например, единица или совокупность единиц основных производственных фондов, управляемых работником данного предприятия, образуют линейное звено управления.

Приведенный ответ на вопрос является достаточно условным и зависит от постановки задачи системного подхода к явлению или объекту. С позиций системного анализа элемент – это предел членения системы, т.е. разделения организационной целостности на составляющие части в зависимости от формулировки цели и ее уточнения в процессе исследования. Все элементы, составляющие структуру системы, могут быть классифицированы следующим образом:

- принимающие входные сигналы из внешней среды (входные полюсы);
- выдающие выходные сигналы (выходные полюсы);
- принимающие или выдающие сигналы от других элементов системы (внутренние элементы).

Структура – производная от категории элемент – совокупность элементов и взаимосвязей между ними, одна из основополагающих категорий для настоящего исследования, поскольку является образующей и конструктивной основой взаимодействия как элементов внутренней среды организаций, так и множества организаций между собой [7]. Взаимодействие организаций, объединенных в структуры, образует уровни организаций различной степени интеграции и агрегирования [5]. Это является предпосылкой, образующим фактором для описания характера взаимодействия открытых социально-экономических систем. Современная глобализация экономики является лишь результатом сложения взаимосвязей многих предприятий и организаций, расположенных в различных национально-экономических образованиях. Между тем взаимодействие происходит в технических, информационных и экономических средах. В организационных структурах отражаются наиболее значимые взаимосвязи между элементами, существующие в течение анализируемого периода времени и отражающие их расположение и порядок взаимодействия. В зависимости от типа структуры последняя может отражать теоретическое взаимоотношение элементов или реальное их расположение (дислокацию). Взаимоотношения могут касаться как элементов различного уровня управления, т.е. как неделимых элементов, так и подсистем более высокого уровня, имеющих «свою» структуру. Наличие структуры обуславливает сам факт существования системы, а ее качество – успешность функционирования системы и достижение или не достижение целей. Соответственно структура – это неотъемлемый атрибут такой организации, как социально-экономическая система. Как правило, структура представляется в виде бесконтурного графа, а также может быть описана посредством матриц, множеств и других языков моделирования. Бесконтурным называется ориентированный граф, не содержащий контуров [21]. Контур – это замкнутый путь в графе, т.е. такая последовательность вершин и дуг, что для всех вершин есть дуга в оргграфе.

Одно из основных качеств организационных структур – их иерархичность. Иерархия – это упорядоченность компонентов по степени важности (многоступенчатость, служебная лестница) [3]. Между уровнями иерархической структуры могут существовать взаимоотношения строгого подчинения компонентов (узлов) нижележащего уровня одному из компонентов вышележащего уровня,

т.е. отношения так называемого древовидного порядка. Такие иерархии называются сильными, или иерархиями типа «дерево». Один и тот же узел нижележащего уровня может быть одновременно подчинен нескольким узлам вышележащего уровня. Такие структуры называются иерархическими со слабыми связями. Между уровнями иерархической структуры могут существовать и более сложные взаимоотношения, например, «страты», «слои», «эшелоны».

Основными видами организационных структур управления (ОСУ) являются линейные, функциональные, линейно-функциональные, матричные, многомерные и др. [1]. Традиционный вид некоторых из них может быть представлен в виде графов и блок-схем. Цифровое представление графов возможно с применением матриц смежности и достижимости [21].

Если система существует, то постоянно изменяются характеристики составляющих ее элементов. Тогда состояние системы – точка в траектории ее развития. Если обозначить множество состояний системы S через Z , то каждое определяется набором действительных чисел. Множество Z называется пространством состояний системы. Аналогом пространства состояний в традиционной интерпретации является трехмерное евклидово пространство, свойства которого описываются аксиомами евклидовой геометрии, где $Z(t)$ есть точка пространства с координатами $z_1, z_2, \dots, z_t, \dots, z_T$.

2. Закон диффузии внешней среды

Как правило, предметом экономического анализа является хозяйственная деятельность организации и определение того, как окружение влияет на нее. Как уже говорилось ранее, процесс развития теории организации можно представить в виде динамик: от рациональных организаций к социальным и от закрытых систем к открытым. Параллельно с этим процессом важным вкладом в развитие теории, являются научные исследования и серии наблюдений, объединенных в последствии Д. Глейком, как «теория хаоса». Это расширяет наше представление о механизмах взаимного воздействия элементов внешней среды на организацию. В частности это выводы в отношении эволюции сложных систем и правил их поведения в будущем [9]. Один из них сформулирован следующим образом: *мельчайшие изменения в настоящем вызывают потрясающие дальнейшие изменения системы.*

Практические рекомендации могут быть таковы: метод исследования должен осуществляться

на уровне отдельных агрегатов организаций как некоторой целостности (агрегативный подход к информационному описанию объекта исследования), но одновременно с этим нужно учитывать пропорции их развития между собой. Это обуславливает, с одной стороны, вероятностный подход к определению достижения состояния организации как системы в целом. С другой стороны, обосновывается важность происходящих событий в частях, ее составляющих, что соотносится к состоянию системы в конкретный период времени или в соответствующей ситуации.

Темпы прироста результатов функционирования (объемов работы) организаций и их объединения, различных уровней агрегирования, неодинаковы [16, с.62]. То есть:

$$T_1^{t+i} \neq T_2^{t+i}; \quad (3)$$

$$T_2^{t+i} \neq T_3^{t+i}, \quad (4)$$

где T_1^{t+i} , T_2^{t+i} , T_3^{t+i} - темпы прироста объема работы соответственно организации (или объединения организаций) самого низкого уровня (УРОВЕНЬ 1), объединения организаций среднего (более высокого уровня) (УРОВЕНЬ 2), объединения организаций самого высокого уровня агрегирования (УРОВЕНЬ 3) на $t+i$ год; t, i - номера года периодов ретроспективы и прогноза ($t \in T, i \in I$) соответственно.

От неравенств (3), (4) перейдем к равенствам вида:

$$T_1^{t+i} = (T_1^{t+i} - T_2^{t+i}) + T_2^{t+i}; \quad (5)$$

$$T_2^{t+i} = (T_2^{t+i} - T_3^{t+i}) + T_3^{t+i}. \quad (6)$$

Подставив (6) в (5), получим:

$$T_1^{t+i} = (T_1^{t+i} - T_2^{t+i}) + (T_2^{t+i} - T_3^{t+i}) + T_3^{t+i} \quad (7)$$

Разности в уравнении (7) обозначим как частные темпы прироста:

$$T_{1,2}^{t+i} = (T_1^{t+i} - T_2^{t+i}); \quad (8)$$

$$T_{2,3}^{t+i} = (T_2^{t+i} - T_3^{t+i}), \quad (9)$$

где $T_{1,2}$, $T_{2,3}$ - частные темпы прироста объема работы «УРОВНЯ 1» по отношению к «УРОВНЮ 2» и «УРОВНЯ 2» по отношению к «УРОВНЮ 3», соответственно.

Таким образом, исходными данными для определения темпов прироста и удельных весов являются объемы работы организаций различного уровня агрегирования. В соответствии с такой постановкой задачи необходимо составить решение в едином процессе увязки динамик изменений различных элементов структуры объединения организаций различных уровней управления. В данном

случае темпы прироста «УРОВНЯ 1» складываются из трех составляющих:

$$T_1^{t+i} = T_{1,2}^{t+i} + T_{2,3}^{t+i} + T_3^{t+i}. \quad (10)$$

А показатели изменений удельного веса рассчитываются как отношения объемов работы организаций (Θ) различных уровней управления в динамике. Соответственно, исследование эмпирических данных позволяет предположить, что соответствующие коэффициенты, *коэффициенты эластичности объемов работы*, прямо пропорциональны частным темпам прироста и обратно пропорциональны приращениям удельных весов объемов работы:

$$\mathfrak{E}_{1,2(2,3)}^{t+i} = \frac{T_{1,2(2,3)}^{t+i}}{\Delta d_{1,2(2,3)}^{t+i}}, \quad (11)$$

или

$$\mathfrak{E}_{1,2(2,3)}^{t+i} = \frac{T_{1(2)}^{t+i} - T_{2(3)}^{t+i}}{\Delta d_{1,2(2,3)}^{t+i}}, \quad (12)$$

В обобщенном виде, поскольку в каждом конкретном случае имеем разное количество уровней, это будет следующая формальная запись:

$$T_{\lambda=i} = \sum_{\lambda=\tau+1}^{\lambda=\eta} T_{\lambda-1,\lambda} + T_{\lambda=\eta}. \quad (13)$$

где $\lambda \in \{\lambda\}$, или $\lambda \in \lambda$ - множество уровней агрегирования работы организаций различного уровня (в системе обозначений, принятых нами в данной работе, $\lambda=1$ (УРОВНЯ 1), $\lambda=2$ (УРОВЕНЬ 2), $\lambda=3$ (УРОВЕНЬ 3); (τ, η) - интервал уровней агрегирования объема работы.

В соответствии с формулой (10) предполагаем наличие равносторонних характеристик изменения темпов прироста объемов работы разных видов организаций и разных уровней агрегирования. Если рассматривать для примера одну из отраслей экономики - транспорт, это зависит от динамики развития пропорций видов транспорта, т.е. от изменений удельного веса объема работы. Показатели, устанавливающие указанные зависимости, названы коэффициентами эластичности уровней агрегирования объемов работы транспорта - $\mathfrak{E}_{1,2}$, $\mathfrak{E}_{2,3}$. По отношению к видам транспорта - это коэффициенты эластичности видов транспорта.

Теорема. Если данной системе соответствует структура, состоящая из двух и более уровней агрегирования, то существует положительное множество значений коэффициентов эластичности.

Или. Наличие двух и более уровней агрегирования в данной системе является достаточным условием существования непустого множества значений коэффициентов эластичности, и при этом все элементы данного множества больше единицы.

То есть:

$$\lambda \geq 2 \Rightarrow \begin{cases} 1) \exists \{\Theta\} = \left\{ \Theta_{\lambda-1, \lambda}^{t+i} \right\}_{\lambda=2}^k; \\ 2) \forall \Theta_{\lambda-1, \lambda}^{t+i} \in \{\Theta\}: \Theta_{\lambda-1, \lambda}^{t+i} > 1. \end{cases}$$

Доказательство. Так как исходными данными для определения темпов прироста удельных весов являются объемы работы (Θ), то выразим через них коэффициент эластичности. Тогда получим:

$$T_{\lambda}^{t+i} = \frac{\Theta_{\lambda}^{t+i} - \Theta_{\lambda}^{(t+i)-1}}{\Theta_{\lambda}^{(t+i)-1}}; \quad (14)$$

$$\Delta d_{(\lambda-1), \lambda}^{t+i} = \frac{\Theta_{\lambda-1}^{t+i}}{\Theta_{\lambda}^{t+i}} - \frac{\Theta_{\lambda-1}^{(t+i)-1}}{\Theta_{\lambda}^{(t+i)-1}}, \quad (15)$$

где Θ_{λ}^{t+i} – объем работы λ -го уровня агрегирования на $t+i$ -й период.

Подставив (14) и (15) в (12), получим:

$$\Theta_{(\lambda-1), \lambda}^{t+i} = \frac{\frac{\Theta_{\lambda-1}^{t+i} - \Theta_{\lambda-1}^{(t+i)-1}}{\Theta_{\lambda-1}^{(t+i)-1}} - \frac{\Theta_{\lambda}^{t+i} - \Theta_{\lambda}^{(t+i)-1}}{\Theta_{\lambda}^{(t+i)-1}}}{\frac{\Theta_{\lambda-1}^{t+i}}{\Theta_{\lambda}^{t+i}} - \frac{\Theta_{\lambda-1}^{(t+i)-1}}{\Theta_{\lambda}^{(t+i)-1}}} \quad (16)$$

или

$$\Theta_{(\lambda-1), \lambda}^{t+i} = \frac{\left(\frac{\Theta_{\lambda-1}^{t+i}}{\Theta_{\lambda-1}^{(t+i)-1}} - 1 \right) - \left(\frac{\Theta_{\lambda}^{t+i}}{\Theta_{\lambda}^{(t+i)-1}} - 1 \right)}{\frac{\Theta_{\lambda-1}^{t+i} \Theta_{\lambda}^{(t+i)-1} - \Theta_{\lambda-1}^{(t+i)-1} \Theta_{\lambda}^{t+i}}{\Theta_{\lambda-1}^{(t+i)-1} \Theta_{\lambda}^{(t+i)-1}}}.$$

Отсюда, после алгебраических преобразований приходим к виду:

$$\Theta_{(\lambda-1), \lambda}^{t+i} = \frac{\Theta_{\lambda}^{t+i}}{\Theta_{\lambda-1}^{(t+i)-1}}, \quad (17)$$

т.е. величина *коэффициента эластичности объемов работы* равна частному от деления объема работы уровня λ в $t+i$ -й год на объем работы меньшего уровня агрегирования $\lambda-1$ предыдущего $(t+i)-1$ -го года и всегда больше 1, что и требовалось доказать.

Для практической иллюстрации расчетов соответствующих значений коэффициента эластичности приводятся данные объемов работы видов транспорта Российской Федерации. Из табл. 1 видно, что наибольший объем перевозок выполняется автомобильным транспортом, а наименьший – воздушным. Но самая большая величина коэффициентов эластичности соответствует воздушному транспорту и, например, по результатам расчета за 2021 г. составляет 6285,38.

Вместе с тем, на основе формулы (17) можно составить принципиальную математическую преамбулу закона и соответствующий ей текст Закона диффузии внешней среды:

Объемы работы организаций (отраслевых объединений организаций) можно определить на основе статистических (известных) данных более низких уровней управления и соответствующих им коэффициентов эластичности.

Закон диффузии внешней среды:

Любые мельчайшие изменения в настоящем организованном целом вызывают потрясающие суммарные изменения в будущем целого более высокого уровня при этом количество рассматри-

Табл. 1

Объем перевозок грузов по видам транспорта России, млн. тонн [22, с.37]

Год	2 015	2019	2 020	2 021
Железнодорожный	1 329	1 399	1 359	1 404
Автомобильный	5 357	5 735	5 405	5 491
Трубопроводный	1 071	1 159	1 061	1 141
Морской	19	23	25	23
Внутренний водный	121	108	109	110
Воздушный	1	1,3	1,3	1,6
Объем в тоннах, всего	7 898	8426	7 960	8 171

Табл. 2

Удельный вес по объему перевозок по видам транспорта России

Год	2 015	2019	2 020	2 021
Железнодорожный	0,1683	0,1660	0,1707	0,1718
Автомобильный	0,6783	0,8606	0,6790	0,6720
Трубопроводный	0,1356	0,1376	0,1333	0,1396
Морской	0,0024	0,0027	0,0031	0,0028
Внутренний водный	0,0153	0,0128	0,0137	0,0134
Воздушный	0,0001	0,0002	0,00016	0,0002

Табл. 3

Коэффициенты эластичности по объему перевозок

Год	2 020	2 021
Железнодорожный	5,689	6,013
Автомобильный	1,388	1,512
Трубопроводный	6,868	7,701
Морской	346,09	326,84
Внутренний водный	73,704	74,281
Воздушный	6123,08	6285,38

Заключение

ваемых периодов перспективы равно количеству отождествляемых уровней управления.

В формализованном виде Закон записывается в следующем виде [15, с.123]:

$$\Theta_{\lambda=\eta}^{t+i-B_0} = \Theta_{\lambda=\tau}^{(t+i)-B_{\{\eta,\tau\}}} \prod_{\lambda=\tau+1, B=(\eta-\tau)-1}^{\lambda=\eta, B=0} \mathfrak{E}_{\lambda-1, \lambda}^{t+i-B},$$

при $\{\lambda, B\} = \{\tau+1, (\eta-\tau)-1\},$
 $\{\tau+2, (\eta-\tau)-2\}, \dots, \{\eta, 0\},$ (18)

где B – величина, устанавливающая соответствие между объемами работы уровней агрегирования и коэффициентами эластичности соответствующего временного периода.

Следствие 1. Если существует λ -1 уровень агрегирования работы по отношению к более высокому уровню, то всегда существует некоторая величина изменения удельного веса работы, не равная нулю, т.е.

$$\Delta d_{(\lambda-1), \lambda}^{t+i} \in \Delta D_{(\lambda-1), \lambda},$$

где $\Delta d_{(\lambda-1), \lambda}^{t+i} \neq 0, \Delta D_{(\lambda-1), \lambda} \neq 0$.

Следствие 2. В соответствии с (17) объем работы любого уровня агрегирования легко определить через объем работы меньшего уровня агрегирования, умножая его на соответствующий коэффициент эластичности:

$$\Theta_2^{t+i} = \Theta_1^{(t+i)-1} \mathfrak{E}_{1,2}^{t+i}, \quad (19)$$

$$\Theta_3^{t+i} = \Theta_2^{(t+i)-1} \mathfrak{E}_{2,3}^{t+i}. \quad (20)$$

В соответствии с (19) и (20) объем работы самого высокого уровня агрегирования можно определить через самый низкий уровень агрегирования:

$$\Theta_3^{t+i} = \Theta_1^{(t+i)-2} \mathfrak{E}_{1,2}^{t+i-1} \mathfrak{E}_{2,3}^{t+i}. \quad (21)$$

В общем виде объем работы, соответствующий любому более высокому уровню агрегирования, можно вычислить через объем работы любого более низкого уровня агрегирования, умножая его на произведение соответствующих коэффициентов эластичности, учитывающее интервал уровней агрегирования (табл. 2,3)

Законы теории организации образуют теоретический фундамент, определяют место и роль соответствующих знаний как самостоятельного научного направления. Общие законы организации были сформулированы основоположником организационной науки А.А. Богдановым [4]. Сформулированный закон диффузии внешней среды относится к числу дополнительных законов теории организации. Основное целеполагание при разработке соответствующего научно-методического аппарата заключается в расширении экономико-математического инструментария для прогнозирования и работы конкретных организаций и их объединений для достижения положительных результатов работы в планируемом периоде.

Объемы работы организаций и их объединений в натуральной или стоимостной форме исчисляются на различных уровнях управления и временных периодах. Однако во многих случаях, особенно в процессе прогнозирования, некоторые элементы статистики остаются неизвестными. Поэтому применение алгоритма и соответствующего методического и математического инструментария Закона диффузии внешней среды и средних значений коэффициентов эластичности может способствовать более точному исчислению соответствующих показателей.

Очевидно, что количественные значения объемов работы в (17) – это только положительные величины. Поэтому значение коэффициента эластичности – только положительная величина. Вместе с тем коэффициент эластичности невозможно вычислить, если отсутствуют объемы работ, выполняемые другими организациями, объединениями организаций. Формула (17) упрощает процедуру вычисления коэффициентов эластичности, так как исключает ряд промежуточных операций, необходимых для определения относительных величин – темпов прироста T и изменений удельного веса Δd . Но основной экономический подход, вытекающий из данного метода расчетов, позволяет в условиях практического использования заранее предпо-

ложить как результаты работы текущего периода работы данного объединения предприятий, например вида транспорта или сельскохозяйственного производства, повлияют на результаты работы отрасли в целом в следующем временном периоде.

Литература

1. Антонов В. Г. Эволюция организационных структур // Менеджмент в России и за рубежом. 2000. №1. С. 25–31.
2. Берталанфи Фон Л. Общая теория систем – критический обзор. // Исследования по общей теории систем. М.: Прогресс. 1969. С. 23–82.
3. Берталанфи Фон Л. История и статус общей теории систем // Системные исследования: Ежегодник. М.: Наука. 1973. С. 20–37.
4. Богданов А.А. Тектология: Всеобщая организационная наука. В 2 кн. Под редакцией Л.И. Абалкина (отв. ред.) и др. М.: Экономика. 1989. 304 с., 352 с.
5. Виссема Х. Стратегический менеджмент и предпринимательство: возможности для будущего процветания / Пер. с англ. М. 2000. 271 с.
6. Гвишиани Д.М. Материалистическая диалектика – философская основа системных исследований. Системные исследования: Методологические проблемы. Ежегодник. М.: Наука. 1979. С. 7–26.
7. Гвишиани Д. М. Организация и управление. – Изд. 3-е перераб. М. Изд-во МГТУ им. М.Э. Баумана. 1998. 331 с.
8. Гибсон Дж. Л., Иванцевич Д. М., Доннелли Д. Х. Организации: поведение, структура, процессы / Пер. с англ. М. 2000. 660 с.
9. Глейк Дж. Хаос: Создание новой науки / Пер. с англ. М. Нахмансона, Е. Барашковой. СПб.: Амфора. 2001. 396 с.
10. Клейнер Г.Б. Производственные функции: Теория, методы, применение. М.: Финансы и статистика. 1986. 239 с.
11. Коган М.С. Система и структура // Системные исследования: Методологические проблемы: Ежегодник. 1983. М.: Наука. 1983. С. 88–105.
12. Лившиц В.Н., Орлова Е.Р., Тищенко Т.И., Мызникова М.Н. Новая модель развития экономики России // Труды ИСА РАН. 2021. Том 71. № 3. С. 3 – 10.
13. Мильнер Б.З. Проблемы исследования и развития организационных систем // Системные исследования: Методологические проблемы. Ежегодник 1984. М. 1984. С. 207–227.
14. Мильнер Б.З., Евенко Л.И., Рапопорт В.С. Системный подход к организации управления. М., 1983. 224 с.
15. Найденов И.В. Теория организации. Калининград: Изд-во КГУ. 2004. 195 с.
16. Найденов И.В. Прогнозирование развития открытых социально-экономических систем на транспорте. М.: Финстатинформ. 2001. 248 с.
17. Найденов И.В., Найденова Г.И. Прогнозирование показателей работы транспорта в Калининградской особой экономической зоне // Труды ИСА РАН. 2018. Том 68. № 3. С. 91 – 96.
18. Острейковский В. А. Теория систем: Учеб. для вузов по спец. «Автом. сист. обр. информ. и упр.». М.: Высш. школа. 1997. 239 с.
19. Темников Ф.Е. Вопросы теории и методологии систем // Тр. моск. энерг. ин-та: Системотехника. Вып. 158. М.. 1973. С. 3–5.
20. Холл А.Д. Опыт методологии для системотехники. М.: Сов. радио. 1975. 448 с.
21. Теория графов: Алгоритмы обработки бесконечных графов / В. А. Евстигнеев, В. Н. Касьянов. – Новосибирск: Наука. Сиб. Предприятие РАН, 1998. 385 с.
22. Транспорт в России. 2022: Стат.сб./Росстат. – М. 2022. 101 с
23. Ivancevich J.M., Donnelly J.H. Organizations: behavior, structure, processes. IRWIN. 1997. 845 с.
24. Morgan G. Images of organization. London. 1989. 520 с.
25. Robbins P.S. Organizational behavior: concepts, controversies, applications. San Diego State University. 1996. 875 с.

Найденов Игорь Витальевич. Западный филиал Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации, г. Калининград, Россия. Профессор, доктор экономических наук. Область научных интересов: анализ экономических процессов на основе использования экономико-математических методов и инструментальных средств, изучение экономических систем на уровне региона в качестве объектов управления, системные преобразования в регионе, формирование и системные преобразования на транспортном комплексе региона. E-mail: naydenovigor@yandex.ru

The action of the law of diffusion of the external environment of organizations as socio-economic systems

I.V. Naydenov

Western branch of the Russian Presidential Academy of National Economy and Public Administration, Kaliningrad, Russia

Abstract: Each organization can be represented as a system, the complexity of which depends on the number of elements of its components and the method of systematization. Collections of organizations form production and industry associations at the regional and national levels. The general theoretical focus on the study of organizations in economic sectors and their presentation as integral socio-economic entities makes it possible to clarify the main task of any research and practical activity of an enterprise. The article is devoted to the study of a specific organizational law - the law of diffusion of the external environment. Methods for calculating the corresponding elasticity coefficients for predicting the volume of work of organizations and their associations at any level of aggregation are considered. This category occupies a central place in the theory of this law, since it is very important for understanding the content and essence of things based on an objective interpretation of the dynamic changes in associations of organizations at various levels of management.

Keywords: *laws of organization theory, methods for describing socio-economic systems, forecasting economic indicators, growth rates in the volume of work of enterprises and organizations.*

DOI: 10.14357/20790279240302 **EDN:** EDBQPO

References

1. Antonov V.G. Evolution of organizational structures // Management in Russia and abroad. 2000. No. 1. P. 25-31.
2. Bertalanffy Von L. General systems theory - a critical review. // Research on general systems theory. M.: Progress. 1969. P. 23 – 82.
3. Bertalanffy Von L. History and status of general systems theory // System Research: Yearbook. M.: Nauka. 1973. P. 20 – 37.
4. Bogdanov A.A. Tectology: General organizational science. In 2 books. Edited by L.I. Abalkin (responsible editor) and others. Moscow: Economica. 1989. 304 p., 352 p.
5. Wissema H. Strategic management and entrepreneurship: opportunities for future prosperity / Transl. from English. M. 2000. 271 p.
6. Gvishiani D.M. Materialistic dialectics is the philosophical basis of systemic research. - System research: Methodological problems. Yearbook. – M.: Nauka. 1979. P. 7-26.
7. Gvishiani D.M. Organization and management. – Ed. 3rd revision. M. Publishing house of MSTU im. M.E. Bauman. 1998. 331 p.
8. Gibson J.L., Ivantsevich D.M., Donnelly D.H. Organizations: behavior, structure, processes / Transl. from English. M. 2000. 660 p.
9. Gleick J. Chaos: The Creation of a New Science / Transl. from English M. Nakhmansona, E. Barashkova. St. Petersburg: Amphora. 2001. 396 p.
10. Kleiner G.B. Production functions: Theory, methods, application. M.: Finance and Statistics. 1986. 239 p.
11. Kogan M.S. System and structure // System research: Methodological problems: Yearbook. 1983. M.: Nauka. 1983. P. 88 – 105.
12. Livshits V.N., Orlova E.R., Tishchenko T.I., Myznikova M.N. A new model for the development of the Russian economy // Proceedings of ISA RAS. 2021. Volume 71. No. 3. P. 3–10.
13. Milner B.Z. Problems of research and development of organizational systems // System research: Methodological problems. Yearbook 1984. M. 1984. P. 207-227.
14. Milner B.Z., Evenko L.I., Rapoport V.S. Systematic approach to the organization of management. M., 1983. 224 p.
15. Naydenov I.V. Organization theory. – Kaliningrad: KSU Publishing House. 2004. 195 p.
16. Naydenov I.V. Forecasting the development of open socio-economic systems in transport. M.: Finstatinform. 2001. 248 p.
17. Naydenov I.V., Naydenova G.I. Forecasting transport performance indicators in the Kaliningrad special economic zone // Proceedings of ISA RAS. 2018. Volume 68. No. 3. P. 91–96.
18. Ostreykovsky V.A. Theory of systems: Textbook. for universities for special purposes “Autom. syst. arr. information and ex.” M.: Higher. school. 1997. 239 p.
19. Temnikov F.E. Questions of theory and methodology of systems // Tr. Moscow energy Institute: Systems Engineering. Vol. 158. M. 1973. P. 3-5.
20. Hall A.D. Experience in methodology for systems engineering. M.: Sov. radio. 1975. 448 p.

21. Graph theory: Algorithms for processing contourless graphs / V.A. Evstigneev, V.N. Kasyanov. Novosibirsk: Science. Sib. Enterprise RAS. 1998. 385 p.
22. Transport in Russia. 2022: Statistical Collection/ Rosstat. M. 2022. 101 p.
23. *Ivancevich J.M., Donnelly J.H.* Organizations: behavior, structure, processes. IRWIN. 1997. 845 c.
24. *Morgan G.* Images of organization. London. 1989. 520 c.
25. *Robbins P.S.* Organizational behavior: concepts, controversies, applications. San Diego State University. 1996. 875 c.

Naydenov Igor Vitalievich. Western branch of the Russian Presidential Academy of National Economy and Public Administration, Kaliningrad, Russia, Professor of the Center for Continuing Education. Professor, Doctor of Economics. Number of published works: over 100 (including 3 monographs). Research interests: analysis of economic processes based on the use of economic and mathematical methods and tools, study of economic systems at the regional level as management objects, systemic transformations in the region, formation and systemic transformations in the transport complex of the region. E-mail: naydenovigor@yandex.ru

Задача генерации выходных форм документов информационных систем

А.В. СОЛОВЬЕВ

Федеральный исследовательский центр «Информатика и управление»
Российской академии наук», г. Москва, Россия

Аннотация. В статье формализована задача генерации выходных (отчетных) форм документов по динамически изменяющимся структурам хранения данных в информационных системах. Предложен подход к автоматизации процесса генерации выходных форм. Рассмотрены структуры данных, необходимые и достаточные для представления цифровизации информационных систем в выходных (отчетных) документах. Преимуществом предложенного подхода является упрощение и существенное ускорение процесса генерации выходных форм, который в настоящее время очень трудоемок.

Ключевые слова: генерация выходных форм документов, структуры хранения данных, структуры данных выходных (отчетных) документов.

DOI: 10.14357/20790279240303 **EDN:** GMDPNM

Введение

В современном мире, охваченном процессами цифровой трансформации, когда, например, цифровые формы становятся ключевым фактором производства (см. [1]), а информационные системы, использующие и обрабатывающие эти данные, становятся неотъемлемой частью производства, крайне важно удержать инициативу и технологическое первенство. В частности, что касается разработки информационных систем (ИС), крайне важной задачей является их быстрая разработка и модификация. Дело в том, что в момент существенных и быстрых цифровых преобразований экономики, перехода на новые технологии, вопрос времени становится крайне важным. В такой ситуации невозможно разрабатывать и модернизировать ИС годами. Скорость разработки и модификации будет решающей в конкурентной борьбе за технологическое преимущество и, как следствие, за устойчивость и конкурентоспособность экономики в целом (см, например, [2-4]).

Конечно, вопросы ускорения разработки и модификации ИС крайне обширны. В данной статье автор рассматривает только два аспекта, а именно:

- быстрое конструирование ИС и структур хранения данных на основе форм входных документов;
- генерация выходных (отчетных) форм документов по динамически изменяющимся структурам хранения данных.

Основной упор в статье сделан на втором аспекте, т.к. первый аспект в настоящее время так или иначе решен, чему посвящен далее краткий обзор.

Ускорение же второго аспекта до настоящего времени остается задачей нерешенной. Формализации данной задачи, рассмотрению структур данных выходных документов, подходу к решению задачи автоматизации генерации выходных форм документов, посвящена данная статья.

1. Краткий обзор источников и проблем быстрой разработки информационных систем

Из краткого обзора литературы следует, что проблема быстрого создания ИС крайне актуальна, решения по созданию конструкторов ИС создавались и создаются. При этом используется огромное количество разных подходов, вплоть до использования методов искусственного интеллекта (ИИ) для создания хранилищ данных для неструктурированных и слабоструктурированных документов и массивов информации.

Только обзор работ за последние годы позволяет создать следующую картину решений и направлений исследования.

Важным направлением при цифровизации экономики является автоматизация процесса соз-

дания гибких производственных системы (см., например, [5]). В частности, в работах описан процесс автоматизации создания структур хранения 2D и 3D данных (баз данных, БД) для таких систем. Используемый подход основан на том, что процесс создания ИС строится от структур входных данных, на основании которых создаются структуры реляционной БД под конкретные производственные задачи.

В работах [6, 7] представлен альтернативный подход, согласно которому создается платформенное решение, позволяющее настраиваться на разные структуры БД и извлекать из них необходимую информацию, в том числе на основе онтологий (см. [7]). Ограничением данного подхода является то, что он может использоваться в достаточно узких предметных областях, где известны заранее параметры, которые требуется извлечь из БД для выполнения функций ИС. Более того, параметры БД, как правило, стандартизованы нормативными документами. Однако такой подход, безусловно, имеет право на существование и практическую ценность.

Отдельных исследований требуют подходы, связанные с автоматизацией процесса построения структур хранения данных в ИС на основе слабоструктурированных и неструктурированных документов. Магистральным направлением в настоящее время здесь является использование методов ИИ для создания БД таких ИС (см., например, [8, 9]). Подход достаточно перспективный, но требующий создания достаточно сложных структур метаданных для своей реализации и нечеткого поиска реквизитов в неструктурированных текстах.

Нельзя не отметить также наличие подходов, связанных с использованием RAD-технологий (RAD – Rapid Application Development) и CASE-средств (см., например, [10-12]). Хотя сами концепции использования RAD и/или CASE возникли еще в конце 80-х годов XX века. Эта концепция до сих пор остается достаточно востребованной. Однако ограничения CASE-средств, в том числе отсутствие универсальности и дороговизна, ресурсоемкость RAD-технологий создает определенные сложности с их массовым внедрением. К тому же идея автоматической генерации программного кода (программирование без программирования) в настоящее время также имеет свои ограничения. Одно дело создать автоматически программы по заданному алгоритму. С этим сейчас справляются средства генеративного интеллекта типа ChatGPT, другое дело автоматическая генерация программ по заданным параметрам бизнес-процессов, требу-

ющих универсальности описания. Хотя здесь тоже есть заметные успехи, несмотря на то, что предложенный в [13] подход натывается на ограничения самой методологии IDEF0, например, отсутствие ветвлений и условных переходов.

И, наконец, еще одним подходом к решению проблемы быстрого создания ИС и структур их БД является разработка достаточно универсальных платформенных систем (см., например, [14, 15]). Достоинством такого является достаточно универсальный продукт, который можно применять в различных предметных областях. Недостатками являются, как правило, большой объем ИС, наличие избыточной функциональности в рамках конкретных применений, недостаточная гибкость для настройки под вкусы и требования конкретных пользователей.

Необходимо отметить, что в Отделе 94 ФИЦ ИУ РАН также был разработан прототип решения для быстрого создания ИС на основе подхода быстрого построения и модификации структур БД ИС на основе входных форм документов и создания платформенного универсального решения. Подробному описанию разработанного прототипа решения будет посвящена отдельная статья.

В качестве вывода для данного раздела необходимо отметить, что в рамках перечисленных подходов в целом решена проблема автоматизации создания и модификации структур БД на основе входных документов и разработки достаточно универсального платформенного решения. Однако до сих пор нерешенной остается задача автоматизации изменения форм выходных документов, созданных в ИС на основе структур БД. В случае изменения структур БД требуется изменения форм выходных документов (отчетов) вручную.

Эта задача не решена, даже несмотря на очень мощное развитие средств генерации отчетов и BI-систем. Эти системы позволяют создавать сложные отчеты с развитой графикой, обобщением данных, снабжены инструментами анализа данных (диаграммы, сводные отчеты, кросс-табличные отчеты и проч.). Однако изменение структур данных требует их перестроения оператором вручную. По мнению автора, для успешного решения задачи необходимо ее правильно формализовать, а также проанализировать структуры данных, необходимые для представления в выходных документах. Этим вопросам и посвящена данная статья.

2. Формализация задачи исследования

В общем виде задачу исследования можно сформулировать так. На вход информационной

системы поступает множество входных данных. В общем случае входные данные имеют структуру явно или неявно выраженную в их форме. ИС при вводе данных представляет их в виде некоторой структуры, чаще выраженной шаблонами входных форм данных, предназначенных для ввода. По формам входных данных ИС осуществляет создание структур для их хранения в своей БД. При изменении входных данных, изменяются их формы, после чего происходит автоматическое или автоматизированное перестроение структур хранения данных. Задача перестроения структур хранящихся данных, согласно выше проведенному обзору, в целом решена. Затем по хранящимся данным, соответственно их структурам хранения, создаются формы выходных документов ИС (отчетные формы). Необходимо разработать механизмы, методы и средства автоматического (автоматизированного) преобразования форм выходных документов ИС при изменении структур хранения данных. Задача такого преобразования пока не решена, что приводит к тому, что на преобразование форм (шаблонов) выходных документов тратится много времени при сопровождении и эксплуатации ИС.

В формализованном виде задачу исследования можно представить так.

Дано:

- 1) Множество входных данных ИС $D = \{d_i\}$.
- 2) Множество форм входных документов ИС для ввода данных $FID = \{fid_i\}$.
- 3) Множество структур хранения данных ИС $DSS = \{dss_i\}$.
- 4) Множество, хранящихся согласно DSS , данных ИС $ISD = \{isd_i\}$.
- 5) Множество способов автоматического преобразования FID в DSS : $IAT: FID \rightarrow DSS$.
- 6) Множество способов автоматического преобразования D в ISD : $IDAT: D \rightarrow ISD$.
- 7) Множество форм выходных документов ИС $FOD = \{fod_i\}$.
- 8) Множество создаваемых по FOD выходных документов $OD = \{od_i\}$.

Найти:

- 1) Множество способов автоматического преобразования DSS в FOD : $OAT: DSS \rightarrow FOD$.
- 2) Множество способов автоматического преобразования ISD в OD : $ODAT: ISD \rightarrow OD$.

Для того, чтобы решить поставленную задачу исследования, необходимо для начала проанализировать структуры данных DSS и FOD .

Структуры хранения данных, как правило, представлены в виде плоских двумерных реляционных таблиц, связанных отношениями 1:N, 1:1,

N:N. Как правило, тип отношений 1:N существенно превалирует над остальными.

Анализ же структур данных FOD требует более детального изучения.

3. Анализ структур данных выходных документов и их форм

Выходной документ (od_i) – это, как правило, отчет, представленный на экране или напечатанный на бумаге. Все современные генераторы отчетов ориентированы на представление данных на плоском бумажном носителе. Для этого в них реализована возможность размещать структуры данных на листах бумаги, как в обычных текстовых редакторах офисных приложений.

В таком случае будет справедливым утверждение о представлении такого выходного документа (od_i) и его формы (шаблона) (fod_i) в виде совокупности семантических блоков, как это сделано в работе автора [16]. Т.е. $OD = \{od_i\}$, $od_i = \bigcup_{(j=1..N)} (B_j)$, каждому od_i из множества OD поставим в соответствие некоторый граф $G(od_i) = (V, E)$, представляющий схему его семантических блоков B_j . V – множество вершин графа, E – множество дуг. $V = \{B_j\}$. $E = \{(B_j, B_k): B_j \rightarrow B_k\}$. Аналогично представим и fod_i .

Будем считать, что od_i (fod_i) разбивается на семантические блоки, если:

- $B_j \cap B_k \neq \emptyset$, то $B_j \subseteq B_k$ или $B_k \subseteq B_j$;
- $\exists B_k \subseteq B_j$, то $B_j = \bigcup B_k$ для всех $B_k \subseteq B_j$.

Семантический блок B_j называется простым (или терминальным), если он не содержит никаких внутренних блоков. На графе семантических блоков $G(od_i)/G(fod_i)$ – это листовые (не имеющие потомков) вершины V .

Назовем рангом $r(B_j)$ семантического блока B_j длину максимальной цепочки вложенных блоков с корнем в B_j . Т.е. $r(B_j) = \max \{m: \exists \{B_{j_1}, \dots, B_{j_m}\}, \text{ что } B_{j_m} \subset \dots \subset B_{j_1} \subset B_j\}$. Если B_j простой, то $r(B_j) = 0$. Иными словами, на графе $G(od_i)$ – это длина ветки с корнем в B_j «минус» единица. С помощью признака $r(B_j) = 0$ удобно пометать листовые вершины (простые блоки).

Назовем повторяемостью $s(B_j)$ семантического блока B_j количество повторений (возможно с разными данными) в od_i . Если блок не повторяется, $s(B_j) = 0$. Т.к. при преобразовании ($ODAT: ISD \rightarrow OD$) выполняется запрос к БД, то, по сути, количество повторов – это количество возвращаемых строк запроса (результат запроса к реляционной БД – всегда плоская двумерная таблица) «минус» 1.

Выходные (отчетные) документы OD и формы FOD , на основании которых эти документы

создаются, представляют данные, хранящиеся в ИС согласно **DSS**, в виде гладкого текста, таблиц, нумерованных или маркированных списков, изображений, рисунков, графических фигур. Частным случаем текстового представления также являются сноски и ссылки.

При этом гладкий текст – это последовательность входящих в него данных B_j . Таблицы – табличное представление данных B_j . Нумерованные и маркированные списки – иерархическое (в общем случае) представление данных B_j . Изображения, рисунки, графические элементы и прочее – контурное представление B_j , характеризующиеся областью изображения с контуром произвольной формы. Все остальные представления данных B_j можно характеризовать как смешанное представление.

Тогда можно утверждать, что все данные B_j представимы в структурах **OD** и **FOD** в виде:

- последовательностей;
- таблиц;
- иерархий;
- контуров;
- смешанное (из указанных выше четырех видов представлений).

Можно утверждать также, что:

- B_j представим в виде последовательности, если $r(B_k) = 0$ для всех $B_k \subseteq B_j$;
- B_j представим в виде таблицы, если $s(B_k) > 0$ для всех $B_k \subseteq B_j$;
- B_j представим в виде иерархии, если $r(B_j) > 0$.

На основании приведенных рассуждений можно утверждать о применимости методов теории множеств, теории графов для организации структур данных **OD** и **FOD**, нацеленных на независимость от реализации БД средствами конкретной СУБД, использующейся в ИС.

Теперь перейдем к описанию подхода к автоматизации процесса генерации выходных форм документов **FOD** и выходных документов **OD** по этим формам.

4. Подход к автоматизации процесса генерации выходных форм и документов

Автоматизацию процесса генерации выходных форм (**FOD**) и документов на их основе (**OD**) логично разбить на два этапа.

Первый этап – собственно автоматизация генерации выходных форм (**FOD**) на основе динамически меняющихся структур хранения данных (**DSS**).

Второй – генерация выходных документов (**OD**) на основе изменившихся форм (**FOD**).

Второй этап гораздо проще поддается автоматизации, т.к. в настоящее время существует достаточно большое количество программных средств, позволяющих создавать выходные документы (отчеты) на основе шаблонов. Такие средства относятся либо к классу генераторов отчетов, либо к классу BI-средств (BI, Business Intelligence), т.е. средств аналитической обработки информации.

Как генераторы отчетов, так и BI-средства обладают развитой функциональностью, позволяющей автоматизировать не только представление данных, хранящихся согласно структурам хранения (**DSS**), но и производить их обработку: суммирование и другие подсчеты, группировки, сортировки, графическое представление в виде графиков, диаграмм, кросс-таблиц и прочее. К средствам генераторов отчетов и BI-средствам можно отнести: Oracle BI Publisher, Crystal Reports, SQL Server Reporting Services, Microsoft Power BI, Qlik, Icube, CyberQuery, 1C:Аналитика и др. К сожалению, формат статьи не позволяет сделать полный обзор данных средств, тем более, что подобные обзоры периодически появляются (см., например, [17]).

В целом же отметим, что в зависимости от нужной функциональности можно использовать то или иное средство при создании ИС. В целом же можно отметить, что указанные выше средства в целом решают проблему поиска способов автоматического преобразования **ISD** в **OD**: **ODAT: ISD → OD**.

Совсем по-другому обстоит дело с первым этапом, т.е. автоматизацией генерации выходных форм (**FOD**) на основе динамически меняющихся структур хранения данных (**DSS**). В настоящее время не существует таких средств, поэтому нужен подход к поиску **OAT: DSS → FOD**.

Если чуть шире описать проблему динамически изменяющихся данных в информационной системе, опирающейся на входные данные (**D**) и формы входных данных (**FID**), то общий алгоритм функционирования ИС будет выглядеть примерно так:

- изменение структуры входных данных (**D**);
- корректировка форм входных документов ИС для ввода данных (**FID**);
- проверка соответствия **FID** и **DSS**;
- автоматическое преобразование **FID** в **DSS**: **IAT: FID → DSS**;
- сохранение истории изменений **DSS**;
- автоматическое преобразование **D** в **ISD**: **IDAT: D → ISD**;
- проверка соответствия **FOD** и **DSS**;
- автоматическое преобразование **DSS** в **FOD**: **OAT: DSS → FOD**;

– автоматическое преобразование *ISD* в *OD*: *ODAT: ISD* → *OD*.

Поскольку выше определена структура данных выходных документов и их форм (*OD* и *FOD*), как граф их семантических блоков B_j , то логично представлять модель соответствия *FOD* и *DSS* как графовую математическую модель. В каждом узле (B_j) этой модели хранится информация о соответствии структур *FOD* и *DSS*, а также метаданные о представлении узлов (B_j) в выходном документе (*OD*), т.е. полиграфическая спецификация B_j . Назовем такую модель – моделью представления *FOD*.

Тогда ИС при изменении структур хранения данных *DSS* (выполнения преобразования *IAT: FID* → *DSS*) выполняет проверку соответствия структур *FOD* и *DSS* с помощью промежуточной модели представления *FOD*. При несовпадении структур осуществляется просмотр истории изменений *DSS*. Потом на основании ретроспективы изменений следует анализ возможных замен имен, изменений размеров и топов полей структуры *DSS*. Возможно, для анализа истории изменений понадобятся методы, использующиеся в системах искусственного интеллекта, такие как нечеткий поиск [18], т.к. в его основе графовые модели представления данных для поиска аналогов в структурах *DSS*.

Поскольку моделью представления *FOD* – это графовая модель, узлами которой могут быть множества, можно утверждать о применимости методов теории множеств и графов, комбинаторного поиска как для организации структур данных *FOD*, так и для организации поиска изменений и проверки соответствий между *FOD* и *DSS*. Кроме того, преобразование графов промежуточной модели также, несомненно, будет лежать в основе способов автоматического преобразования *DSS* в *FOD*: *OAT: DSS* → *FOD*.

Заключение

В данной статье выполнен краткий обзор источников и проблем быстрой разработки информационных систем. Формализована задача генерации выходных (отчетных) форм документов по динамически изменяющимся структурам хранения данных в информационных системах.

Предложены и рассмотрены структуры данных, необходимые и достаточные для представления цифровых данных информационных систем в выходных (отчетных) документах. Сделан вывод о применимости методов теории множеств и графов, комбинаторного поиска как для организации структур данных и выходных документов, так и для их хранения.

Предложен подход к автоматизации процесса генерации выходных форм. Показана необходимость использования промежуточной графовой модели для анализа соответствия структур данных хранения и форм выходных документов и автоматизации генерации выходных форм. Преимуществом предложенного подхода является упрощение и существенное ускорение процесса генерации выходных форм, который в настоящее время достаточно трудоемок из-за необходимости проведения проверки и выполнения модернизации форм выходных документов вручную.

Литература

1. Программа «Цифровая экономика Российской Федерации». Утверждена распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р. М.: 2017. 88 с.
2. Bartodziej C.J. The Concept Industry 4.0 // Springer Fachmedien Wiesbaden: Wiesbaden, Germany. 2017. P. 27–50.
3. Schilz A., Rehbein M. Digitization of Cultural Heritage // Handbook Industry 4.0: Law, Technology, Society; Springer: Berlin, Heidelberg. 2022. P. 1193–1212.
4. Aheleroff S., Huang H., Xu X., Zhong R.Y. Toward sustainability and resilience with Industry 4.0 and Industry 5.0 // Frontiers in Manufacturing Technology. 2022. 2. 951643. <https://doi.org/10.3389/fmtec.2022.951643>.
5. Abdullayev G. Algorithmically and programming management of design process of flexible manufacture system // Системная инженерия и информационные технологии. 2021. Т. 3, № 1 (5). С. 20–28.
6. Арзучинцев А.В., Кедрин В.С., Кедрина М.С. Архитектура иерархически модифицируемо-пересекающейся базы данных экологических параметров // Вестник Бурятского государственного университета. Математика, информатика. 2022. №1. С. 3-17.
7. Taberko V., Ivaniuk D., Zotov N., Orlov M., Pupena O., Lutska N. Principles of building a system for automating the activities of a process engineer based on an ontological approach within the framework of the Industry 4.0 concept // Открытые семантические технологии проектирования интеллектуальных систем. 2021. № 5. С. 209-218.
8. Вотинова Е.С., Чуприна С.И. Автоматизация построения баз данных из неструктурированных текстовых документов // Математика и междисциплинарные исследования – 2021. [Электронный ресурс]: материалы Всероссийской научно практической конференции молодых

- ученых с международным участием (г. Пермь. 18–20 октября 2021 г.) / гл. ред. А. П. Шкарапу-та; Пермский государственный национальный исследовательский университет. Пермь, 2021. – С. 119-122. URL: www.psu.ru/files/docs/science/books/sborniki/mmi-2021.pdf. (дата обращения 15.03.2024).
9. Пальмов С.В., Шамрин Д.А. Функциональная структура «Фермы данных» // Евразийское Научное Объединение. 2021. № 5-1 (75). С. 42-45. DOI: 10.5281/zenodo.4926581.
 10. Кондратьев В.Ю., Кутыкова Е.М. Прототипное проектирование ЭИС (RAD-технология) // Информационное общество: современное состояние и перспективы развития. Сборник материалов XIV международного форума. Краснодар. 2021. С. 429-431.
 11. Lyashenko I.I. About the use of CASE-technologies in the process of designing information systems // Вестник Инновационного Евразийского университета. 2022. № 2 (86). С. 126-133.
 12. Лайпанов А.А., Гебенев Р.М. Проектирование информационных систем с использованием CASE-технологий // Тенденции развития науки и образования. 2021. № 80-2. С. 118-121. DOI: 10.18411/trnio-12-2021-89.
 13. Басманов С.Н., Вакалюк А.А. Разработка подхода к созданию гибкого пользовательского интерфейса на основе преобразования IDEF0-диаграммы // Современные наукоемкие технологии. 2020. № 5. С. 20-25.
 14. Черныш Б.А., Картамышев А.С. Разработка ядра интегрированной информационной системы // Программные продукты и системы (Software & Systems). 2021. № 2. С. 237-244. DOI: 10.15827/0236-235X.134.237-244.
 15. Автоматизированная система управления предприятием // Рули24. [Электронный ресурс]: URL: <https://ruli24.ru/> (дата обращения 15.03.2024).
 16. Соловьев А.В. Математическая модель электронного документа долговременного хранения // Информационные технологии и вычислительные системы. №2. 2022. С.30-36. DOI: 10.14357/20718632220204.
 17. King Tim. The 19 Best Data Catalog Tools and Software for 2024 // Solutions Review. 2024. URL: <https://solutionsreview.com/data-management/the-best-data-catalog-tools-and-software/> (дата обращения 19.03.2024).
 18. Kang Bo-Yeong, Dae-Won Kim, Hae-Jung Kim. Fuzzy Information Retrieval Indexed by Concept Identification // Lecture Notes in Computer Science, vol. 3658, Springer Berlin / Heidelberg. 2005. P. 179–186. DOI:10.1007/11551874_23.

Соловьев Александр Владимирович. Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия. Главный научный сотрудник. Доктор технических наук. Область научных интересов: системный анализ, системы управления базами данных, теория надежности, математическое моделирование, долговременное хранение электронных документов. E-mail: soloviev@isa.ru

The problem of generating output forms of information systems documents

A.V. Solovyev

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, Moscow, Russia

Abstract. The article formalizes the task of generating output (reporting) forms of documents based on dynamically changing data storage structures in information systems. An approach to automating the process of generating output forms is proposed. The data structures necessary and sufficient for representing digital data of information systems in output (reporting) documents are considered. The advantage of the proposed approach is the simplification and significant acceleration of the process of generating output forms, which is currently quite labor-intensive.

Keywords: *generation of output document forms, digital data, data structures of output (reporting) documents.*
DOI: 10.14357/20790279240303 **EDN:** GMDPNM

References

1. The Government of the Russian Federation. 2017. PROGRAM "Digital Economy of the Russian Federation" approved by the order of at July 28, 2017. № 1632-р: 88.
2. *Bartodziej C.J.* 2017. The Concept Industry 4.0. Springer Fachmedien Wiesbaden: Wiesbaden, Germany, pp. 27–50.
3. *Schilz A., Rehbein M.* 2022. Digitization of Cultural Heritage. Handbook Industry 4.0: Law, Technology, Society; Springer: Berlin, Heidelberg, pp. 1193–1212.
4. *Aheleroff S., Huang H., Xu X., Zhong R.Y.* 2022. Toward sustainability and resilience with Industry 4.0 and Industry 5.0. *Frontiers in Manufacturing Technology*, 2, 951643. <https://doi.org/10.3389/fmtec.2022.951643>.
5. *Abdullayev G.* 2021. Algorithmically and programming management of design process of flexible manufacture system. *Systems Engineering and Information Technology*, vol. 3, № 1 (5): 20–28.
6. *Arguchintsev A.V., Kedrin V.S., Kedrina M.S.* 2022. Architecture of a hierarchically modified-intersecting database of environmental parameters. *Bulletin of Buryat State University. Mathematics, computer science*, №1: 3-17.
7. *Taberko V., Ivaniuk D., Zotov N., Orlov M., Pupena O., Lutska N.* 2021. Principles of building a system for automating the activities of a process engineer based on an ontological approach within the framework of the Industry 4.0 concept // Open semantic technologies for designing intelligent systems. № 5: 209-218.
8. *Votinova E.S., Chuprina S.I.* 2021. Automation of database construction from unstructured text documents. *Mathematics and interdisciplinary research - 2021: materials of the All-Russian scientific and practical conference of young scientists with international participation* (Perm, October 18–20, 2021) / Ch. ed. A. P. Shkaraputa; Perm State National Research University. Perm: 119-122. URL: www.psu.ru/files/docs/science/books/sborniki/mmi-2021.pdf. (Accessed 15.03.2024).
9. *Palmov S.V., Shamrin D.A.* 2021. Functional structure of the Data Farm. *Eurasian Scientific Association*, № 5-1 (75): 42-45. DOI: 10.5281/zenodo.4926581.
10. *Kondratyev V.Yu., Kutykova E.M.* 2021. Prototype design of EIS (RAD technology). *Information society: current state and development prospects. Collection of materials from the XIV International Forum*. Krasnodar: 429-431.
11. *Lyashenko I.I.* 2022. About the use of CASE-technologies in the process of designing information systems. *Bulletin of the Innovative Eurasian University*, № 2 (86): 126-133.
12. *Laipanov A.A., Gebenov R.M.* 2021. Design of information systems using CASE technologies. *Trends in the development of science and education*, № 80-2: 118-121. DOI: 10.18411/trnio-12-2021-89.
13. *Basmanov S.N., Vakalyuk A.A.* 2020. Developing an approach to creating a flexible user interface based on IDEF0 diagram transformation. *Modern science-intensive technologies*, № 5: 20-25.
14. *Chernysh B.A., Kartamyshev A.S.* 2021. Development of the core of an integrated information system. *Software & Systems*, № 2: 237-244. DOI: 10.15827/0236-235X.134.237-244.
15. Automated enterprise management system. Ruli24. URL: <https://ruli24.ru/> (Accessed 15.03.2024).
16. *Solovyev A.V.* 2022. Mathematical model of an electronic document for long-term storage. *Information technology and computing systems*, №2: 30-36. DOI: 10.14357/20718632220204.
17. *King Tim.* 2024. The 19 Best Data Catalog Tools and Software for 2024. *Solutions Review*. URL: <https://solutionsreview.com/data-management/the-best-data-catalog-tools-and-software/> (Accessed 19.03.2024).
18. *Kang Bo-Yeong, Dae-Won Kim, Hae-Jung Kim.* 2005. Fuzzy Information Retrieval Indexed by Concept Identification. *Lecture Notes in Computer Science*, vol. 3658, Springer Berlin / Heidelberg, pp. 179–186, DOI:10.1007/11551874_23.

Solovyev A.V. Chief Researcher, Doctor of Technical Sciences. Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia. E-mail: soloviev@isa.ru

Динамические системы

Решение дифференциально-алгебраических систем уравнений с помощью аппроксимации Паде матричной экспоненты

Ю.А. БУРЦЕВ

Южно-Российский государственный политехнический университет
им. М.И. Платова, г. Новочеркасск, Россия

Аннотация. Рассмотрено применение новых высокоточных численных методов на основе матричной экспоненты к решению линейных дифференциально-алгебраических систем уравнений. Расчет состояния системы на каждом шаге интегрирования в зависимости от порядка метода сводится к решению одной или нескольких систем линейных алгебраических уравнений. Методы основаны на разложении аппроксимации Паде матричной экспоненты на простейшие дроби. Предложенные формулы позволяют исключить преобразование дифференциально-алгебраической системы уравнений в систему обыкновенных дифференциальных уравнений на этапе решения задачи. Новые методы отличаются простотой и требуют в несколько раз меньше вычислительной работы, чем методы типа Рунге-Кутты, эквивалентные по области устойчивости и точности.

Ключевые слова: численные методы, дифференциально-алгебраические системы уравнений, матричная экспонента, аппроксимация Паде.

DOI: 10.14357/20790279240304 **EDN:** НКТОУЕ

Введение

В предыдущей публикации [1] представлено семейство новых A -устойчивых и L -устойчивых [2] высокоточных численных методов решения классической задачи Коши для линейных систем обыкновенных дифференциальных уравнений (СОДУ). Однако эти же методы с помощью несложных преобразований можно применять и непосредственно для решения дифференциально-алгебраических систем уравнений (ДАСУ), которые часто встречаются на практике. В этих системах имеются линейные дифференциальные уравнения первого порядка, а также линейные алгебраические уравнения. Предлагаемые методы для преобразования ДАСУ в СОДУ используют варианты метода пространства состояний и метода ε -вложения [2, с. 414-422], при-

способленные для новых методов решения СОДУ. Но эти преобразования нужны лишь на этапе вывода уравнений и формул. Окончательные формулы для решения ДАСУ имеют понятную простую структуру и выводятся один раз.

1. Анализ состояния проблемы. Актуальность темы

Самые простые и часто используемые способы привести СОДУ или ДАСУ к системе линейных алгебраических уравнений (СЛАУ) заключаются в применении конечно-разностных формул неявного метода Эйлера, метода трапеций или метода 2-го порядка на основе формул дифференцирования назад (ФДН). Однако эти методы имеют невысокую точ-

ность, а два последних – особенно большие погрешности дают при решении жестких, колебательных и жестко-колебательных систем. Между тем, такие системы часто встречаются на практике. Методы на основе ФДН с порядком больше 2 и явные методы без применения специальных приемов имеют ограничения на шаг интегрирования, что неудобно во многих практических случаях. Для решения ДАСУ и СОДУ классическими неявными методами высоких порядков точности (например, методами типа Рунге-Кутты) проблему представляют большие вычислительные затраты, а также сложность теории и алгоритмов. Для численно-аналитических методов необходима сложная процедура преобразования ДАСУ в СОДУ. Известные компьютерные программы для расчетов технических систем нередко дают решения с большими погрешностями или даже качественно неверные [3, 4].

Очень обстоятельные исследования в области ДАСУ были проделаны Ю. Е. Бояринцевым и его коллегами в работах [5 – 10] и других. Для решения ДАСУ предложен эффективный метод на основе так называемых базовых матриц, однако этот метод требует сложных дополнительных построений.

Глубокий анализ различий свойств и методов решения ДАСУ и СОДУ на примерах неявного метода Эйлера и методов на основе ФДН представлен в статье [11]. Отмечены многие еще нерешенные проблемы, особенно для ДАСУ с индексом 3 и выше. Сингулярно-возмущенные задачи образуют особый класс задач, содержащих параметр ε . Когда этот параметр мал, соответствующая СОДУ является жесткой. Когда ε стремится к нулю, СОДУ переходит в ДАСУ. Методы построения асимптотических приближений для сингулярно возмущенных задач оптимального управления предложены в работах [12, 13]. Однако эти методы используют очень сложный математический аппарат и ориентированы на специфические задачи теории управления. Анализ существующих методов решения СОДУ и ДАСУ также можно найти в работах [2, 3, 14].

Таким образом, создание простых высокоточных экономичных по вычислениям методов решения ДАСУ является актуальной задачей. Разработка новых методов была вызвана практической необходимостью расчетов электрических цепей [15].

2. Постановка задачи

Рассмотрим ДАСУ

$$\begin{bmatrix} \mathbf{V} \\ \mathbf{0} \end{bmatrix} \mathbf{x}' = \begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{w} \\ \mathbf{g} \end{bmatrix}, \quad (1)$$

где $\begin{bmatrix} \mathbf{V} \\ \mathbf{0} \end{bmatrix}$ и $\begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix}$ – некоторые постоянные в пре-

делах шага интегрирования квадратные матрицы одного размера, причем число строк матриц $\mathbf{V}$ и $\mathbf{P}$ равно числу элементов вектора $\mathbf{w}$; $\mathbf{x}$ – неизвестная вектор-функция времени t , $\mathbf{w}$ и $\mathbf{g}$ – заданные вектор-функции. Задача заключается в том, чтобы получить семейство простых высокоточных экономичных по вычислениям A -устойчивых и L -устойчивых численных методов для ее решения, исключая преобразование ДАСУ в СОДУ на этапе решения задачи.

Методы на основе диагональных аппроксимаций Паде матричной экспоненты [2, 16] (у которых степень числителя равна степени знаменателя) являются A -устойчивыми (не имеют ограничений на шаг интегрирования) [2]. Но они склонны к раскачке паразитных колебаний, если число точек интегрирования недостаточно для адекватного расчета жесткой части решения. Эти методы хорошо подходят для задач с длительными слабозатухающими колебаниями. Методы на основе субдиагональных аппроксимаций Паде матричной экспоненты (степень знаменателя на единицу больше степени числителя) являются L -устойчивыми (A -устойчивыми с затухающей до нуля на бесконечности функцией устойчивости) [1]. Они хорошо демпфируют паразитные колебания, хотя имеют некоторую методическую погрешность по затуханию решения [2]. Поэтому L -устойчивые методы более приспособлены для жестких и жестко-колебательных задач. Однако высокоточные (с порядком точности 3 и более) L -устойчивые методы годятся также для колебательных задач.

Будем считать, что матрица $\begin{bmatrix} \mathbf{V} \\ -\mathbf{Q} \end{bmatrix}$ обратима, матрица $\mathbf{A} = \begin{bmatrix} \mathbf{V} \\ -\mathbf{Q} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{P} \\ \mathbf{0} \end{bmatrix}$ имеет простую струк-

туру (N линейно независимых собственных векторов, где N – размерность матрицы), все собственные значения матрицы $\mathbf{A}$ различны и имеют отрицательную действительную часть, за исключением нулевого собственного значения кратности n , которому соответствуют n линейно независимых собственных векторов. Также примем, что функция $\mathbf{g}$ дифференцируема – далее мы будем рассматривать полиномиальное представление свободного члена уравнения. ДАСУ такого типа возникают в некоторых приложениях, в частности, при расчетах электрических цепей. Особые случаи и общие условия разрешимости задачи могут быть предметом дальнейших исследований.

Мы рассмотрим два способа решения – с помощью дифференцирования алгебраических уравнений и предельного перехода. Первый использует более очевидные преобразования уравнений и позволяет в случае необходимости в явном виде найти матрицу $\mathbf{A}$ эквивалентной СОДУ и ее собственные значения и таким образом ответить на вопрос об устойчивости задачи. Этот способ также дает возможность получить аналитические выражения для матричной экспоненты и аналитические решения. Второй способ по существу представляет собой вариант метода ε -вложения. Он дает более простые формулы для численного решения задачи.

3. Приведение ДАСУ к СОДУ с помощью дифференцирования алгебраических уравнений

Продифференцируем нижнюю матричную строку системы (1). Получим:

$$\begin{bmatrix} \mathbf{V} \\ -\mathbf{Q} \end{bmatrix} \mathbf{x}' = \begin{bmatrix} \mathbf{P} \\ \mathbf{0} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{w} \\ \mathbf{g}' \end{bmatrix}. \quad (2)$$

$$\text{Обозначим } \begin{bmatrix} \mathbf{V} \\ -\mathbf{Q} \end{bmatrix} = \mathbf{G}, \quad \begin{bmatrix} \mathbf{P} \\ \mathbf{0} \end{bmatrix} = \mathbf{H}_0, \quad \begin{bmatrix} \mathbf{w} \\ \mathbf{g}' \end{bmatrix} = \mathbf{f}^+,$$

тогда система (2) примет вид:

$$\mathbf{G}\mathbf{x}' = \mathbf{H}_0\mathbf{x} + \mathbf{f}^+. \quad (3)$$

Так как согласно принятому условию матрица $\mathbf{G}$ обратима, преобразуем систему (3) к СОДУ в форме Коши:

$$\mathbf{x}' = \mathbf{G}^{-1}\mathbf{H}_0\mathbf{x} + \mathbf{G}^{-1}\mathbf{f}^+. \quad (4)$$

В соответствии с ранее принятым обозначением $\mathbf{G}^{-1}\mathbf{H}_0 = \mathbf{A}$. Для СОДУ (4) с матрицей $\mathbf{A}$ применимы известные критерии существования и устойчивости решения, основанные на свойствах собственных значений матрицы $\mathbf{A}$. Однако в классической теории, как правило, не рассматриваются вырожденные матрицы СОДУ и нулевые собственные значения. Обратим внимание на то что нижняя часть матрицы $\mathbf{H}_0$ состоит из нулей. По этой причине среди собственных значений матрицы $\mathbf{A}$ также есть нуль, причем в общем случае его кратность больше 1.

Тем не менее, случай кратных чисто нулевых собственных значений для рассматриваемой матрицы простой структуры не представляет проблемы. Пусть $\lambda_1, \lambda_2, \dots, \lambda_m$ – ненулевые собственные значения матрицы $\mathbf{A}$, $\lambda_{m+1} = \lambda_{m+2} = \dots = \lambda_N = 0$ – нулевые собственные значения, $\mathbf{U}$ – фундаментальная матрица, столбцы которой $\mathbf{u}_k$ представляют собой собственные векторы матрицы $\mathbf{A}$. Тогда аналитическое выражение матричной экспоненты для СОДУ (4) имеет вид

тическое выражение матричной экспоненты для СОДУ (4) имеет вид

$$\exp(t\mathbf{A}) = \mathbf{U} \begin{bmatrix} \exp(\lambda_1 t) & 0 & \dots & 0 & 0 \\ 0 & \exp(\lambda_2 t) & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \mathbf{U}^{-1},$$

а решение однородной СОДУ (4) (при $\mathbf{f}^+ = 0$) можно представить в форме:

$$\mathbf{x}(t) = \sum_{k=1}^N c_k \mathbf{u}_k \exp(\lambda_k t) = \sum_{k=1}^m c_k \mathbf{u}_k \exp(\lambda_k t) + \sum_{k=m+1}^N c_k \mathbf{u}_k,$$

где c_k – некоторые коэффициенты, $\mathbf{c} = \mathbf{U}^{-1}\mathbf{x}(0)$. Так как согласно принятому условию $\text{Re}(\lambda_k) < 0$ при $1 \leq k \leq m$, то первая сумма стремится к нулю с ростом t . Вторая сумма от t не зависит. Таким образом, СОДУ (4) устойчива, но не асимптотически.

Дифференцирование алгебраических уравнений как способ приведения ДАСУ к СОДУ, известно. Однако при этом, как правило, переменные, которых нет под знаком производной, исключаются из системы [2, 14]. С одной стороны, это исключение уменьшает размер полученной СОДУ. С другой стороны, оно требует дополнительных и зачастую непростых преобразований системы и затем отдельного расчета исключенных переменных. Предлагаемый способ не делает такого исключения.

4. Решение ДАСУ после дифференцирования алгебраических уравнений

Будем строить решение на основе матричной экспоненты [2, 17]. Как это обычно делается, положим начальный момент времени $t_0 = 0$, тогда в общем случае переменная t может принимать любое значение от 0 до некоторого малого шага интегрирования Δt . При расчете последовательных шагов по времени будем иметь в виду вычисление $\mathbf{x}(\Delta t)$. Воспользуемся методом решения СОДУ на основе разложения аппроксимации Паде матричной экспоненты на простейшие дроби, предложенным в статье [1]. Вначале для краткости и ясности изложения рассмотрим L -устойчивую субдиагональную аппроксимацию третьего порядка точности R_{12} [1, 2, 16] (степень числителя 1, степень знаменателя 2):

$$\exp(t\mathbf{A}) \approx R_{12}(t\mathbf{A}) = (6\mathbf{E} - 4t\mathbf{A} + (t\mathbf{A})^2)^{-1}(6\mathbf{E} + 2t\mathbf{A}) = 2\text{Re}[y_1(t\mathbf{A} - z_1\mathbf{E})^{-1}],$$

где $z_1 = 2 - i\sqrt{2}$ – корень многочлена $6 - 4z + z^2$ (корень знаменателя), $y_1 = 1 + i\frac{5}{\sqrt{2}}$ – коэффи-

циент, $\mathbf{E}$ – единичная матрица. Затем представим решения для других аппроксимаций.

В работе [1] показано, что решение СОДУ в форме Коши

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t) \quad (5)$$

в случае, если функция в правой части в пределах каждого шага интегрирования $[0; \Delta t]$ (локально)

задана в виде многочлена $\mathbf{f}(t) = \sum_{m=0}^3 \mathbf{f}_m t^m$, с по-

мощью аппроксимации Паде матричной экспоненты R_{12} может быть представлено в виде:

$$\mathbf{x}(t) = 2\text{Re}[(t\mathbf{A} - z_1\mathbf{E})^{-1}(y_1\mathbf{x}(0) + t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1 t + a_{12}\mathbf{f}_2 t^2 + a_{13}\mathbf{f}_3 t^3))], \quad (6)$$

$$\text{где } a_{10} = -\frac{1}{2} + i\sqrt{2}; a_{11} = -\frac{1}{2} + \frac{1}{2\sqrt{2}}i; a_{12} = -\frac{1}{2};$$

$$a_{13} = -\frac{1}{2} - \frac{1}{2\sqrt{2}}i.$$

Переходя от СОДУ (5) к СОДУ (4), подставим в формулу (6) вместо матрицы $\mathbf{A}$ матрицу $\mathbf{G}^{-1}\mathbf{H}_0$ и вместо функции $\mathbf{f}(t)$ функцию $\mathbf{G}^{-1}\mathbf{f}^*(t)$, сделаем преобразования:

$$\begin{aligned} \mathbf{x}(t) &= 2\text{Re}[(t\mathbf{G}^{-1}\mathbf{H}_0 - z_1\mathbf{E})^{-1}(y_1\mathbf{x}(0) + \\ &+ \mathbf{G}^{-1}t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1 t + a_{12}\mathbf{f}_2 t^2 + a_{13}\mathbf{f}_3 t^3))] = \\ &= 2\text{Re}[(t\mathbf{G}\mathbf{G}^{-1}\mathbf{H}_0 - z_1\mathbf{G}\mathbf{E})^{-1}\mathbf{G}(y_1\mathbf{x}(0) + \\ &+ \mathbf{G}^{-1}t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1 t + a_{12}\mathbf{f}_2 t^2 + a_{13}\mathbf{f}_3 t^3))] = \\ &= 2\text{Re}[(t\mathbf{H}_0 - z_1\mathbf{G})^{-1}(y_1\mathbf{G}\mathbf{x}(0) + \\ &+ t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1 t + a_{12}\mathbf{f}_2 t^2 + a_{13}\mathbf{f}_3 t^3))]. \end{aligned}$$

Далее заметим, что согласно системе (1) выполняется равенство $\mathbf{Q}\mathbf{x} + \mathbf{g} = \mathbf{0}$ и поэтому

$$\begin{aligned} \mathbf{G}\mathbf{x}(0) &= \begin{bmatrix} \mathbf{V} \\ -\mathbf{Q} \end{bmatrix} \mathbf{x}(0) = \begin{bmatrix} \mathbf{V}\mathbf{x}(0) \\ -\mathbf{Q}\mathbf{x}(0) \end{bmatrix} = \begin{bmatrix} \mathbf{V}\mathbf{x}(0) \\ \mathbf{g}(0) \end{bmatrix} = \\ &= \begin{bmatrix} \mathbf{V} \\ \mathbf{0} \end{bmatrix} \mathbf{x}(0) + \begin{bmatrix} \mathbf{0} \\ \mathbf{g}(0) \end{bmatrix} = \mathbf{G}_0\mathbf{x}(0) + \mathbf{g}_0, \quad \mathbf{G}_0 = \begin{bmatrix} \mathbf{V} \\ \mathbf{0} \end{bmatrix}, \end{aligned}$$

$$\mathbf{g}_0 = \begin{bmatrix} \mathbf{0} \\ \mathbf{g}(0) \end{bmatrix}. \text{ По этой причине вместо вектора}$$

$-\mathbf{Q}\mathbf{x}(0)$, который в результате приближенного характера расчетов на очередном шаге интегрирования получается с некоторой погрешностью, для вычисления $\mathbf{G}\mathbf{x}(0)$ лучше взять точно известный вектор $\mathbf{g}(0)$. Этот прием к тому же сократит вычислительную работу. Для однородной системы $\mathbf{g}(0) = \mathbf{0}$, в этом случае вектор $\mathbf{g}_0$ не нужен. Итак, окончательно, для ДАСУ (1) вектор состояния системы $\mathbf{x}(t)$ с помощью аппроксимации матричной экспоненты R_{12} может быть получен по формуле:

$$\mathbf{x}(t) = 2\text{Re}[(t\mathbf{H}_0 - z_1\mathbf{G})^{-1}(y_1(\mathbf{G}_0\mathbf{x}(0) + \mathbf{g}_0) + t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1 t + a_{12}\mathbf{f}_2 t^2 + a_{13}\mathbf{f}_3 t^3))]. \quad (7)$$

5. Приведение ДАСУ к СОДУ предельным переходом и решение СОДУ

Обозначим $\begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix} = \mathbf{H}$, $\begin{bmatrix} \mathbf{w} \\ \mathbf{g} \end{bmatrix} = \mathbf{f}$, тогда система (1) примет вид:

$$\mathbf{G}_0\mathbf{x}' = \mathbf{H}\mathbf{x} + \mathbf{f}. \quad (8)$$

Матрица $\mathbf{G}_0$ необратима, поэтому добавим к ней некоторую матрицу $\varepsilon\mathbf{D}$, такую, чтобы матрица $\mathbf{G}(\varepsilon) = \mathbf{G}_0 + \varepsilon\mathbf{D}$ была обратима, здесь ε – малое число. По существу, таким образом мы применяем вариант метода ε -вложения [2]. Подставим в систему (8) вместо матрицы $\mathbf{G}_0$ матрицу $\mathbf{G}(\varepsilon)$ и преобразуем ее к СОДУ в форме Коши:

$$\mathbf{x}' = \mathbf{G}(\varepsilon)^{-1}\mathbf{H}\mathbf{x} + \mathbf{G}(\varepsilon)^{-1}\mathbf{f}. \quad (9)$$

Подобно тому как это сделано в п. 4, переходя от СОДУ (5) к СОДУ (9), подставим в формулу (6) вместо матрицы $\mathbf{A}$ матрицу $\mathbf{G}(\varepsilon)^{-1}\mathbf{H}$ и вместо функции $\mathbf{f}(t)$ функцию $\mathbf{G}(\varepsilon)^{-1}\mathbf{f}(t)$, сделаем преобразования для аппроксимации Паде матричной экспоненты R_{12} :

$$\begin{aligned} \mathbf{x}(t) &= 2\text{Re}[(t\mathbf{G}(\varepsilon)^{-1}\mathbf{H} - z_1\mathbf{E})^{-1}(y_1\mathbf{x}(0) + \\ &+ \mathbf{G}(\varepsilon)^{-1}t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1 t + a_{12}\mathbf{f}_2 t^2 + a_{13}\mathbf{f}_3 t^3))] = \\ &= 2\text{Re}[(t\mathbf{G}(\varepsilon)\mathbf{G}(\varepsilon)^{-1}\mathbf{H} - z_1\mathbf{G}(\varepsilon)\mathbf{E})^{-1}\mathbf{G}(\varepsilon)(y_1\mathbf{x}(0) + \\ &+ \mathbf{G}(\varepsilon)^{-1}t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1 t + a_{12}\mathbf{f}_2 t^2 + a_{13}\mathbf{f}_3 t^3))] = \\ &= 2\text{Re}[(t\mathbf{H} - z_1\mathbf{G}(\varepsilon))^{-1}(y_1\mathbf{G}(\varepsilon)\mathbf{x}(0) + \\ &+ t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1 t + a_{12}\mathbf{f}_2 t^2 + a_{13}\mathbf{f}_3 t^3))]. \end{aligned}$$

Окончательно, переходя к пределу при $\varepsilon \rightarrow 0$, получим для ДАСУ (1) вектор состояния системы $\mathbf{x}(t)$ в момент времени t :

$$\mathbf{x}(t) = 2\text{Re}[(t\mathbf{H} - z_1\mathbf{G}_0)^{-1}(y_1\mathbf{G}_0\mathbf{x}(0) + t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1 t + a_{12}\mathbf{f}_2 t^2 + a_{13}\mathbf{f}_3 t^3))]. \quad (10)$$

Таким образом, для расчета вектора состояния системы $\mathbf{x}(t)$ нужно решать комплексные СЛАУ с матрицей $t\mathbf{H}_0 - z_1\mathbf{G}$ или $t\mathbf{H} - z_1\mathbf{G}_0$ если матрицы ДАСУ (1) изменяются часто, либо умножать векторы на обратную матрицу $(t\mathbf{H}_0 - z_1\mathbf{G})^{-1}$ или $(t\mathbf{H} - z_1\mathbf{G}_0)^{-1}$, если матрицы ДАСУ (1) не меняются на большом количестве шагов интегрирования. При этом этап преобразования ДАСУ в СОДУ в ходе решения не нужен, используются исходные заданные матрицы.

Формулы (7) и (10) дают один и тот же результат, так как системы уравнений $\mathbf{Q}\mathbf{x} + \mathbf{g} = \mathbf{0}$ и $-\mathbf{Q}\mathbf{x}' = \mathbf{g}'$ имеют решения, которые могут отличаться только на некоторые константы, но они однозначно определяются из начальных условий. Вопрос обратимости матриц $t\mathbf{H}_0 - z_1\mathbf{G}$ и $t\mathbf{H} - z_1\mathbf{G}_0$

зависит от конкретного типа задач и в этой работе не исследуется. Вероятнее всего, для задач, относящихся к реальным физическим системам, эти матрицы обратимы.

6. Решение ДАСУ для других аппроксимаций Паде матричной экспоненты

Приведем теперь решения вида (10) для нескольких других конкретных аппроксимаций Паде, а также формулы для решения в общем виде. Для решений вида (7) формулы получаются аналогичным образом, однако они представляют скорее теоретический интерес, так как они сложнее формул вида (10). В рассмотренном выше примере функция $\mathbf{f}(t)$ была многочленом 3-й степени. В общем случае степень этого многочлена может быть любой, не превышающей порядок точности аппроксимации Паде матричной экспоненты, который равен сумме степеней числителя и знаменателя аппроксимации. Все формулы выводятся аналогично рассмотренному примеру с аппроксимацией матричной экспоненты R_{12} . Формулы для СОДУ вида (5) получены в статье [1].

Аппроксимация

$$R_{01}(t\mathbf{A}) = -(t\mathbf{A} - \mathbf{E})^{-1} = y_1(t\mathbf{A} - z_1\mathbf{E})^{-1}$$

(неявный метод Эйлера,

$$z_1 = 1, y_1 = -1, a_{10} = -1, a_{11} = -1):$$

$$\mathbf{x}(t) = (t\mathbf{H} - z_1\mathbf{G}_0)^{-1}(y_1\mathbf{G}_0\mathbf{x}(0) + t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1t)).$$

Аппроксимация

$$R_{11}(t\mathbf{A}) = (2\mathbf{E} - t\mathbf{A})^{-1}(2\mathbf{E} + t\mathbf{A}) = -\mathbf{E} + y_1(t\mathbf{A} - z_1\mathbf{E})^{-1}$$

(метод трапеций, $z_1 = 2, y_1 = -4, a_{10} = -2, a_{11} = -1, a_{12} = -1$):

$$\mathbf{x}(t) = -\mathbf{x}(0) + (t\mathbf{H} - z_1\mathbf{G}_0)^{-1}(y_1\mathbf{G}_0\mathbf{x}(0) + t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1t + a_{12}\mathbf{f}_2t^2)).$$

Аппроксимация

$$R_{22}(t\mathbf{A}) = (12\mathbf{E} - 6t\mathbf{A} + (t\mathbf{A})^2)^{-1}(12\mathbf{E} + 6t\mathbf{A} + (t\mathbf{A})^2) =$$

$$\mathbf{E} + 2\text{Re}[y_1(t\mathbf{A} - z_1\mathbf{E})^{-1}]$$

$$(z_1 = 3 - i\sqrt{3}, y_1 = 6 + i6\sqrt{3}, a_{10} = i2\sqrt{3},$$

$$a_{11} = -\frac{1}{2} + i\frac{\sqrt{3}}{2}, a_{12} = -\frac{1}{2} + i\frac{1}{2\sqrt{3}}, a_{13} = -\frac{1}{2}):$$

$$\mathbf{x}(t) = \mathbf{x}(0) + 2\text{Re}[(t\mathbf{H} - z_1\mathbf{G}_0)^{-1}(y_1\mathbf{G}_0\mathbf{x}(0) + t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1t + a_{12}\mathbf{f}_2t^2 + a_{13}\mathbf{f}_3t^3))].$$

Аппроксимация

$$R_{23}(t\mathbf{A}) = (60\mathbf{E} - 36t\mathbf{A} + 9(t\mathbf{A})^2 - (t\mathbf{A})^3)^{-1}(60\mathbf{E} + 24t\mathbf{A} + 3(t\mathbf{A})^2) =$$

$$= y_1(t\mathbf{A} - z_1\mathbf{E})^{-1} + 2\text{Re}[y_2(t\mathbf{A} - z_2\mathbf{E})^{-1}]$$

(коэффициенты для краткости не приводим):

$$\mathbf{x}(t) = (t\mathbf{H} - z_1\mathbf{G}_0)^{-1}(y_1(\mathbf{G}_0\mathbf{x}(0) + t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1t + a_{12}\mathbf{f}_2t^2 + a_{13}\mathbf{f}_3t^3)) +$$

$$+ 2\text{Re}[(t\mathbf{H} - z_2\mathbf{G}_0)^{-1}(y_2(\mathbf{G}_0\mathbf{x}(0) + t(a_{20}\mathbf{f}_0 + a_{21}\mathbf{f}_1t + a_{22}\mathbf{f}_2t^2 + a_{23}\mathbf{f}_3t^3))]).$$

В общем случае формулы для расчета состояния системы в момент времени t можно записать так: степень числителя аппроксимации Паде k , степень знаменателя $j > 1$, степень полинома для представления функции $\mathbf{f}(t)$ равна M , коэффициенты для всех аппроксимаций разные, определяются по описанной в статье [1] методике. Для нечетных степеней знаменателя непарный действительный корень знаменателя обозначен z_1 .

Диагональные аппроксимации порядка $2j$, степень знаменателя $j = k$ четная:

$$\mathbf{x}(t) = \mathbf{x}(0) + 2\text{Re}\sum_{n=1}^j \left[(t\mathbf{H} - z_{2n-1}\mathbf{G}_0)^{-1} \left(y_{2n-1}\mathbf{G}_0\mathbf{x}(0) + t\sum_{m=0}^M a_{2n-1,m}\mathbf{f}_m t^m \right) \right].$$

Диагональные аппроксимации порядка $2j$, степень знаменателя $j = k$ нечетная:

$$\mathbf{x}(t) = -\mathbf{x}(0) + (t\mathbf{H} - z_1\mathbf{G}_0)^{-1} \left(y_1\mathbf{G}_0\mathbf{x}(0) + t\sum_{m=0}^M a_{1,m}\mathbf{f}_m t^m \right) +$$

$$+ 2\text{Re}\sum_{n=1}^{j-1} \left[(t\mathbf{H} - z_{2n}\mathbf{G}_0)^{-1} \left(y_{2n}\mathbf{G}_0\mathbf{x}(0) + t\sum_{m=0}^M a_{2n,m}\mathbf{f}_m t^m \right) \right].$$

Субдиагональные аппроксимации вида $j = k + 1$, порядок $2j - 1$, степень знаменателя j четная:

$$\mathbf{x}(t) = 2\text{Re}\sum_{n=1}^j \left[(t\mathbf{H} - z_{2n-1}\mathbf{G}_0)^{-1} \left(y_{2n-1}\mathbf{G}_0\mathbf{x}(0) + t\sum_{m=0}^M a_{2n-1,m}\mathbf{f}_m t^m \right) \right].$$

Субдиагональные аппроксимации вида $j = k + 1$, порядок $2j - 1$, степень знаменателя j нечетная:

$$\mathbf{x}(t) = (t\mathbf{H} - z_1\mathbf{G}_0)^{-1} \left(y_1\mathbf{G}_0\mathbf{x}(0) + t\sum_{m=0}^M a_{1,m}\mathbf{f}_m t^m \right) +$$

$$+ 2\text{Re}\sum_{n=1}^{j-1} \left[(t\mathbf{H} - z_{2n}\mathbf{G}_0)^{-1} \left(y_{2n}\mathbf{G}_0\mathbf{x}(0) + t\sum_{m=0}^M a_{2n,m}\mathbf{f}_m t^m \right) \right].$$

7. Решение ДАСУ для задач, связанных с графами

Существует по крайней мере один класс задач, который позволяет проделать дополнительные несложные преобразования с полученными формулами для решения ДАСУ и существенно сократить вычисления. Это задачи расчета токов (потоков) в ребрах некоторого графа и потенциалов его вершин – расчет переходных процессов в электрических цепях, а также процессов в гидравлических, тепловых, механических и других системах. К такому виду можно привести и некоторые полевые задачи. Матрицу $t\mathbf{H} - z_k\mathbf{G}_0$ для за-

дач, связанных с графами, масштабированием строк можно привести к симметричному виду:

$$\mathbf{W}_k = \begin{bmatrix} -\mathbf{R}_k & \mathbf{K}^T \\ \mathbf{K} & \mathbf{0} \end{bmatrix},$$

где $\mathbf{R}_k$ – диагональная матрица сопротивлений ребер графа, $\mathbf{K}$ – матрица инцидентности графа, k – номер корня знаменателя аппроксимации Паде. Масштабирование строк производится умножением на симметризирующую диагональную матрицу $\mathbf{S}_k$: $\mathbf{W}_k = \mathbf{S}_k(t\mathbf{H} - z_k\mathbf{G}_0)$.

Продолжим решение на примере аппроксимации Паде R_{12} . Вставим в формулу (10) матрицу $\mathbf{S}_1$:

$$\mathbf{x}(t) = 2\text{Re}[(\mathbf{S}_1(t\mathbf{H} - z_1\mathbf{G}_0))^{-1}\mathbf{S}_1(y_1\mathbf{G}_0\mathbf{x}(0) + t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1t + a_{12}\mathbf{f}_2t^2 + a_{13}\mathbf{f}_3t^3))].$$

Введем обозначения:

$$\mathbf{q} = \mathbf{S}_1(y_1\mathbf{G}_0\mathbf{x}(0) + t(a_{10}\mathbf{f}_0 + a_{11}\mathbf{f}_1t + a_{12}\mathbf{f}_2t^2 + a_{13}\mathbf{f}_3t^3)),$$

$$\mathbf{q} = \begin{bmatrix} \mathbf{e}_+ \\ \mathbf{j}_+ \end{bmatrix}, \quad \mathbf{x}_+ = \mathbf{W}_1^{-1}\mathbf{q}, \quad \mathbf{x}_+ = \begin{bmatrix} \mathbf{i}_+ \\ \boldsymbol{\varphi}_+ \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{i} \\ \boldsymbol{\varphi} \end{bmatrix}, \quad (11)$$

где $\mathbf{e}_+$, $\mathbf{j}_+$, $\mathbf{i}_+$, $\boldsymbol{\varphi}_+$ – некоторые вспомогательные векторы; $\mathbf{i}$ – вектор токов; $\boldsymbol{\varphi}$ – вектор потенциалов. Это позволит записать формулу (10) в виде:

$$\mathbf{x}(t) = 2\text{Re}[\mathbf{x}_+]. \quad (12)$$

Можно значительно сократить вычисления, если воспользоваться методом дополнения по Шуру. Симметрия матрицы $\mathbf{W}_1$ облегчает эту задачу. С учетом введенных обозначений запишем систему уравнений для вычисления вектора $\mathbf{x}_+$ в виде:

$$\begin{bmatrix} -\mathbf{R}_1 & \mathbf{K}^T \\ \mathbf{K} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{i}_+ \\ \boldsymbol{\varphi}_+ \end{bmatrix} = \begin{bmatrix} \mathbf{e}_+ \\ \mathbf{j}_+ \end{bmatrix}.$$

Далее выполним преобразования, фактически используем блочный метод Гаусса:

$$\begin{bmatrix} -\mathbf{E} & \mathbf{R}_1^{-1}\mathbf{K}^T \\ \mathbf{K} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{i}_+ \\ \boldsymbol{\varphi}_+ \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1^{-1}\mathbf{e}_+ \\ \mathbf{j}_+ \end{bmatrix},$$

$$\begin{bmatrix} -\mathbf{E} & \mathbf{R}_1^{-1}\mathbf{K}^T \\ \mathbf{0} & \mathbf{K}\mathbf{R}_1^{-1}\mathbf{K}^T \end{bmatrix} \begin{bmatrix} \mathbf{i}_+ \\ \boldsymbol{\varphi}_+ \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1^{-1}\mathbf{e}_+ \\ \mathbf{j}_+ + \mathbf{K}\mathbf{R}_1^{-1}\mathbf{e}_+ \end{bmatrix}.$$

Диагональная матрица $\mathbf{R}_1$ обращается элементарно. Нижняя матричная строка представляет собой отдельную систему уравнений для вектора $\boldsymbol{\varphi}_+$, в электротехнике это известно как метод узловых потенциалов:

$$\mathbf{K}\mathbf{R}_1^{-1}\mathbf{K}^T\boldsymbol{\varphi}_+ = \mathbf{j}_+ + \mathbf{K}\mathbf{R}_1^{-1}\mathbf{e}_+. \quad (13)$$

Векторы $\mathbf{e}_+$ и $\mathbf{j}_+$ находятся как составляющие вектора $\mathbf{q}$ согласно обозначениям (11). Решив систему (13), находим вектор $\boldsymbol{\varphi}_+$, далее вычисляем вектор $\mathbf{i}_+$ по формуле: $\mathbf{i}_+ = \mathbf{R}_1^{-1}(\mathbf{K}^T\boldsymbol{\varphi}_+ - \mathbf{e}_+)$.

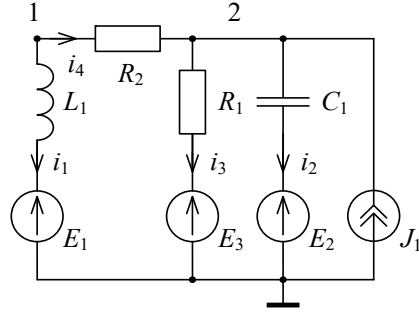


Рис. 1. Схема электрической цепи

Зная векторы $\mathbf{i}_+$ и $\boldsymbol{\varphi}_+$, находим вектор состояния системы $\mathbf{x}(t)$ по формуле (12). Можно рассчитывать не все составляющие векторов $\mathbf{i}_+$ и $\mathbf{i}$, а только те, которые нужны в конкретной задаче. Обязательны к вычислению только составляющие вектора $\mathbf{i}$, соответствующие ненулевым столбцам матрицы $\mathbf{G}_0$.

8. Пример решения ДАСУ для электрической цепи

Рассмотрим электрическую цепь, показанную на рис 1. Запишем исходную ДАСУ цепи:

$$\mathbf{G}_0\mathbf{x}' = \mathbf{H}\mathbf{x} + \mathbf{f}.$$

Соответствующие матрицы имеют вид

$$\mathbf{G}_0 = \begin{bmatrix} L_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & C_1 \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{bmatrix},$$

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -R_1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -R_2 & 1 & -1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 0 & 0 \end{bmatrix},$$

$$\mathbf{x} = \begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ i_4 \\ \varphi_1 \\ \varphi_2 \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} -E_1 \\ C_1 E'_2 \\ -E_3 \\ 0 \\ 0 \\ -J_1 \end{bmatrix},$$

$$\mathbf{W}_1 = \begin{bmatrix} \frac{-z_1 L_1}{\Delta t} & 0 & 0 & 0 & 1 & 0 \\ 0 & \frac{-\Delta t}{z_1 C_1} & 0 & 0 & 0 & 1 \\ 0 & 0 & -R_1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -R_2 & 1 & -1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 0 & 0 \end{bmatrix},$$

$$\mathbf{S}_1 = \frac{1}{\Delta t} \begin{bmatrix} 1 & & & & & \\ & \frac{-\Delta t}{z_1 C_1} & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & 0 & & & 1 & \\ & & & & & 1 \end{bmatrix},$$

$$\mathbf{K} \mathbf{R}_1^{-1} \mathbf{K}^T = \begin{bmatrix} \frac{\Delta t}{z_1 L_1} + \frac{1}{R_2} & \frac{-1}{R_2} \\ \frac{-1}{R_2} & \frac{z_1 C_1}{\Delta t} + \frac{1}{R_1} + \frac{1}{R_2} \end{bmatrix}.$$

Матрицу $\mathbf{K} \mathbf{R}_1^{-1} \mathbf{K}^T$ можно составить прямо по схеме цепи согласно известным в электротехнике правилам для метода узловых потенциалов.

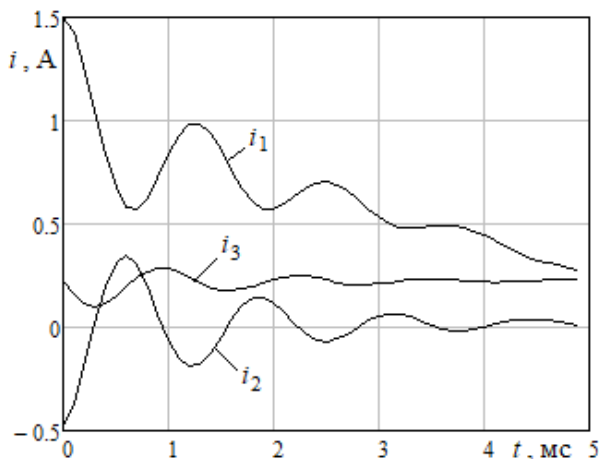


Рис. 2. Токи электрической цепи

На рис. 2 представлены графики токов электрической цепи, показанной на рис. 1. Расчет произведен с помощью аппроксимации Паде матричной экспоненты R_{12} по формуле (10).

Параметры цепи:

$$R_1 = 180 \text{ Ом}, R_2 = 0,5 \text{ Ом},$$

$$L = 0,01 \text{ Гн}, C = 4 \text{ мкФ},$$

$$E_1(t) = 10 + 2 \cdot 10^3 t - 3 \cdot 10^5 t^2 + 5 \cdot 10^7 t^3 \text{ В},$$

$$E_2(t) = 20 - 3 \cdot 10^3 t + 4 \cdot 10^5 t^2 - 6 \cdot 10^7 t^3 \text{ В},$$

$$E_3(t) = -30 + 2 \cdot 10^3 t - 5 \cdot 10^5 t^2 + 7 \cdot 10^7 t^3 \text{ В}.$$

Ток источника тока $J_1(t)$, а также составляющие остальных токов и потенциалов, обусловленные действием источников ЭДС и источников тока без учета влияния начальных значений, для аналитического решения вычислены исходя из заданной составляющей тока катушки, обусловленной только действием источников: $i_{1F}(t) = 1 - 200t + 3 \cdot 10^4 t^2 - 4 \cdot 10^6 t^3 \text{ А}$. Этот же ток источника использовался как заданный для численного решения. Функции источников заданы на всем расчетном интервале, глобально. По ним рассчитаны локальные коэффициенты полиномов для каждого шага интегрирования. Начальные условия: $\mathbf{x}(0) = (1,5; -0,486; 0,222; -1,5; 9,25; 10)^T$.

Расчетный интервал – 5 мс, число точек интегрирования – 50, шаг интегрирования – 100 мкс. Постоянная времени затухания переходного процесса – 1,390 мс, период собственных колебаний процесса – 1,268 мс. Число точек интегрирования на постоянную времени – 13,9; число точек интегрирования на период собственных колебаний – 12,7. Относительная среднеквадратическая погрешность максимальная у тока i_2 – 0,76 %; минимальная у тока i_3 – 0,14 %. Погрешности при решении этой же задачи методом трапеций: максимальная у тока i_2 – 9,5 %; минимальная у тока i_3 – 1,7 %. Сравнение проводилось с аналитическим решением.

Заключение

Предложены новые экономичные по вычислениям, простые по теории и практической реализации высокоточные методы решения линейных дифференциально-алгебраических систем уравнений. На этапе решения задачи не требуется преобразование этих систем в системы обыкновенных дифференциальных уравнений. Методы позволяют решать однородные системы, а также системы, у которых свободный член в пределах шага интегрирования представлен в форме полинома, степень которого не выше чем порядок точности соответствующего метода. Так как полином является удобной формой аппроксимации для весьма разнообразных функций, это придает методам универсальность. Методы пригодны для решения задач без какой-либо настройки, не имеют ограничений на шаг интегрирования. По точности и устойчивости новые методы эквивалентны известным методам решения систем обыкновенных дифференциальных уравнений соответствующих порядков,

однако требуют в несколько раз меньше вычислительной работы [1]. Методы одношаговые, это позволяет легко изменять шаг интегрирования.

Предложенные методы позволяют решать задачи на каждом шаге интегрирования с малыми вычислительными затратами без подготовительных процедур, что дает возможность использовать их как основу для решения систем уравнений с переменными коэффициентами и нелинейных систем. Методы проверены на практике большим количеством численных экспериментов. Они не требуют глубоких математических знаний и доступны инженерам с обычным техническим образованием.

Подробные исследования свойств и условий применения предложенных в настоящей работе методов еще не проводились и являются актуальными. В частности, пока неясно влияние на решение индекса ДАСУ.

Литература

1. Бурцев Ю.А. Решение задачи Коши высокоточными методами на основе аппроксимации Паде матричной экспоненты // Труды ИСА РАН. 2024. № 1. С. 3-11.
2. Хайрер Э., Нёрсетт С., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Жесткие и дифференциально-алгебраические задачи. / Пер. с англ., М.: Мир, 1999. 685 с. (E. Hairer, G. Wanner. Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems. Second Revised Edition. Springer Verlag Berlin Heidelberg 1991, 1996.)
3. Гридин В.Н., Михайлов В.Б., Шустерман Л.Б. Численно-аналитическое моделирование радиоэлектронных схем. М.: Наука. 2008. 339 с.
4. Бурцев Ю.А. Сравнение программы расчета электрических цепей на основе модифицированного табличного метода с известными аналогами // Известия высших учебных заведений. Электромеханика. 2013. № 4. С. 8-13.
5. Бояринцев Ю.Е., Корсуков В.М. Применение разностных методов к решению регулярных систем обыкновенных дифференциальных уравнений // Вопросы прикладной математики. – Иркутск : Изд. СЭИ СО АН СССР. 1975.
6. Бояринцев Ю.Е. Регулярные и сингулярные системы линейных обыкновенных дифференциальных уравнений. Новосибирск : Наука, Сиб. отделение, 1980. 222 с.
7. Бояринцев Ю.Е. Методы решения вырожденных систем обыкновенных дифференциальных уравнений. Новосибирск: Наука. 1988. 158 с.
8. Бояринцев Ю.Е., Чистяков В.Ф. Алгебро-дифференциальные системы: Методы решения и исследования. Новосибирск: Наука. 1998. 224 с.
9. Бояринцев Ю.Е. Линейные и нелинейные алгебро-дифференциальные системы. Новосибирск: Наука. 2000. 222 с.
10. Бояринцев Ю.Е., Орлова И.В. Пучки матриц и алгебро-дифференциальные системы. Новосибирск: Наука. 2006. 124 с.
11. Petzold L. Differential/Algebraic Equations are not ODE'S. SIAM Journal of Scientific and Statistical Computing. Vol. 3. No. 3. September 1982.
12. Курина Г.А. Новый алгоритм построения асимптотического решения сингулярно возмущенных задач оптимального управления с пересекающимися траекториями вырожденного уравнения состояния. // Прикладная математика & Физика. 2023. Т. 55. № 4. С. 313–329.
13. Курина Г.А. Проекторный подход к алгоритму Бутузова–Нефедова асимптотического решения одного класса сингулярно возмущенных задач в критическом случае // Журнал вычислительной математики и математической физики. 2020. Т. 60. № 12. С. 2073–2084
14. Скворцов Л.М. Численное решение обыкновенных дифференциальных и дифференциально-алгебраических уравнений. М.: ДМК Пресс. 2018. 230 с.
15. Burtsev Y. High Precision Methods Based on Pade Approximation of Matrix Exponent for Numerical Analysis of Stiff-Oscillatory Electrical Circuits. Proceedings - 2020 International Conference on Industrial Engineering, Applications and Manufacturing, ICIEAM 2020; IEEE inc.
16. Бейкер Дж. мл., Грейвс-Моррис П. Аппроксимации Паде. / Пер. с англ., Москва: Мир, 1986. 502 с. (G. A. Baker, Jr., P. Graves-Morris. Pade approximants. London, Addison-Wesley Publishing Company. 1981. 502 p.)
17. Гантмахер Ф.Р. Теория матриц. Издание второе, дополненное. М.: Наука. 1966. 576 с.

Бурцев Юрий Алексеевич. Федеральное государственное бюджетное образовательное учреждение высшего образования «Южно-Российский государственный политехнический университет им. М.И. Платова», г. Новочеркасск, Россия. Доцент. Кандидат технических наук. Область научных интересов: численные методы решения систем обыкновенных дифференциальных уравнений и дифференциально-алгебраических систем уравнений, теория электрических цепей. E-mail: proton36@yandex.ru

Solving Differential-Algebraic Equation Systems with Pade Approximation of Matrix Exponent

Y. A. Burtsev

Platov South-Russian State Polytechnic University, Russia

Abstract. A set of new numerical methods for solving linear differential-algebraic equation systems is developed. Homogenous systems can be solved, and nonhomogeneous systems with piecewise-polynomial right-hand side function. Calculation of system state at each integration step requires solving one or several (depending on the method order) systems of linear algebraic equations. The methods are based on decomposition of Pade approximation of the matrix exponent to simplest fractions. The proposed formulas make possible to avoid conversion of the differential-algebraic equation system to the ordinary differential equation system at the stage of system solving. The new methods are equivalent to some well-known Runge-Kutta type methods like Radau and Lobatto methods in terms of accuracy and steady areas. However, new methods are much more simple in theory and practical implementation, and they require several times less computational work. Methods with diagonal Pade approximations are A -stable, and methods with subdiagonal Pade approximations are L -stable. New methods can be used for solving stiff, oscillative and stiff-oscillative systems.

Keywords: numerical methods, differential-algebraic equation systems, matrix exponent, Pade approximation.

DOI: 10.14357/20790279240304 **EDN:** HKTOYE

References

1. Burtsev Y.A. Solving of Cauchy Problem with High Precision Methods Based on Matrix Exponent // Trudy Instituta sistemnogo analiza Rossiyskoy akademii nauk (ISA RAN) (Proceedings of the Institute for Systems Analysis Russian Academy of Sciences (ISA RAS)). 2024; 1:3-11 (In Russ.).
2. Hairer E., Wanner G. Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems. Second Revised Edition. Springer Verlag Berlin Heidelberg; 1991, 1996. 685 p.
3. Gridin V.N., Mikhajlov V.B., Shusterman L.B. Numerical-analytical modelling of radioelectronic schemes. Moscow: Science Publishing; 2008. 339 p. (In Russ.).
4. Burtsev Y.A. Comparing of Program for Calculation of Electrical Circuits Based on Modified Table Method with Known Analogs. Izvestiya vysshih uchebnyh zavedenij. Elektromekhanika (Russian Electromechanics). 2013; 4:8-13 (In Russ.).
5. Boyarintsev Y.E., Korsukov V.M. Applying of Finite-difference Methods to Solving Regular Systems of Ordinary Differential Equations // Voprosy prikladnoj matematiki (Questions of Applied Mathematic. Irkutsk: Publishing of Siberian Energy Institute, Siberian Section of Academy of Sciences of the USSR), 1975. (In Russ.).
6. Boyarintsev Y.E. Regular and Singular Systems of Linear Ordinary Differential Equations. Novosibirsk: Science Publishing, Siberian Section; 1980. 222 p. (In Russ.).
7. Boyarintsev Y.E. Methods for Solving of Degenerate Systems of Ordinary Differential Equations. Novosibirsk: Science Publishing, 1988. 158 p. (In Russ.).
8. Boyarintsev Y.E. Algebro-Differential Systems: Methods of Solving and Research. Novosibirsk: Science Publishing, 1998. 224 p. (In Russ.).
9. Boyarintsev Y.E. Linear and Non-linear Algebro-Differential Systems. Novosibirsk: Science Publishing, 2000. 222 p. (In Russ.).
10. Boyarintsev Y.E., Orlova I.V. Sheaves of Matrixes and Algebro-Differential Systems. Novosibirsk: Science Publishing, 2006. 124 p. (In Russ.).
11. Petzold L. Differential/Algebraic Equations are not ODE'S. SIAM Journal of Scientific and Statistical Computing. Vol. 3, No. 3, September 1982.
12. Kurina G.A. New Algorithm for Constructing Asymptotic Solution of Singular Indignant Problems of Optimal Control with Intersecting Trajectories of degenerate State Equation. // Prikladnaya Matematika & Fizika (Applied Mathematics and Physics, 2023, vol. 55, # 4, p. 313-329.) (In Russ.).
13. Kurina G.A. Projector Approach to Butuzov-Nefedov Algorithm for Asymptotic Solution of One Class Singular Indignant Tasks in Critical Case // Zhurnal vychislitelnoj matematiki i matematicheskoy fiziki (Journal of Computational Mathematics and Mathematical Physics, 2020, vol. 60, # 12, p. 2073-2084.) (In Russ.).
14. Skvortsov L.M. Numerical solving of ordinary differential and differential-algebraic equations. Moscow: DMK Press Publishing; 2018. 230 p. (In Russ.).
15. Burtsev Y.A. High Precision Methods Based on Pade Approximation of Matrix Exponent for Nu-

- merical Analysis of Stiff-Oscillatory Electrical Circuits. Proceedings - 2020 International Conference on Industrial Engineering, Applications and Manufacturing, ICIEAM 2020; IEEE inc.
16. *Baker G.A. Jr., Graves-Morris P.* Pade approximants. London: Addison-Wesley Publishing Company; 1981. 502 p.
17. *Gantmaher F.R.* Theory of Matrixes. Second Edition, Expanded. Moscow: Science; 1966. 576 p. (In Russ)

Burtsev Yuri Alexeevich. Philosophy doctor, senior lecturer. Platov South-Russian State Polytechnic University (NPI). Scientific interests: Numerical methods for solving systems of ordinary differential equations and systems of differential-algebraic equations, theory of electrical circuits. Email: proton36@yandex.ru

Компьютерный анализ текстов

Контекстно-независимый метод быстрой детекции текста для распознавания номеров телефонов

А.В. ГАЙЕР^{I,II}

^I Федеральный исследовательский центр «Информатика и управление»
Российской академии наук», г. Москва, Россия;

^{II} ООО «Смарт Энджинс Сервис», г. Москва, Россия

Аннотация. Современные методы детекции текста на изображениях основаны на вычислительно затратных моделях глубокого обучения и требуют большое количество данных для обучения, в том числе реальных. В случае поиска текста в произвольных сценах, процесс сбора и аннотирования настоящих данных для обучения крайне трудозатратен и дорог из-за высокой вариативности возможных сцен. В данной работе представлен новый метод детекции текста на произвольных изображениях, который не требует для обучения фотографий текста в реальных сценах и может быть обучен на простых синтетических данных в виде строк. Предложенная нейросетевая модель в 42 раза меньше, чем детектор текста в одной из лучших в плане качества и скорости работы системе распознавания текста PaddleOCR (84 КБ против 3.6 МБ), что делает ее отличным выбором для мобильных устройств. Модель была протестирована в составе системы распознавания номеров телефонов, где с ее помощью удалось достичь 80,35% правильно распознанных номеров.

Ключевые слова: глубокое обучение, детекция объектов, сегментация изображений, детекция текста.

DOI: 10.14357/20790279240305 **EDN:** HREWAU

Введение

Поиск текста на изображении представляет собой классическую задачу компьютерного зрения. Подходы, решающие данную проблему, сегодня активно применяются в задачах распознавания документов [1], перевода текста в режиме дополненной реальности, а также в задачах распознавания кодифицированных строк, т.е. имеющих фиксированный формат записи: номер телефона или банковской карты, платежные реквизиты, серийные номера деталей на производстве и т.д. Общей целью при постановке большинства задач, связанных с поиском текста, является упрощение ввода данных, уменьшение количества ошибок ввода и повышение удобства использования прикладных систем.

Современные методы поиска текста на изображении используют подходы глубокого обучения, для которых необходимы большие массивы обучающих данных. И если для задачи распознавания документов можно успешно использовать синтез данных [2], то для задачи поиска текста в произвольной сцене это сделать гораздо сложнее. Так, для задачи детекции номера телефона нужно учесть возможные поверхности, на которых может быть текст, а также сами сцены съемки, т.е. контекст. Т. о., есть необходимость разработки таких моделей поиска текста в произвольных сценах, которые можно обучить на простых синтезированных данных.

Уже более 10 лет для решения задач компьютерного зрения активно используются смартфоны

и планшеты, а также иные мобильные устройства со встроенной камерой. Мощность современных мобильных процессоров позволяет запускать глубокие нейронные сети прямо на устройстве, что важно для безопасной работы с персональными данными в виде документов и платежных реквизитов. Однако вычислительная сложность алгоритмов поиска текста на изображении, а также размер используемых для этого моделей все еще играют важную роль в виду двух факторов: а) большинство устройств относятся к бюджетному классу, мощность которых ограничена; б) энергоэффективные модели экономят расход заряда аккумулятора, который также ограничен.

В данной работе представлен новый нейросетевой подход поиска текстовых строк на изображении, который не завязан на контексте сцены при обучении и может быть эффективно применен на мобильных устройствах: модель детектора содержит всего 21 тысячу параметров и весит 84 КБ против 3,6 МБ компактного детектора текста в одной из лучших систем распознавания PaddleOCR. Обученный детектор был протестирован в задаче распознавания номера телефона на реальных данных, где он показал свою эффективность – количество правильно найденных и при этом распознанных номеров равняется 80,35%. Примеры целевых изображений задачи приведены на рис. 1. Входное изображение содержит номер телефона, который необходимо локализовать и распознать. Предложенный подход детекции текста работает с условием, что искомая строка имеет формат, позволяющий отличить ее от других строк на изображении. Тем самым, после этапа распознавания найденных строк минимизируется негативный эффект в виде ложных срабатываний детектора текста. В то же время, это позволяет значительно упростить модель по количеству обучаемых параметров.

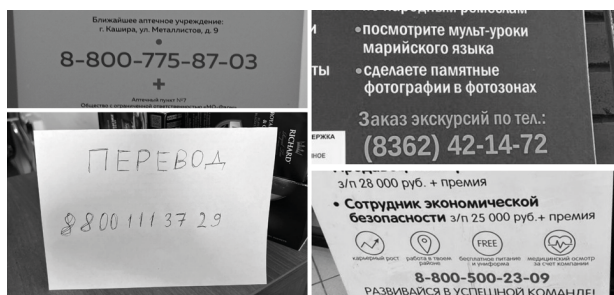


Рис. 1. Примеры изображений целевой задачи

1. Методы детекции текста на изображении

Ранние работы по поиску текста на изображении в основном рассматривали задачу распознавания сканов документов и предлагали подходы,

основанные на методах обработки изображений. Предложенные в работах [3–5] алгоритмы поиска текста используют фильтрацию изображения, морфологические операции, бинаризацию и поиск связанных компонент для выделения текстовых блоков. В работе [6] для поиска длинных текстовых строк на изображении документа используется преобразование Хафа и анализ проекций изображения для оценки наклона документа и последующего извлечения строк. Такие подходы показывают хорошие результаты для сканов документов в хорошем качестве с учетом подбора параметров алгоритмов под размер шрифта, однако в сложных сценах с неоднородным фоном они малоприменимы.

С развитием методов глубокого обучения, нейросетевые модели значительно повысили эффективность и применимость поиска текста в сценах, отличных от распознавания сканов документов. Лучшие на сегодня модели используют в качестве извлекателя признаков из изображения сверточные архитектуры нейросетей с пирамидой масштабов [7]. Выходом сети могут быть карта плотности вероятности текста [8–11] или непосредственно коллекция объектов (текстовых строк или отдельных слов). Для последнего сегодня, как правило, используются архитектуры типа трансформер с механизмом внимания [12–14]. Приведенные модели способны находить текст в произвольных сценах, однако требуют значительных вычислительных ресурсов и объема памяти. Так, например, CRAFT [8] весит 79 МБ, а более легкий детектор DBNet [10] – 53 МБ. В то же время, модели с использованием модуля трансформер имеют куда больший размер: 522,8 МБ у DPText-DETR [14]. Время работы моделей поиска текста в публикациях зачастую замеряют на высокопроизводительных видеокάρтах настольных ПК. Доля работ, которые фокусируются на вычислительно эффективных методах поиска текста на центральных процессорах, относительно невелика. В качестве примера такой работы можно отнести PaddleOCR [15]: предложенный ими детектор текста является оптимизированной версией DBNet [10] и имеет размер в 3,6 МБ.

В работе [16] был предложен подход для поиска текстовых регионов на изображениях в социальных медиа для модерации. В его основе лежит простой LeNet-подобный классификатор квадратных областей изображения на текст и фон. Будучи примененным к картинке целиком, на выходе получается карта плотности вероятности текстовых регионов. Особенностью алгоритма является то, что для его обучения не требуются настоящие картинки в полном размере: классификатор обучается

на нарезанных частях изображения, тем самым в модель не закладывается контекст сцены и повышается робастность к компоновке текста и фона. Данный метод «склеивает» строки в абзацы, поэтому для распознавания строк необходим дополнительный алгоритм, который сначала их извлечет из найденных текстовых блоков.

2. Постановка задачи

Из обзора литературы видно, что задача быстрой и вычислительно эффективной детекции текстовых строк на изображении является актуальной. Применение моделей глубокого обучения на мобильных устройствах накладывает ограничения на их вычислительную сложность, энергоэффективность и размер. Также важным является тот факт, что большинство моделей детекторов дообучаются на настоящих данных. Т.к. их сбор и разметка для обучения весьма трудоемкая и дорогостоящая процедура, то есть потребность в методах, для обучения которых можно использовать только синтетические данные.

Постановка задачи заключается в следующем. Имеется входное изображение, содержащее номер телефона. Имеется также алгоритм распознавания текстовых строк и алгоритм сверки распознанной строки на соответствие формату номера телефона. Найденные некоторой моделью поиска текстовые строки распознаются и проверяются на соответствие формату, среди которых затем выбираются наиболее похожие на номер телефона строки. Блок-схема алгоритма распознавания номеров телефонов приведена на рис. 2. В данной работе предлагается подход для быстрой детекции текстовых строк (выделенный блок) при условии, что ложные срабатывания детектора текста будут отфильтрованы за счет распознавания и проверки соответствия строки определённому формату.

Необходимо разработать метод поиска текстовых строк на изображении, который удовлетворяет следующим критериям:

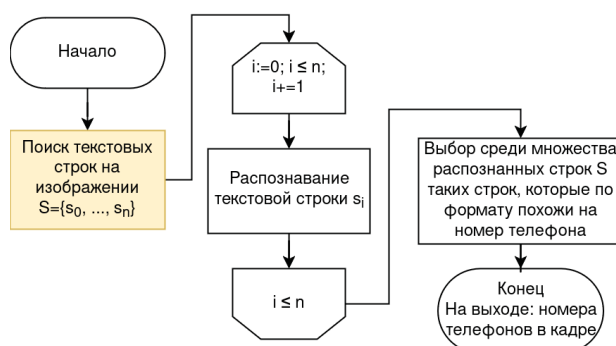


Рис. 2. Блок-схема алгоритма работы системы распознавания номеров телефонов на изображении

- 1) Быстрая скорость работы и компактный размер модели в сравнении с существующими детекторами текста общего вида;
- 2) Для обучения не требуются настоящие контекстно-зависимые данные, т.е. изображения сцен с разметкой в виде геометрических зон для вклейки текста или текстовое наполнение строк.

Важно отметить, что количество ложных срабатываний детектора не является оптимизируемой величиной т.к. по результату сверки распознанной строки с форматом номера телефона можно будет отфильтровать ложные срабатывания.

3. Предлагаемый метод

Для решения задачи предлагается использовать нейросетевую модель классификации зон изображения на «текст» и «не текст». Архитектура сети приведена в табл. 1. Она состоит только из сверточных слоев, что дает возможность применять ее к изображениям произвольного размера. В качестве функции активации используется $\text{symrelu}(x) = \max(-1, \min(1, x))$. Модель содержит всего 21 тысячу обучаемых параметров и крайне компактна: в вещественном

Табл. 1

Архитектура предлагаемой модели детектора текста.

#	Тип слоя	Параметры	Функция активации
1	Conv	8 фильтров 4x4, шаг 1x1, без отступов	Symrelu
2	Conv	8 фильтров 1x3, шаг 1x2, без отступов	Symrelu
3	Conv	8 фильтров 3x1, шаг 2x1, без отступов	Symrelu
4	Conv	8 фильтров 3x3, шаг 2x2, отступы 1x1	Symrelu
5	Conv	12 фильтров 3x3, шаг 1x1, без отступов	Symrelu
6	Conv	16 фильтров 3x3, шаг 1x1, без отступов	Symrelu
7	Conv	32 фильтров 3x2, шаг 1x1, без отступов	Symrelu
8	Conv	48 фильтров 3x1, шаг 1x1, без отступов	Symrelu
9	Conv	64 фильтров 3x1, шаг 1x1, без отступов	Symrelu
10	Conv	3 фильтра 5x1, шаг 1x1, без отступов	Softmax

типе (4 байта на значение) веса будут иметь размер всего в 84 КБ.

Обучающая выборка состоит из синтезированных текстовых строк. Синтез происходил путем печати случайных последовательностей символов различными шрифтами на изображения фонов. Преимуществом таких данных является простота их генерации и отсутствие контекста сцены. Важность независимости от контекста при обучении можно продемонстрировать следующим примером. Допустим, для обучения сети для поиска номера телефона были собраны настоящие фото с рынков, мастерских, домов быта и т.п. мест. В преобладающем большинстве случаев, номер телефона для денежного перевода пишут на листе бумаги или картона. При этом количество примеров, когда его пишут на необычной поверхности, такой как стекло или меловая доска, будет значительно меньше. Обучение на такой выборке может привести к переобучению сети под наиболее частый случай, т.е. номеров, записанных на листе бумаги. Таким образом, при обучении на настоящих данных возникает проблема несбалансированной выборки. В предложенном же методе, сеть учится на строках выявлять признаки текста без зависимости от сцены и прочего контекста.

Нейронная сеть при обучении классифицирует каждый столбец входной картинки строки с учетом его окрестности (размером с рецептивное поле сети) на 3 класса: «фон», «символ» и «частично символ». Класс «частично символ» включает в себя стыки отдельных символов или горизонтальные пространства между близкими строками – т.е. области, обладающие признаками текста, но не принадлежащие центру отдельного символа. Похожий подход применялся в детекторе текста CRAFT [8], где аналогичный класс использовался как соединяющее звено между отдельными символами в словах и помогал уменьшить количество ложных срабатываний в высокочастотных областях изображения, имеющих некоторые признаки текстовой строки в целом, но не отдельных символов.

Рассмотрим алгоритм формирования векторов-ответов для изображений текстовых строк в обучающей выборке. Вектор-ответ для изображения $W \times H \times C$ имеет размерность $W \times 1 \times 1$. Порядок его заполнения следующий. Сперва он инициализируется значениями -1. Затем, в геометрический центр по оси X каждого символа ставится индекс класса «символ», т.е. 1. Затем, между соседними символами с вероятностью p_{pol} , в случайной выбранной точке отрезка, указывается индекс 2, соответствующий классу «частично символ». Наконец, проставляются индексы точек фона (значение 0). Они соответствуют как пробелам в строке, так и пустым пространствам слева или справа от строки. Точки фона выбираются случайно, но с условием, что их суммарное количество во всей обучающей выборке должно быть равно $k_{bg} \times$ (количество отдельных символов в данных). Иллюстрация заполненного вектора-ответа для изображения приведена на рис. 3. Важным моментом алгоритма является то, что в итоге часть точек вектора-ответа будет иметь индекс -1. Во время обучения сети, функция ошибки не вычисляется в таких точках (ее значение будет равно нулю). Таким образом, необязательность присваивания меток классов всем точкам вектора-ответа делает возможным балансировку количества точек разных классов за счет параметров p_{pol} и k_{bg} . За счет балансирования кол-ва точек классов «символ» и «фон» можно добиться необходимой устойчивости сети к сложному фону. В данной работе были использованы значения $p_{pol} = 0,3$ и $k_{bg} = 2$.

Как было сказано ранее, предложенная нейронная сеть имеет полносверточную архитектуру, а потому ее можно применять к изображениям нефиксированного размера. Будучи примененной к произвольному изображению, выход сети будет представлять собой изображение, каждый пиксель которого соответствует некоторой области на входном изображении и содержит вероятность его принадлежности текстовой строке.

Результат детекции в виде прямоугольников найденных текстовых строк получается на этапе

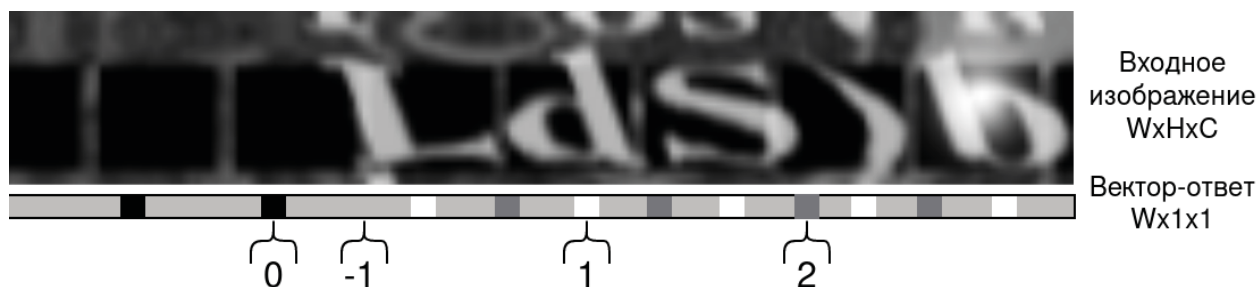


Рис. 3. Пример входного изображения текстовой строки при обучении и вектора-ответа для нее

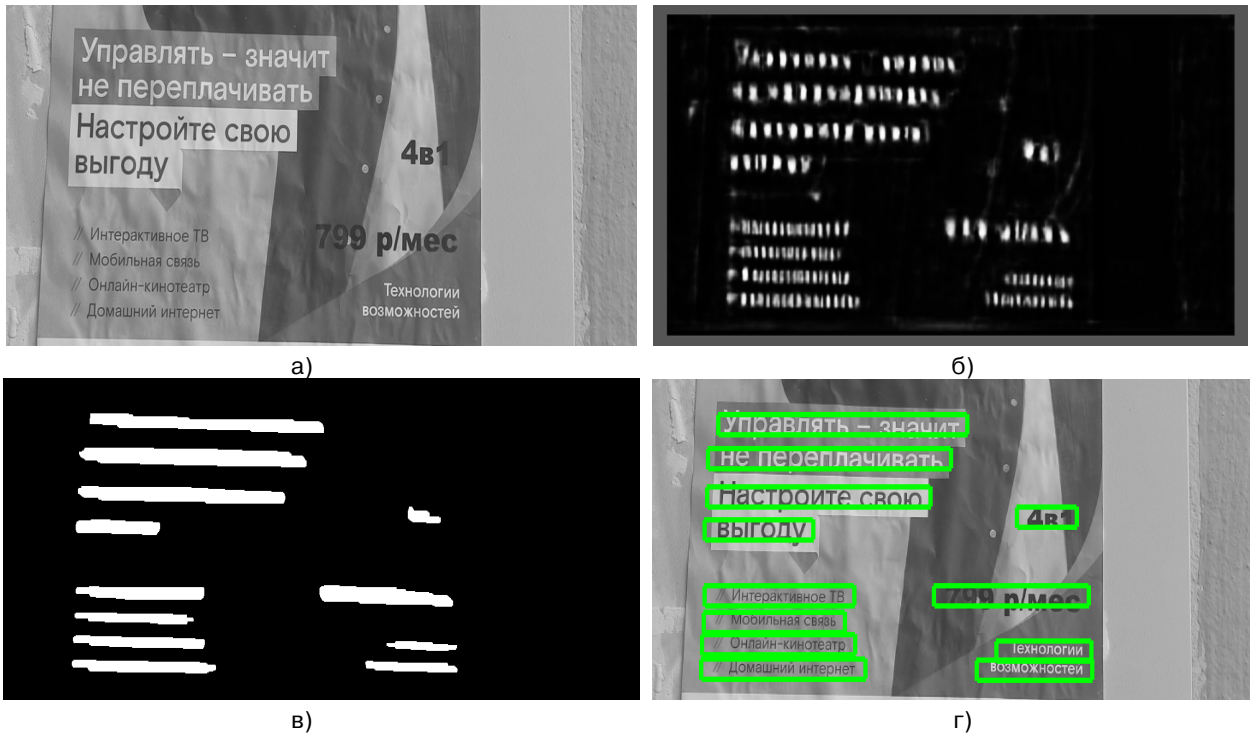


Рис. 4. Пример работы предложенного метода детекции текста: а – исходное изображение; б – выход сети в виде тепловой карты текста; в – постобработка выхода сети с помощью морфологических операций и бинаризации; г – результат детекции текста

постобработки выхода сети. Выход сети преобразуется в одноканальное изображение путем взятия значения уверенности класса «символ» в каждой точке. Затем оно приводится к размеру входа и к нему применяются морфологические операции и бинаризация по фиксированному порогу 0,5. После применяется алгоритм поиска связанных компонент, и для каждой компоненты берется ее обрамляющая рамка, которая и является результатом детекции строки. Подробно процесс детекции изображен на рис. 4.

4. Экспериментальные результаты

Обучение модели и замеры вычислительной эффективности производились на ПК с процессором AMD Ryzen 3970X и видеокартой Nvidia RTX 3090. Обучение производилось на 500 тысячах сгенерированных строках размера 256x29x1 в течение 100 эпох. Используемая функция потерь – классическая функция перекрестной энтропии, используемая в задачах классификации. В качестве алгоритма оптимизации использовался стохастический градиентный спуск (SGD) с моментами, скорость обучения была 0.01, значение моментов – 0,9, размер минибатча – 256. Для увеличения репрезентативности данных использовалась аугментация на

лету в виде накладывания шумов, бликов, размытия и небольших сдвигов изображения по вертикальной оси [17].

Предложенный метод был протестирован в задаче распознавания номера телефона в кадре. Закрытый набор данных содержал 5264 изображения номеров телефонов в разных сценах, такие как рынки, мастерские, объявления, визитки и т.д. В данных были как печатные, так и рукописные номера. Процесс распознавания был следующий: предложенная модель поиска текста выделяла все возможные текстовые строки на изображении, которые затем распознавались сторонней моделью. Среди распознанных строк в ответ выбиралась та, которая больше всего похожа на номер телефона по формату. Качество распознавания оценивалось как отношение числа правильно распознанных номеров к общему числу. Результат замера качества приведен в табл. 2.

Табл. 2

Результат распознавания номеров телефонов на изображениях с предлагаемой моделью детектора текста

Тип данных	Качество распознавания
Печатные номера	88,33%
Рукописные номера	77,02%
Любые номера	80,35%

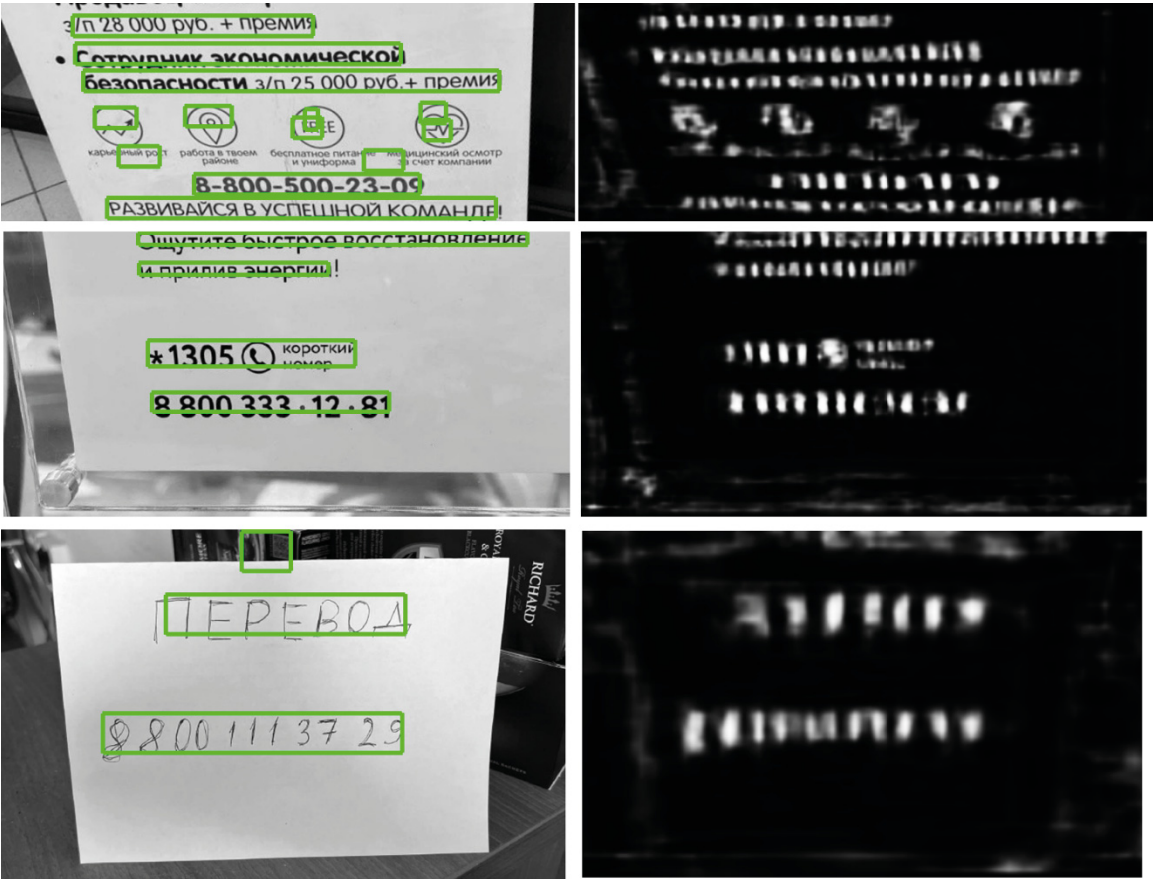


Рис. 5. Примеры работы предложенной модели поиска текста на реальных данных. Слева приведены исходные изображения с найденными рамками строк, справа – тепловая карта текста на основе выхода нейронной сети

Примеры работы предложенного детектора приведены на рис. 5.

Время работы предложенного метода на ПК с процессором AMD Ryzen 3970X на изображении 905x1280 составляет всего 199 мс. Предложенная модель была реализована в коммерческом ПО Smart Code Engine, с использованием проприетарной библиотеки обработки изображений min*. Для исполнения сверточных слоев в модели использовался алгоритм p-im2col [18], позволяющий настроить оптимальное потребление оперативной памяти. Сравнение вычислительной эффективности с другими детекторами текста представлено в табл. 3. Предложенная модель значительно превосходит

CRAFT и компактный детектор текста PaddleOCR по размеру модели и пиковому потреблению оперативной памяти (RAM). Рассматриваемые модели находятся в разных нишах как по отличающимся на порядки размерам, так и по способу обучения, из-за чего сравнение предложенного метода с PaddleOCR и CRAFT в плане качества работы не является релевантным.

Недостатком метода являются ложные срабатывания на фоне, которые фильтруются на стадии распознавания. Пример некорректной работы приведен на рис. 6. Также недостатком метода является низкая устойчивость к разномасштабному тексту – для стабильной работы с маленьким и большим

Табл. 3

Сравнение вычислительной эффективности предложенного метода детекции текста с другими детекторами

Модель	Время работы, мс	Размер модели, МБ	Пиковое потребление RAM, МБ
CRAFT [7]	1529	79,3	1100
PaddleOCR [14]	200	3,6	415,6
Предложенная модель	199	0,084	101

текстом необходимо обрабатывать вход в разных масштабах и агрегировать результаты детекции, например алгоритмом подавления не-максимумов.



Рис. 6. Пример ложных срабатываний предложенного подхода поиска текста. Лишние найденные рамки на фоне удаляются после этапа распознавания строк

Заключение

В работе представлен быстрый и компактный детектор текста для систем распознавания номеров телефонов. За счет фиксированного формата таких строк можно фильтровать ложные срабатывания детектора на сложном фоне. В сравнении с детектором текста CRAFT, предложенная модель имеет на 3 порядка меньший размер и в 11 раз меньшее потребление оперативной памяти. В сравнении с компактным детектором текста PaddleOCR, предложенная модель меньше в 42 раза (84 КБ против 3,6 МБ) и потребляет 101 МБ оперативной памяти против 415 МБ. Отличительной чертой подхода является независимость от контекста сцены и формата строк при обучении – вместо больших изображений сцен используются сгенерированные изображения строк текста. Таким образом, модель обучается только образам «текст» и «не текст», без учета дополнительной информации сцены. Эффективность метода была продемонстрирована в задаче распознавания номеров телефонов в произвольных сценах – количество правильно найденных и распознанных номеров на закрытом наборе реальных данных составило 80,35%.

В качестве будущей работы по развитию предложенного метода можно выделить повышение устойчивости детектора к разномасштабному тексту, а также уменьшение количества ложных срабатываний, что приведет к меньшему количеству запусков распознающей модели и ускорит систему распознавания номеров телефонов.

Литература

1. *Arlazarov V.L., Slavin O.A.* Issues of recognition and verification of text documents. ITiVS 3. P. 55–61. 2023. DOI: 10.14357/20718632230306.
2. *Bulatov K.B., Emelyanova E.V., Tropin D.V., Skoryukina N.S., Chernyshova Y.S., Sheshkus A.V., Usilin S.A., Ming Z., Burie J.C., Luqman M.M., Arlazarov V.V.* Midv-2020: A comprehensive benchmark dataset for identity document analysis. Computer Optics 46(2). P. 252–270 (2022). DOI: 10.18287/2412-6179-CO-1006.
3. *Okun O., Yan Y., Pietikainen M.* Robust text detection from binarized document images. In: 2002 International Conference on Pattern Recognition. Vol. 3. P. 61–64 vol.3 (2002). <https://doi.org/10.1109/ICPR.2002.1047795>.
4. *Diem M., Kleber F., Sablatnig R.* Text line detection for heterogeneous documents. In: 2013 12th International Conference on Document Analysis and Recognition. P. 743–747 (2013). <https://doi.org/10.1109/ICDAR.2013.152>.
5. *dos Santos R.P., Clemente G.S., Ren T.I., Cavalcanti G.D.* Text line segmentation based on morphology and histogram projection. In: 2009 10th International Conference on Document Analysis and Recognition. P. 651–655 (2009). <https://doi.org/10.1109/ICDAR.2009.183>.
6. *Gatos B., Papamarkos N., Chamzas C.* Skew detection and text line position determination in digitized documents. Pattern Recognition 30(9), 1505–1519 (1997). [https://doi.org/10.1016/S0031-3203\(96\)00157-4](https://doi.org/10.1016/S0031-3203(96)00157-4).
7. *Lin T.Y., Dollár P., Girshick R., He K., Hariharan B. and Belongie S.* “Feature Pyramid Networks for Object Detection,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017. P. 936–944. DOI: 10.1109/CVPR.2017.106.
8. *Baek Y., Lee B., Han D., Yun S., Lee H.* Character region awareness for text detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). P. 9357–9366 (06 2019), DOI: 10.1109/CVPR.2019.00959.
9. *Chen Z., Wang J., Wang W., Chen G., Xie E., Luo P., Lu T.* Fast: Faster arbitrarily-shaped text detector with minimalist kernel representation. In: arXiv (2021), 2111.02394.
10. *Liao M., Wan Z., Yao C., Chen K., Bai X.* Real-time scene text detection with differentiable binarization. Proceedings of the AAAI Conference on Artificial Intelligence 34(07), 11474–11481 (Apr 2020). <https://doi.org/10.1609/aaai.v34i07.6812>.
11. *Liao M., Zou, Z., Wan Z., Yao C., Bai X.* Real-time scene text detection with differentiable

- binarization and adaptive scale fusion. arXiv (2022), 2202.10304.
12. Zhang S.X., Zhu X., Yang C., Yin X.C. Arbitrary shape text detection via boundary transformer. *IEEE Transactions on Multimedia* 26. P. 1747–1760 (2022), <https://api.semanticscholar.org/CorpusID:248693243>.
 13. Bu Q., Park S., Khang M. & Cheng Y. (2024). SRFormer: Text Detection Transformer with Incorporated Segmentation and Regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(2). P. 855-863. <https://doi.org/10.1609/aaai.v38i2.27844>.
 14. Ye M., Zhang J., Zhao S., Liu J., Du B., Tao D. Dptext-detr: towards better scene text detection with dynamic points in transformer. In: *Proceedings of the AAAI Conference on Artificial Intelligence. AAAI'23/IAAI'23/EAAI'23*, AAAI Press (2023). <https://doi.org/10.1609/aaai.v37i3.25430>, <https://doi.org/10.1609/aaai.v37i3.25430>.
 15. Li C., Liu W., Guo R., Yin X., Jiang K., Du Y., Du Y., Zhu L., Lai B., Hu X., Yu D., Ma Y. Pp-ocrv3: More attempts for the improvement of ultra lightweight ocr system. ArXiv abs/2206.03001 (2022), <https://api.semanticscholar.org/CorpusID:249431435>.
 16. Layek A.K., Mandal S., Ghosh S. (2020). A Fast Approach for Text Region Detection from Images on Online Social Media. In: Das, A., Nayak, J., Naik, B., Pati, S., Pelusi, D. (eds) *Computational Intelligence in Pattern Recognition. Advances in Intelligent Systems and Computing*, vol 999. Springer, Singapore. https://doi.org/10.1007/978-981-13-9042-5_31.
 17. Gayer A.V., Sheshkus A.V. and Chernyshova Y.S. “Augmentation on the fly for the neural networks learning,” *Trudy ISA RAN (Proceedings of ISA RAS)*, vol. 68, Спецвыпуск № S1. P. 150-157, 2018, DOI: 10.14357/20790279180517.
 18. Trusov A.V., Limonova E.E., Nikolaev D.P. and Arlazarov V.V. “p-im2col: Simple Yet Efficient Convolution Algorithm with Flexibly Controlled Memory Overhead,” *IEEE Access*, vol. 9. P. 168162-168184, 2021. DOI: 10.1109/ACCESS.2021.3135690.

Гайер Александр Вячеславович. Федеральный исследовательский центр «Информатика и управление» Российской академии наук, г. Москва, Россия. Младший научный сотрудник. ООО «Смарт Энджинс Сервис», г. Москва, Россия. Научный сотрудник-программист. Область научных интересов: глубокое обучение, детекция объектов. E-mail: agayer@smartengines.com

Context-independent fast text detection method for recognizing phone numbers

A.V. Gayer^{I,II}

^I Smart Engines Service LLC, Moscow, Russia

^{II} Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia

Abstract. Modern methods for detecting text in images are based on computationally expensive deep learning models and require a large amount of training data, including real data. In the case of text retrieval in arbitrary scenarios, the process of collecting and annotating real data for training is extremely labor-intensive and expensive due to the high variability of possible scenes. This paper presents a new method for detecting text in arbitrary images, which does not require photographs of text in real scenes to be trained and can be trained on simple synthetic data in the form of strings. The proposed neural network model is 42 times smaller than the text detector in one of the best text recognition systems in terms of quality and speed, PaddleOCR (84 KB versus 3.6 MB), which makes it an excellent choice for mobile devices. The model was tested as part of a phone number recognition system, where with its help it was possible to achieve 80.35% of correctly recognized numbers.

Keywords: *deep learning, object detection, image segmentation, text detection*

DOI: 10.14357/20790279240305 **EDN:** HREWAU

References

1. *Arlazarov V.L., Slavin O.A.* Issues of recognition and verification of text documents. ITiVS 3. P. 55–61. 2023. DOI: 10.14357/20718632230306.
2. *Bulatov K.B., Emelyanova E.V., Tropin D.V., Skoryukina N.S., Chernyshova Y.S., Sheshkus A.V., Usilin S.A., Ming Z., Burie J.C., Luqman M.M., Arlazarov V.V.* Midv-2020: A comprehensive benchmark dataset for identity document analysis. Computer Optics 46(2), 252–270 (2022). DOI: 10.18287/2412-6179-CO-1006.
3. *Okun O., Yan Y., Pietikainen M.* Robust text detection from binarized document images. In: 2002 International Conference on Pattern Recognition. Vol. 3. P. 61–64 vol.3 (2002). <https://doi.org/10.1109/ICPR.2002.1047795>.
4. *Diem M., Kleber F., Sablatnig R.* Text line detection for heterogeneous documents. In: 2013 12th International Conference on Document Analysis and Recognition. P. 743–747 (2013). <https://doi.org/10.1109/ICDAR.2013.152>.
5. *dos Santos R.P., Clemente G.S., Ren T.I., Cavalcanti G.D.* Text line segmentation based on morphology and histogram projection. In: 2009 10th International Conference on Document Analysis and Recognition. P. 651–655 (2009). <https://doi.org/10.1109/ICDAR.2009.183>.
6. *Gatos B., Papamarkos N., Chamzas C.* Skew detection and text line position determination in digitized documents. Pattern Recognition 30(9), 1505–1519 (1997). [https://doi.org/10.1016/S0031-3203\(96\)00157-4](https://doi.org/10.1016/S0031-3203(96)00157-4).
7. *Lin T.Y., Dollár P., Girshick R., He K., Hariharan B. and Belongie S.* “Feature Pyramid Networks for Object Detection,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017. P. 936–944. DOI: 10.1109/CVPR.2017.106.
8. *Baek Y., Lee B., Han D., Yun S., Lee H.* Character region awareness for text detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). P. 9357–9366 (06 2019), DOI: 10.1109/CVPR.2019.00959.
9. *Chen Z., Wang J., Wang W., Chen G., Xie E., Luo P., Lu T.* Fast: Faster arbitrarily-shaped text detector with minimalist kernel representation. In: arXiv (2021), 2111.02394.
10. *Liao M., Wan Z., Yao C., Chen K., Bai X.* Real-time scene text detection with differentiable binarization. Proceedings of the AAAI Conference on Artificial Intelligence 34(07), 11474–11481 (Apr 2020). <https://doi.org/10.1609/aaai.v34i07.6812>.
11. *Liao M., Zou, Z., Wan Z., Yao C., Bai X.* Real-time scene text detection with differentiable binarization and adaptive scale fusion. arXiv (2022), 2202.10304.
12. *Zhang S.X., Zhu X., Yang C., Yin X.C.* Arbitrary shape text detection via boundary transformer. IEEE Transactions on Multimedia 26, 1747–1760 (2022), <https://api.semanticscholar.org/CorpusID:248693243>.
13. *Bu Q., Park S., Khang M. & Cheng Y.* (2024). SRFormer: Text Detection Transformer with Incorporated Segmentation and Regression. Proceedings of the AAAI Conference on Artificial Intelligence, 38(2), 855–863. <https://doi.org/10.1609/aaai.v38i2.27844>.
14. *Ye M., Zhang J., Zhao S., Liu J., Du B., Tao D.* Dptext-detr: towards better scene text detection with dynamic points in transformer. In: Proceedings of the AAAI Conference on Artificial Intelligence. AAAI’23/IAAI’23/EAAI’23, AAAI Press (2023). <https://doi.org/10.1609/aaai.v37i3.25430>, <https://doi.org/10.1609/aaai.v37i3.25430>.
15. *Li C., Liu W., Guo R., Yin X., Jiang K., Du Y., Du Y., Zhu L., Lai B., Hu X., Yu D., Ma Y.* Pp-ocrv3: More attempts for the improvement of ultra lightweight ocr system. ArXiv abs/2206.03001 (2022), <https://api.semanticscholar.org/CorpusID:249431435>.
16. *Layek A.K., Mandal S., Ghosh S.* (2020). A Fast Approach for Text Region Detection from Images on Online Social Media. In: Das, A., Nayak, J., Naik, B., Pati, S., Pelusi, D. (eds) Computational Intelligence in Pattern Recognition. Advances in Intelligent Systems and Computing, vol 999. Springer, Singapore. https://doi.org/10.1007/978-981-13-9042-5_31.
17. *Gayer A.V., Sheshkus A.V. and Chernyshova Y.S.* “Augmentation on the fly for the neural networks learning,” Trudy ISA RAN (Proceedings of ISA RAS), vol. 68, Спецвыпуск № S1. P. 150–157, 2018, DOI: 10.14357/20790279180517.
18. *Trusov A.V., Limonova E.E., Nikolaev D.P. and Arlazarov V.V.* “p-im2col: Simple Yet Efficient Convolution Algorithm with Flexibly Controlled Memory Overhead,” IEEE Access, vol. 9. P. 168162–168184, 2021. DOI: 10.1109/ACCESS.2021.3135690.

Gayer A.V. Junior researcher at the Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia. Researcher-programmer at Smart Engines Service LLC, Moscow, Russia. Number of publications: 13. Research interests: deep learning, object detection. E-mail: agayer@smartengines.com

Информатика сообществ и формирование социальных сетей

Данные как вызов

В.И. Тищенко

Федеральный исследовательский центр «Информатика и управление»
Российской академии наук», г. Москва, Россия

Аннотация. В статье рассмотрены проблемы анализа природы данных, возникающие на границе знания и бытия. Показана противоречивость представления науки о данных как совокупности технологий и алгоритмов, предназначенных для решения задач обработки больших массивов слабоструктурированных данных, прогнозирования нераспознаваемых ранее корреляций. Представлен этимологический и когнитивный анализ трансформации понятия «данные». Проанализированы концептуальные основания становления новой эпистемологической парадигмы отображения окружающего мира, позволяющей рассматривать по-новому проблему данных.

Ключевые слова: наука о данных, *data science*, эпистемология модели познания.

DOI: 10.14357/20790279240306 **EDN:** LHREBR

Введение

На современном этапе всесторонней цифровизации социального бытия и повсеместного экспоненциального роста количества данных «производство» («добыча»), как в обыденном, так и научном сознании, рассматривается как нечто очевидное, само собой разумеющееся. И, как следствие, для этого термина в сообществе специалистов сформировалось вполне рабочее и устоявшееся понимание [1]. *Данные* – это форма представления информации, фактов или понятий, зарегистрированных восприятий или измерений, пригодных для передачи, общения, интерпретации или обработки, в т. ч., с помощью цифровых устройств. Данные обычно рассматривают как мельчайшие единицы фактической информации,

которые могут быть использованы в качестве основы для расчетов, рассуждений или обсуждения. Данные могут варьироваться от абстрактных идей до конкретных измерений [2-5]. И, казалось бы, не совсем понятно, что может в настоящее время дать для развития информационных технологий когнитивный анализ, казалось бы, устоявшегося понятия «данные».

1. Многозначность взглядов на природу данных

Однако при всей кажущейся очевидности истолкования этого термина, значение данных как источника новых знаний, существующего на стыке знания и бытия и «возникающего» в моменты реги-

страции актов восприятия или измерений, не раскрывается в приведенных выше определениях и, с очевидностью, требует дополнительного рассмотрения [6]. К необходимости дополнительного осмысления природы данных, возможно, также подталкивает, то, что понятие «данные» нередко рассматривается как синоним понятия «информация», или элементов нового знания, и в то же время существует мнение, «данные больше, чем знания» [4].

Как ни странно, более глубокому рассмотрению этого понятия в немалой степени мешает проведенная в 2015 году стандартизация термина, предпринятая Международной организацией по стандартизации (ISO): «Данные – представляют поддающееся многократной интерпретации представление информации в формализованном виде, пригодном для передачи, связи, или обработки»¹ [ISO/IEC 2382:2015]².

Другими словами, это закрепило в общепринятом употреблении, представление о *данных* как наборе дискретных, или непрерывных значений, которые передают информацию, описывающую количество, качество, факты, статистику, другие базовые смысловые единицы или просто последовательности символов, которые в дальнейшем могут быть интерпретированы формально [9,10].

В академической среде стремление осмыслить природу данных как источника знаний тем не менее имеет определенную историю. В конце 80-х годов прошлого века это движение привело к построению т. н. «информационной иерархии» в форме четырехуровневой модели DIKW (данные, информация, знания, мудрость) соотношения данных и действительности [11]. Однако в дальнейшем интеллектуальный анализ данных, процессов их «создания» и сохранения, а главное аппаратное развитие информационных технологий показали, что модель DIKW предлагает сильно искаженную и упрощенную картину «извлечения» знания из данных. Попыткой уйти от схематизма модели DIKW и стало формирование направления, получившего название *наука о данных* (Data Science, англ.).

И, хотя начало исследований в области анализа данных обычно связывают с именами Д. Неймана и К. Пирсона, собственно Data Science как академическая наука начала развиваться лишь с 1996

г., даты создания CODATA (Комитета по данным для науки и техники), а также издания соответствующих научных журналов [12,13].

В основе этого направления находится интеллектуальный анализ данных и, соответственно, подготовка, отбор, очистка данных и адекватная интерпретация результатов [14]. Другими словами, Data Science исследует жизненный цикл *цифровых данных* — от появления до преобразования и представления в других областях знаний (Peter Naur).

Очевидно, что концептуальной основой Data Science является представление, что данные, и, соответственно их анализ, все же начинаются непосредственно со сбора и систематизации показаний, сгенерированных различными источниками в процессе преобразования потока независимых переменных. При этом данные отождествляются с зафиксированными наборами значений качественных или количественных переменных, которыми «обладают» и которыми «характеризуются» исследуемые процессы и события, существующие как бы независимо от наблюдателя. *Данные* – ведь, это то, что *дано*, *найденно*. И такое понимание представляется естественным, ведь слово *данные* в русском языке берет свое начало от праславянского слова **dati* – ‘*дар*’ (индоевропейское **dō*); ср. английское слово *data*, производное от латинского слова *datum* – *данная вещь* [15].

При этом основные усилия, по мнению IT-специалистов, затрачиваются при обработке данных на сбор, «очистку» и *систематизацию* полученных показателей. Ведь, пока *данные* не вписаны в тот или иной контекст, не структурированы, они не могут рассматриваться как исходный материал для аналитики, оставаясь как бы «сырыми» показаниями. Только в результате применения интеллектуальных алгоритмов из данных «извлекается» информация и получают «умные» («smart») данные [16]³. Но тогда анализ данных, по сути, рассматривается как способ нахождения адекватных алгоритмов, которые собственно и преобразуют, по мнению Data аналитиков, «сырые» зафиксированные переменные (которые ошибочно иногда определяют, как данные) в *данные*, пригодные для извлечения информации [17]. И следствием этого оказывается «естественный» вывод, что наука о данных – это разработка методов по обработке данных в условиях больших объемов.

В результате такого взгляда на проблему данных, вроде бы, исчезает необходимость обсуждать природу данных как таковых, их происхождение,

¹ «Data – a reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or communication, or processing» [7].

² В нашей стране разработан и утвержден аналог приведенного выше международного стандарта – ГОСТ 33707-2016 «Информационные технологии. Словарь»; согласно этому ГОСТу – «данные – предоставление информации в формальном виде, пригодном для передачи, интерпретации или обработки людьми, или компьютерами» [8, с. 21.].

³ Не удивительно, что здесь сразу вспоминается пирамидальная структура модели DIKW, где в основании лежит слой «данных», над которым располагается уменьшенный по сравнению с предыдущим слой «информация». И последующие слои – «знание» и «мудрость».

поскольку они представляют собой результаты «обработки» воздействий, зарегистрированных датчиками измерительных устройств. И, соответственно, наука о данных, рассматриваемая с этой точки зрения, обсуждает не столько собственно данные, сколько возможности их поиска, выделения и агрегирования.

А между тем данные, сгенерированные разнообразными способами, не существуют сами по себе, и не появляются из ниоткуда. Ведь даже зарегистрированные показания (которые определяют, как источник данных) должны быть «порождены» регистрирующими устройствами, *выбранными* для описания (регистрации) «предстоящей» реальности согласно той или иной концептуальной модели исследуемой системы, будь она неосознанная или значимо формализована. Ведь анализ – это не ответ на вопрос, это «задание вопроса», нахождение связи между наблюдаемыми процессами. Поэтому наука о данных не является статистическим или компьютерным анализом обнаружения и выделения корреляций, которые могут свидетельствовать о наличии закономерностей и взаимосвязей в исследуемых процессах. И, соответственно, не «корреляции вытесняют каузальность», а каузальность порождает корреляции.

Данные без модели, которая «настраивает» регламент, процедуру, а самое главное, средства и методы сбора данных, или, по сути, определяет регистрирующие средства, приборы и датчики – всего лишь шум. Исследователь всегда воспринимает сведения, показания как определенный результат анализа реальности в ее целостности, рассмотрении в определенном «ментальном» ракурсе. Это означает, что в действительности никаких «сырых» данных (неструктурированных показателей), которые потом преобразуются в «smart» данные, не существует. Сама процедура регистрации, записи, описания значения переменных и есть способ представления и определения совокупности показаний, которые становятся данными уже на этапе, предшествующему их генерированию.

Так что же тогда такое *данные*? С одной стороны, они рассматриваются как нечто, созданное нашим восприятием, инструментами, или технологиями. В то же время, очевидно, что они «существуют» как бы вне нас со всем нашим инструментальным миром познавательной деятельности, как нечто *данное*. Чтобы разрешить это противоречие, полагаю, следует обратить внимание на то, как в языке, в семантической эволюции развития этого термина, отражается когнитивная «процедура» раскрытия самой природы термина *данные* как границы бытия и знания того, как реализуется переход от любых фактов к исчисляемому.

2. Когнитивная трансформация понятия данные

Диахронический анализ понятия «данные» следует начать с обращения к наиболее популярным (как бы «классическим») словарям. Ведь включение слов или фразеологизмов в словарь означает не только фиксацию статуса их как элемента общеупотребительной лексики, но как и процедуру их нормирования (нормализации). Поэтому крайне важно рассмотреть, как толкуют словари слово «данные» в том или ином языке. Так, согласно «*Оксфордскому словарю современного английского языка*»⁴ (1993), *данные* – это числовые и нечисловые значения характеристик кого-либо (чего-либо), операции с которыми проводит компьютер или какое-нибудь другое подобное устройство; данные представляют собой набор значений качественных или количественных переменных; отдельные фрагменты информации⁵. В «Новом словаре русского языка»⁶ (2000) термин *данные* определяются как «сведения, факты, характеризующие кого-либо, что-либо, необходимые для каких-либо выводов, решений» [19, с. 6]. Таким образом, в соответствии со словарным значением слово *данные* представляет собой набор значений качественных или количественных переменных.

Между тем подобное, как нередко считается, общепринятое истолкование *данных*, как основы вычисления при решении математических задач, в действительности возникло и сформировалось относительно недавно. Оно представляет собой результат семантической и когнитивной трансформации этого слова, полученной в процессе его исторического «освоения». Реконструкция происхождения слова *данные* показывает, что в русском языке оно восходит, как было упомянуто ранее, к этимологическому праславянскому гнезду **dati*, представляющему важный смысловой фрагмент в русской языковой картине мира.

Используя, как в быденном, так и в научном дискурсе это слово, мы, в то же время, не отдаем себе отчет в антропоцентрическом контексте его применения. В действительности, усваивая это слово в самом раннем возрасте посредством соци-

⁴ The Oxford Dictionary of Current English. by Della Thompson (Оксфордский словарь современного английского языка) содержит более 140 000 определений и статей, написанных для доступного освещения современного английского языка, источник всех наиболее часто используемых слов, значений, происхождения, употребления и правописания. [18, с. 4]

⁵ Перевод выполнен мною, ВТ.

⁶ «В толковом словаре Ефремовой насчитывается более 140 тысяч словарных статей толкование и значений большинства слов, примерно 11 тысяч значений слов являются отсылочными, что делает онлайн-словарь более удобным.» [19].

альной активности и обыденной речевой практики, мы воспринимаем его как отражение, некий естественный ментальный *образ* множества элементов (характеристик) реальности, которые существуют независимо от нас. Объективно! И в этом мы усматриваем основания реальности нашего познания действительности (ее освоение).

Тем самым, в нашей речи, и, соответственно, в сознании *данные* с самого начала овладения речевой деятельностью, как бы «противостоят» нам, нашей языковой личности (субъекту языка, языкового общения и речевой практики). Истоки этого противостояния заложены в исторической социальной деятельности, осваивающей и одновременно *созидающей* мир культуры, мир вещей, создаваемых и творимых *человеком* для человека. Перенимая все те специфически человеческие способы деятельности, которые предметно зафиксированы в формах создаваемых вещей, *осваиваясь* в мире этих вещей, мы усваиваем и опредмеченный в них общественно-человеческий разум с его логикой, выраженной в речевой деятельности, национальном языке (Ильенков Э. В.). Поэтому в ходе реконструкции отдельных элементов языка, словоупотребления мы раскрываем становление того или иного понятия как отражение истории практической жизни человека, формирования «следов» мышления. В поисках этих «следов», которые получили отражение в эволюции смыслов понятия *данные*, попробуем воспроизвести этимологическую реконструкцию данного слова.

Беря за точку семантического отсчета индоевропейский корень **dō-*, наш (русский) язык в своем развитии как бы порождает множество «семантико-словообразовательных центров **dati* – ‘*дать, давать, дарить*’, **danь* – ‘*дань, подать*’ – объект обязательной передачи в качестве платы’, **darь* – ‘*дар, подарок*’ – объект добровольной передачи в знак признательности, дружбы и установления мирных отношений’» [20]. Несмотря на очевидное многообразие элементов данного этимологического гнезда, мы видим, что в их смысловой основе лежит преобразованная семантика древнеиндоевропейского глагола *dā* – ‘*давать, дарить что-л.*’. Другими словами, значение понятия «данные» предстает перед нами как *нечто*, изначально независимо существующее и определяемое как бы без участия *того*, кому это нечто будет вручено, *подарено* (дано!). Но в то же время как нечто, которое уже исторически освоено и «вычленено» (выделено) из действительности, и, как результат, «присвоено» кем-то перед тем, как оно будет им передано.

Подобное семантическое «звучание» смысла слова *данные*, отмечается не только в русском или

родственных славянских языках, но и во множестве иных языков, которые восходят к индоевропейскому. Так, например, английское *data*, производное от латинского *datum* – ‘*данная вещь*’, берет свое начало от *dare* – ‘*давать*’. В этой связи, любопытно отметить, что древнеиндоевропейский корень **dō-*, в каком-то смысле, сохраняется в английском языке и в настоящее время в виде глагола неопределенной формы *do* – *делать*. Причем третья форма прошедшего времени этого глагола – *done* – и по произношению, и по правописанию коррелирует с русским словом *данные*.

В качестве еще одного примера раскрытия историчности смысла слова *данные* как отражения реальных («вещных») отношений в форме ‘*дать, давать, дарить*’ можно привести вывод, сделанный французским лингвистом Э. Бенвенистом. Так, он утверждал, что в хеттском языке корень **dō-* имел значение ‘*брать*’ и один из глагольных дериватов древнеиндоевропейского *dā* – *dāra* означал ‘*брать в жены, жениться*’. Из этого им был сделан вывод, что в индоевропейском языке, к которому восходит и хеттский язык, значения ‘*давать*’ и ‘*брать*’ синкретично соединены в одной словесной форме универсального обозначения обмена, занимающего важное место в праиндоевропейской картине мира [20, с. 82].

Важным аспектом рассмотрения возникновения и трансформации понятия *данные* представляется также анализ количественной формы использования этого слова. Она находится в прямой зависимости от ментальной среды того дискурса, контекста применения слова *данные*.

Так, если следовать «Энциклопедическому словарю Брокгауза и Ефрона», одно из сочинений Эвклида, представляющее продолжение его «Элементов», носило название «*Данные*» (δεδομένα). При этом Эвклид называет *данными* все то, что на основании теорем, доказанных в «Элементах» (предшествующего труда), непосредственно следует из условий задачи. Например, прямая, проводимая из точки по касательной к кругу, – есть *данная* по величине и положению [21, с. 133]. Другими словами, для Эвклида, заложившего основы математики, в содержании понятия *данные* отсутствовала количественная характеристика, но присутствовала деятельностьная (практическая) характеристика. Иначе говоря, *данные* – то, что мы изначально создали прежде, чем было начато математическое действие. Причины этого также обусловлены ментальными особенностями эпохи эллинизма.

Как видно из приведенного анализа, язык, речевая деятельность неразрывно связаны с историческими формами мышления, формами менталь-

ной активности. Эту зависимость не всегда удается выявить в явном виде, но в любом случае анализ семантики и грамматики словоупотребления показывает исторические особенности развития как обыденного, так и научного тезауруса.

В этой связи, кратко рассмотрим эволюцию использования существительных *data*, *datum* в английском языке, который можно рассмотреть, как языковую основу информатики. Слово *data* в значении, близком сегодняшнему, стало использоваться в начале 1600-х годов, будучи заимствованным из латинского языка, языка науки XVII-го и более ранних веков. Слово *data* произошло от латинского *datum* – «данная вещь» (см. латинское *dare* – «давать»). В английском языке оно применяется как во множественном, так и в единственном числе. Исчисляемое существительное *datum* означает «исходный факт», «заданная величина».

Первый случай использования термина «data» (*данные*) в англоязычном тексте был обнаружен в трактате 1630 г. Уильяма Баттена (William Batten), английского вице-адмирала, при составлении им навигационной таблицы, выполненной для вычисления местонахождения корабля по азимуту положения Солнца. У. Баттен называет цифры в своей таблице «данными», потому что, очевидно, он воспринимал их как данность положения корабля, поэтому сделанные записи не следует подвергать сомнению. Другими словами, по его мнению, *данные* описывают «аксиоматические» и, следовательно, неоспоримые утверждения или числа, *данные* («предоставленные») для расчета [22].

Несколько позже, в 1646 году, слово *data* (*datum*) в истолковании «куча данных» встречается в богословском трактате Генри Хаммонда (Henry Hammond) в форме утверждения – «не набор цифр, а список богословских положений, признанных истинными ради аргументации». В этом контексте использование при описании перечисленных богословских суждений слово «data» следует воспринимать как «данные», аналогично использованию У. Баттеном чисел уже как *данных* (вычисленных и внесенных) в его таблице. По мнению Генри Хаммонда слово Божие не подлежит ни сомнению, ни эмпирической проверке, его следует принимать как *данность*, как «аксиому». И понятие «данные» (терминологически как *data*) входит в английский язык как когнитивный «инструмент», сигнализирующий о результате освоения нового знания («информации»), которое следует воспринимать как данность, а не подвергать сомнению, будь то в области математики или теологии.

Осознание и «включение» в язык количественного значения терминов *datum* (*data*) произо-

шло позже в результате введения в сферу знания методов статического анализа демографических характеристик территорий Великобритании. В 1798 году Джон Синклер (Sir John Sinclair) в своем «Статистическом отчете о Шотландии» ввел в английский язык термин «*statistics*». Однако он считал, что «*statistics*» (статистика) – это собрание фактов, которые совершенно не обязательно представлены (и даже в основном не были представлены) в числовом виде [23].

И только с 1829 года слово *statistics* стало употребляться в английских документальных источниках в качестве «*собранных и классифицированных числовых данных*». Широкое распространение статистических методов в дальнейшем привело к тесной смысловой привязке термина *data* к термину *statistics* и, соответственно, пониманию *данных* как *числовых значений*. В результате с 1897 года в документальных источниках на английском языке слово *data* стало использоваться в новом значении – «*факты в числовом виде, собранные для дальнейшего использования*» [24].

В XX веке спектр значений слова *data* существенно расширяется и постепенно оформляется представление о *данных* как качественных и количественных переменных, отдельных фрагментах информации. В речевой практике появились и закрепились такие термины как «обработка данных» (*data processing*), «база данных» (*database*), «ввод данных» (*data entry*). И, наконец, в 1974 г. Питер Науэр (Peter Naur) публикует в Швеции и Соединенных Штатах книгу «Краткий обзор компьютерных методов» («Concise Survey of Computer Methods»). В этой монографии методы обработки данных соответствуют понятию *данные*, который был предложен Международной федерацией по обработке информации (IFIP). А именно: «Данные – это представление фактов или идей формализованным образом, способное передаваться или манипулироваться каким-либо способом»⁷. Исходя из этого, сам Питер Науэр определяет «науку, имеющую дело с данными»⁸, как Data Science. Пожалуй, это самое раннее сформулированное представление о Data science, как науке, работающей с данными после их «установления».

Описанная выше эволюция когнитивных смыслов понятия *данные* показывает, что анализ данных, определяемый таким образом и называемый в последствии, и как извлечение знаний, и как извлечение паттернов из данных, «добыча дан-

⁷ «A representation of facts or ideas in a formalized manner capable of being communicated or manipulated by some process» [3, c.9].

⁸ «Data science is the science of dealing with data, once they have been established, while the relation of data to what they represent is delegated to other fields and sciences» [3 c.13].

ных», и, наконец, сбор информации, в действительности отражает не столько процедуру выделения «полезных» данных из массивов зарегистрированных показателей в целом, сколько преобразование с помощью конкретных алгоритмов массивов независимых переменных, полученных при «участии» неких ментальных «усилий», в поток «исходящих» прогностических или информационных параметров. Между тем, трактовка понятия *данные*, как я считаю, зависит от того, как мы представляем процесс познания. Если в его основу мы помещаем классическую модель научного знания, согласно которой предметом исследования оказывается выделенная совокупность объектов, то наука о данных действительно имеет дело лишь с результатом их выделения. Но если мы рассматриваем науку о данных как процесс «освоения» и представления явлений и событий в когнитивном формате, то мы взаимодействуем с новым видением реальности. Возникающая в рамках науки о данных новая эпистемологическая установка подразумевает иной способ постановки исследовательских вопросов относительно производства знания, что ранее казалось не очень нужным, поскольку этот процесс рассматривался как естественное взаимодействие «наблюдателя» с реальностью. Тем самым феномен *данных*, особенно остро проявившийся в условиях всесторонней цифровизации социального бытия, ставит под сомнение устоявшуюся классическую схему представления об исследовательском процессе, как интерпретации, отображении «предстоящей» реальности.

3. Трансформация парадигмы познания

Необходимость формирования новой исследовательской парадигмы (познавательной деятельности в целом) в результате всесторонней цифровизации, получил отражение в утверждении – «большие данные» нуждаются в «большой теории». Крис Андерсон, еще в 2008 году, отмечал, что развитие инновационных технологий анализа данных указывает на новую стадию формирования знания, в то время как существующий научный метод знаменует «конец теории» [25]. Столь радикальный вывод был сделан им на основании того, что наука о данных, реализуя цели по извлечению знания из обрабатываемых данных, предстает как совокупность аналитических технологий выявления корреляций, отражающих наличие системных закономерностей и взаимосвязей в больших массивах данных.

Несмотря на привлекательность и, казалось бы, очевидность такого вывода, следует вновь отметить, что анализ данных начинается не с ис-

пользования статистических «технологий» или процедур обработки полученных данных, а с формулировки задач, которые мы хотим решить с помощью анализа данных. И сбор данных с последующей их обработкой с одной стороны – не более чем инструмент для реализации *правильно поставленной задачи*, а, с другой – некая модель, способ реализации нашей задачи, которая выражает проблему, требующую разрешения. При этом в основе, используемой/предложенной модели, лежит уже существующий в нашем совокупном знании о мире образ реальности, имманентно содержащий представление о возможности некой корреляции, которую мы намерены извлечь из «сформированного и заданного» нами набора значений, сгенерированных как «следствие» существующей корреляции. Поэтому реализуемая модель получения данных отражает, несмотря на кажущуюся непредвзятость, заложенные в нее цели, задачи, исследовательскую идеологию⁹.

В качестве разъясняющей аналогии любопытно привести описание процесса «возникновения» произведения искусства известного историка искусств Нильса Бюттнера (Nils Büttner). Создание любого произведения искусства, по его мнению, начинается с *'inventio'* (изобретения или обнаружения сущности), далее наступает стадия *'dispositio'* (этап выбора, оценки, классификации и структурирования полученного материала). И, наконец, завершается все *'elecutio'* (приданием формальным и материальным аспектам подходящей формы) [26]. Как мы видим, творческий процесс создания артефакта неотличимо совпадает с процедурой «формирования» некоего набора данных, представляющих фрагмент реальности. И это понятно, поскольку в основе этих различных форм человеческой деятельности, лежит процесс не просто освоения реальности, но воссоздание ментальных форм «присвоенной» реальности.

Предлагаемый взгляд на природу *данных* как источник знаний, «возникающий» на стыке знания и бытия в результате процесса «освоения» исследуемого фрагмента реальности, отражает формирование новой эпистемологической парадигмы, меняющей привычные процедуры интерпретации окружающего мира. Он знаменует новый подход к оценке произошедшего и прогнозу грядущего [27]. В отличие от предшествующих этапов развития науки о данных, формирующаяся парадигма характеризуется не только тем, что рассматривает современную эпоху в качестве цифровой эры анализа

⁹ И мы можем высказать «крамольное утверждение» – *данные* – это, по сути, модель представления осваиваемой в процессе осознания реальности во множестве ее особенностей и характеристик.

данных. И даже не тем, что разнообразные цифровые устройства или цифровые сети порождают посредством «заложенных» в них моделей обширные и динамичные потоки цифровых данных. Она ориентируется на то, что современное пространство социальной (человеческой) активности, преобразующей реальность и выраженной в совокупности форм создаваемых вещей, предстает как среда эффективных коммуникаций не только между людьми, но и различными предметами (Bruno Latour).

Датафикация социального пространства (*data everywhere*), в котором происходит становление и существование познаваемой реальности, приводит к тому, что мы рассматриваем реальный мир, с одной стороны, как множество данных, а с другой, как виртуальную среду коммуникационных взаимодействий различных сущностей (людей и предметов). В результате *данные* на современном этапе отражают цифровую реальность, которая оказывается порождением не только нашего знания (исследовательских практик), но и нашей социальной (человеческой) активности, «возникшей» в результате личностной социализации.

В контексте предлагаемой концепции *данные* рассматриваются как результат сложного процесса «разворачивания» процедуры структурирования предметного пространства, которое предстает перед исследователем как пространство социальной/познавательной коммуникации участников научного познания. И это пространство рассматривается в качестве конструкта, задающего концептуальный каркас исследуемого процесса научного познания. Соответственно, элементами пространства познания выступают не формы приращения знания (научное открытие, решение проблемы, фальсификация теории и т.п.), и не визуализированные формы этих приращений знания, документы, а *непосредственные* акты взаимодействия, которые и предстают в качестве анализируемых данных. Причем акты этого взаимодействия всегда конкретно историчны, и требуют дальнейшего исследования.

Заключение

Отсутствие достаточно ясного концептуального представления о том, как возникают данные на границе знания и бытия проявляется определенной многозначностью существующих определений понятия *данные*. Традиционно данные рассматриваются как первичные зарегистрированные воздействия, поступающие на органы чувств субъекта или датчики измерительного прибора. Однако определение данных как формы или способа

представления регистрируемых характеристик реальности не раскрывает их взаимосвязь с процедурами взаимодействия познающего субъекта с действительностью.

Вместе с тем, разработчики информационных технологий нередко рассматривают регистрируемые данные в качестве «сырых», не вписанных в тот или иной контекст. Только после соответствующего структурирования и выбора определенного формата представления *данные*, по их мнению, трансформируются в *исходный* «материал» для осуществления интеллектуальных операций, обеспечивающих осмысление и интерпретацию первичных зарегистрированных воздействий.

Между тем, представление о противостоянии *неструктурированных* и трансформированных («умных») данных не отражает реальную процедуру формирования знания в виде данных, «генерированных» различными источниками в процессе преобразования потока независимых переменных. Значения качественных или количественных переменных, которые характеризуют исследуемые процессы и события реальности, не могут быть получены независимо от теорий и моделей мышления исследователя. Он всегда привносит в исследование какую-либо концептуальную модель, будь она в имманентном виде, или формализована. Именно эта модель отражает совокупность исследовательских вопросов и способы их инструментальной постановки. И, соответственно, анализ начинается не с процесса обработки полученных данных, а с формулировки задачи, вопросов, которые мы хотим решить с помощью анализа данных.

Эпоха Data Science, проявляющаяся в повсеместном проникновении компьютерных технологий знаменует наступление нового эпистемологического этапа в познании реальности, так называемой «четвертой парадигмы» [The Fourth Paradigm]. Цифровизация социальной действительности, в которой происходит становление и существование познаваемой реальности, приводит к тому, что мы рассматриваем реальный мир, с одной стороны как множество данных, а с другой, как виртуальную среду коммуникационных взаимодействий не только между людьми, но и различными предметами (Bruno Latour). Результатом этого возникает цифровая реальность, которая, с одной стороны, есть порождение не только нашего знания (исследовательских практик), но и нашей социальной (человеческой) активности. И тем самым по-новому формулируется проблема данных как процедуры преодоления границы бытия и знания.

Литературы

1. *Gil Press*. A Very Short History of Data Science // Пост на сайте Forbes, updated Apr 14, 2022 [Электронный ресурс]. URL: <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/> Дата обращения 17.06.2024
2. Журавлева Е.Ю. Эпистемический статус цифровых данных в современных научных исследованиях. Вопросы философии. № 2. 2012. С. 113–123.
3. *Naur Peter*. Concise Survey of Computer Methods. Lund, Sweden. Studentlitteratur. 1974.
4. *Tuomi Ilkka*. Data Is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory // Journal of Management Information Systems. 2000. 6 (3). P. 103–117. DOI: 10.1080/07421222.1999.11518258. Дата обращения 17.06.2024.
5. «What are Data?» Australian Bureau of Statistics // [Электронный ресурс]. URL: <https://web.archive.org/web/20190419010315/http://abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+-+what+are+data>. Дата обращения 17.06.2024.
6. Смирнов Г.А. О месте данных в структуре познания // Труды ИСА РАН. 2019. Том 69 (3). С. 55–67. DOI: 10.14357/20790279190305
7. ISO/IEC 2382:2015 Information technology — Vocabulary. [Электронный ресурс]. URL: <https://webstore.iec.ch/publication/22380>. Дата обращения 18.06.2024.
8. ГОСТ 33707-2016 (ISO/IEC 2382:2015) Информационные технологии. СЛОВАРЬ. (ISO/IEC 2382:2015, MOD). М.: Стандартинформ. 2016.
9. OECD Glossary of Statistical Terms. OECD. 2008. DOI: [org/10.1787/9789264055087-en](https://doi.org/10.1787/9789264055087-en)
10. Data vs Information-Difference and Comparison [Diffen] [Электронный ресурс]. URL: https://www.diffen.com/difference/Data_vs_Information. Дата обращения 17.06.2024.
11. *Weinberger D*. The problem with the data-information-knowledge-wisdom hierarchy. Harvard Business Review. 2010. V.2/2. Feb 02. [Электронный ресурс]. URL: <https://hbr.org/2010/02/data-is-to-info-as-info-is-not>. Дата обращения 17.06.2024.
12. *Smith F. Jack*. Data Science as an Academic Discipline (англ.). Data Science Journal, 2006. 5(19). P. 163–164 // DOI:10.2481/dsj.5.163. Дата обращения 17.06.2024.
13. Тищенко В.И. Data science: новый этап сетевого анализа // Системный анализ и информационные технологии. САИТ-2019. Труды Восьмой международной конференции. 2019. М.: ФИЦ ИУ РАН. С. 370–377. DOI: 10.14357/SAIT2019048
14. *Fayyad U., Piatetsky-Shapiro G., Smyth P*. From Data Mining to Knowledge Discovery in Databases // AI Magazine. 1996; 17(3) [Электронный ресурс]. URL: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1230>. Дата обращения 17.06.2024.
15. Пятаева Н.В. Этимологическое гнездо *dati в праславянском языке (реконструкция, словообразование, семантика) // Acta Universitatis Lodziensis. Folia Linguistica Rossica. 2017. (14). P. 113–123. DOI: <https://doi.org/10.18778/1731-8025>
16. *Goldman Jeremy*. Big Data is So. 2016. We Need Smart Data. Connecting the dots is more important than ever. // Блог Inc. Grow. 2017. March 21. // [Электронный ресурс]. URL: <https://www.inc.com/jeremy-goldman/big-data-is-so-2016-we-need-smart-data.html>. Дата обращения 24.05.2023.
17. *Breiman L*. Statistical Modeling: The Two Cultures Statistical Science 2001; 16(3), 199–231 [Электронный ресурс]. URL: <http://www2.math.uu.se/~thulin/mm/breiman.pdf>. Дата обращения 17.06.2024.
18. Oxford Dictionary of Current English / Edited by Della Thompson: 2nd Edition, Oxford University Press, 1993.
19. Ефремова Т.Ф. Новый словарь русского языка. Толково-словообразовательный. М.: Русский язык. 2000.
20. Пятаева Н.В., Булина Е.В., Ибатуллина Е.А. Дань, дар, подарки и духовные дары в динамике русской языковой картины мира // Актуальные проблемы филологии и педагогической лингвистики. 2018. 29(1). P. 103–112. DOI: 10.29025/2079–6021-2018-1(29)-103-112.
21. Эвклид. Энциклопедический словарь. Т. XL. С.-Петербург. 1904; 133.
22. *Hanlon Aaron R*. Data at the Dawn of the Anthropocene // The River Rail. 2019 [Электронный ресурс]. URL: [tps://brooklynrail.org/special/River_Rail_Colby/river-rail/Data-at-the-Dawn-of-the-Anthropocene](https://brooklynrail.org/special/River_Rail_Colby/river-rail/Data-at-the-Dawn-of-the-Anthropocene). Дата обращения 18.06.2024.
23. Сеница А.Л. Первое (старое) статистическое обозрение Шотландии как источник демографических данных о населении Шотландии конца XVIII в. // Вестник Томского государственного университета. 2018; 432. P. 159–169 [Электронный ресурс]. URL: <https://www.openrepository.ru/article?id=279663>. Дата обращения 19.06.2024.

24. Кузнецов С., Константинов А., Скворцов Н. Ценность ваших данных. М.: Альпина ПРО. 2022.
25. Anderson Chris. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete // Wired. June 23, 2008 [Электронный ресурс]. URL: <https://www.wired.com/2008/06/pb-theory>. Дата обращения 18.06.2024.
26. Бюттнер Н. «Таинственный Босх: кошмары Средневековья в картинах художника. М.: Эксмо. 2019.
27. The fourth paradigm: data-intensive scientific discovery, ed. by T. Hey, S. Tansley, K. Tolle, Kristin Tolle, Jim Gray. Microsoft Research, Redmond, Washington [Электронный ресурс]. URL: https://www.researchgate.net/publication/229529541_The_Fourth_Paradigm_Data-Intensive_Scientific_Discovery. Дата обращения 18.06.2024.

Тищенко Виктор Иванович. Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия. Заведующий отделом. Кандидат философских наук. Область научных интересов: системный анализ, информатика сообществ, виртуальные сообщества. E-mail: vtichenko@mail.ru

Data as a challenge

V. I. Tishchenko

Federal Research Center Informatics and Management” of the Russian Academy of Sciences”, Moscow, Russia

Abstract: The paper deals with the problem of analyzing the nature of data arising at the boundary of knowledge and being. It shows the contradictory view of data science as a set of technologies and algorithms designed to solve problems of processing large amounts of poorly structured data, predicting correlations unrecognised earlier. The etymological and cognitive analysis of the transformation of the concept “data” is presented. The conceptual foundations of the formation of a new epistemological paradigm for displaying the surrounding world, which allows us to consider the problem of data in a new way.

Keywords: *data science, epistemology of the cognitive model.*

DOI: 10.14357/20790279240306 **EDN:** LHREBR

References

1. Gil Press. A Very Short History of Data Science // Forbes, updated Apr 14, 2022; (In Russ). Available from: <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/> [Accessed 17 June 2024].
2. Guravleva E.Ju. Epistemic status of digital data in modern scientific research. Voprosy filosofii, 2012; 2: 113—123 (In Russ).
3. Naur Peter. Concise Survey of Computer Methods. Lund, Sweden. Studentlitteratur. 1974.
4. Tuomi Ilkka. Data Is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory // Journal of Management Information Systems. 2000. 6(3): 103—117. DOI: 10.1080/07421222.1999.11518258. [Accessed 17 June 2024].
5. What are Data? Australian Bureau of Statistics // Available from: <https://web.archive.org/web/20190419010315/http://abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+-+what+are+data>. [Accessed 17 June 2024].
6. Smirnov G.A. On the place of data in the structure of cognition. Trudy ISA RAN. 2019; 69 (3). 55-67. (In Russ). DOI: 10.14357/20790279190305.
7. ISO/IEC 2382:2015 Information technology — Vocabulary. Available from: <https://webstore.iec.ch/publication/22380>. [Accessed 17 June 2024].
8. GOST 33707-2016 (ISO/IEC 2382:2015) Information technologies. Vocabulary. (ISO/IEC 2382:2015, MOD). Moscow: Standartinform. 2016.
9. OECD Glossary of Statistical Terms. OECD. 2008. DOI: [org/10.1787/9789264055087-en](https://doi.org/10.1787/9789264055087-en)
10. Data vs Information-Difference and Comparison [Diffen Available from: https://www.diffen.com/difference/Data_vs_Information. [Accessed 17 June 2024].

11. *Weinberger D.* The problem with the data-information-knowledge-wisdom hierarchy. *Harvard Business Review*. 2010. V.2/2. Feb 02. Available from: <https://hbr.org/2010/02/data-is-to-info-as-info-is-not>. [Accessed 17 June 2024].
12. *Smith F. Jack.* Data Science as an Academic Discipline (англ.). *Data Science Journal*, 2006; 5(19). 163—164 // DOI:10.2481/dsj.5.163. [Accessed 17 June 2024].
13. *Tishchenko V.I.* Data science: new stage of network analysis // *Sistemnyi Analis. SAIT-2019. Trudy Vosmoi megdunarodnoi konfercii*. 2019; M.: FRC CSC RAS. 370-377. (In Russ) DOI: 10.14357/SAIT2019048
14. *Fayyad U., Piatetsky-Shapiro G., Smyth P.* From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 1996; 17(3) // Available from: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1230>. [Accessed 17 June 2024].
15. *Pyataeva N.V.* Etymological nest *dati in the Proto-Slavic language (reconstruction, word formation, semantics) // *Acta Universitatis Lodzianensis. Folia Linguistica Rossica*, 2017; (14), 113–123. (In Russ). DOI: org/10.18778/1731-8025
16. *Goldman Jeremy.* Big Data is so. 2016. We Need Smart Data. Connecting the dots is more important than ever. // *Blog Inc. Grow*. 2017; March 21. Available from: <https://www.inc.com/jeremy-goldman/big-data-is-so-2016-we-need-smart-data.html>. [Accessed 17 June 2024].
17. *Breiman L.* Statistical Modeling: The Two Cultures *Statistical Science* 2001; 16(3), 199-231 // <http://www2.math.uu.se/~thulin/mm/breiman.pdf>. [Accessed 17 June 2024].
18. *Oxford Dictionary of Current English*. Edited by Della Thompson: 2nd Edition, Oxford University Press, 1993.
19. *Efremova T.F.* New dictionary of the Russian language. Explanatory and word-formative. 2000; M.: Russk. Yaz. (in Russ).
20. *Pyataeva N.V., Bulina E.V., Ibatullina E.A.* Tribute, gift, gifts and spiritual gifts in the dynamics of the Russian language picture of the world. *Aktualnii problem filologii i pedagogicheskoi lingvistiki*. 2018; 29(1). 103-112. (In Russ) DOI: 10.29025/2079-6021-2018-1(29)-103-112.
21. *Euclid. S.* *Encyclopedicheskii Slovar*. XL. S. – Peterburg. 1904; 133.
22. *Hanlon Aaron R.* Data at the Dawn of the Anthropocene // *The River Rail*. 2019. Available from: https://brooklynrail.org/special/River_Rail_Colby/river-rail/Data-at-the-Dawn-of-the-Anthropocene. [Accessed 17 June 2024].
23. *Sinitisa A.L.* The first (old) statistical review of Scotland as a source of demographic data on the population of Scotland at the end of the 18th century. *Vestnik Tomskogo gosudarstvennogo universiteta*. 2018; 432. 159-169 (In Russ) Available from: <https://www.openrepository.ru/article?id=279663>. [Accessed 18 June 2024].
24. *Kuznetsov S., Konstantinov A., Skvortsov N.* Value of your data. M.: Alpina PRO. 2022.
25. *Anderson Chris.* The End of Theory: The Data Deluge Makes the Scientific Method Obsolete // *Wired*. June 23. 2008; Available from: <https://www.wired.com/2008/06/pb-theory>. [Accessed 18 June 2024].
26. *Byutner N.* Mysterious Bosch: nightmares of the Middle Ages in the artist's paintings. 2019. M.: EKSMO. (In Russ)
27. *The fourth paradigm: data-intensive scientific discovery*, ed. by T. Hey, S. Tansley, K. Tolle, Kristin Tolle, Jim Gray. Microsoft Research, Redmond, Washington Available from: https://www.researchgate.net/publication/229529541_The_Fourth_Paradigm_Data-Intensive_Scientific_Discovery. [Accessed 19 June 2024].

Tishchenko Viktor I. Federal Research Center “Informatics and Management” of the Russian Academy of Sciences”, Moscow, Russia. Head of Department. Ph.D. Area of scientific interests: system analysis, community informatics, virtual communities. E-mail: vtichenko@mail.ru

Исследование распределения расстояний между графами с упорядоченными вершинами

А.А. Рогов, Н.Д. Москин, Р.В. Воронов, К.А. Кулаков

Петрозаводский государственный университет, г. Петрозаводск,
Россия

Аннотация. В статье развивается подход, основанный на вероятностной модели генерации объектов, между которыми находится расстояние. Рассматривается распределение расстояний между графами с упорядоченными вершинами на основе максимального общего подграфа. Одним из возможных вариантов применения подобных расстояний является задача стилистической диагностики текстов. Представлено два алгоритма подсчета расстояния на множестве графов. Один из них заключается в генерации и полном переборе всех пар деревьев, второй – эвристический. Это приближенный метод сбора статистики, где перебирается заданное число пар псевдослучайных деревьев, так как полный перебор может занимать много времени. С помощью этих алгоритмов были найдены матрицы расстояний между деревьями с малым и большим числом вершин n . Результаты экспериментов показали, что при малых n значение метрики не превосходит 0,5. При больших n среднее значение метрики слабо растет и стабилизируется в точке 0,587. Гипотеза о соответствии распределения нормальному закону при $n=100$ была отвергнута с помощью критерия Пирсона на уровне значимости 0,1.

Ключевые слова: граф, сравнение, метрика, максимальный общий подграф, упорядочение вершин.

DOI: 10.14357/20790279240307 **EDN:** NCISUO

Введение

Графовые модели в настоящее время используются очень активно при моделировании различных объектов. При этом возникает задача сравнения двух графовых моделей. Существуют различные метрики для определения расстояния (сходства) между ними [1]. Однако получив значения расстояния, невозможно его оценить по критерию «далеко-близко». Для этого требуется оценка частоты его встречаемости между парами графовых моделей. Решение данной проблемы позволит ответить на вопрос о том, является ли отличие между графовыми моделями существенным или возникшим случайным образом. Одним из методов получения такой оценки является поиск всевозможных числовых значений расстояний между двумя графовыми моделями. При этом большим расстоянием будем называть такое, равное которому и больше которого встречаются редко. С другой стороны, графовые модели можно называть похожими, если расстояние, меньшее расстоянию между этими моделями, встречается редко. В данной статье развивается

подход, основанный на вероятностной модели генерации объектов, между которыми находится расстояние, ранее описанный в работах [2–4].

Одним из возможных вариантов применения подобных расстояний является задача стилистической диагностики (подробнее об этом в [1]). Например, для представления синтаксической структуры текста используются так называемые деревья зависимостей (деревья подчинения) и деревья составляющих, сравнивая которые мы можем определить сходство или различие сюжетов. Это требуется, например, для решения задачи атрибуции, в частности, плагиата. Другим важным вариантом применения анализа расстояний является сравнение классификации текстов, полученных с разным набором признаков. Это позволяет выявить наиболее информативные признаки.

Остановимся более подробно на деревьях зависимостей в терминах работы [5]. Если P – конечное линейно упорядоченное множество, то всякое дерево D , для которого P служит множеством узлов, называется деревом синтаксического подчинения на P . Если P – это множество точек не-

которой цепочки x , говорят, что D – дерево синтаксического подчинения для x . Дерево зависимостей $\langle P; \rightarrow \rangle$ для цепочки x называется проективным, если для любых трех его узлов α, β, γ цепочки x из того, что $\alpha \rightarrow \beta$ и γ лежит между α и β , следует, что γ зависит от α . Дерево зависимостей $\langle P; \rightarrow \rangle$ называется слабо проективным, если для любых его четырех узлов $\alpha, \beta, \gamma, \delta$ цепочки x из того, что $\alpha \rightarrow \beta$ и $\gamma \rightarrow \delta$ следует, что пары α, β и γ, δ не разделяют друг друга. При изображении деревьев зависимостей слабая проективность означает возможность провести ребра так, чтобы никакие два из них не пересекались.

В качестве примера рассмотрим дерево зависимостей, построенное на основе текстового фрагмента из творчества Ф. М. Достоевского (роман «Идиот»): «Мы уже сказали сейчас что сам генерал хотя был человек и не очень образованный, а, напротив, как он сам выражался о себе, «человек самоучный», но был, однакоже, опытным супругом и ловким отцом» [6]. При этом нумерация вершин дерева соответствует порядку появления соответствующих слов в тексте (рис. 1): первая вершина – первому слову, вторая вершина – второму слову и т. д. В [6] показано, как с помощью деревьев зависимостей и вычисленных на их основе диагностических параметров можно распознать индивидуальный стиль таких известных писателей, как Ф. М. Достоевский, А. П. Чехов, М. А. Шолохов, Л. Н. Толстой и др. Говоря о перспективах подобных исследований И. П. Севбо отмечает, что с помощью теоретико-графовых моделей важно исследовать «многообразие деревьев зависимостей в однородном тексте (перечень типов структур)», «закономерности чередования структур, следующих друг за другом в связном тексте» и пр. В этом случае можно поставить задачу исследования расстояний между деревьями с упорядоченными вершинами.

В работе предлагается способ построения количественных оценок для понятий «редко» и «часто» на основе вероятностного распределения значений мер близости.

1. Метрики на множестве графов

Рассмотрим расстояния на множестве графов, которые позволяют оценить насколько те или иные структуры «похожи» друг на друга. Как правило, эта мера выражает степень неточностей, которые возникают при нахождении изоморфизма графов или подграфов. Одним из способов оценки возможных ошибок сравнения является максималь-

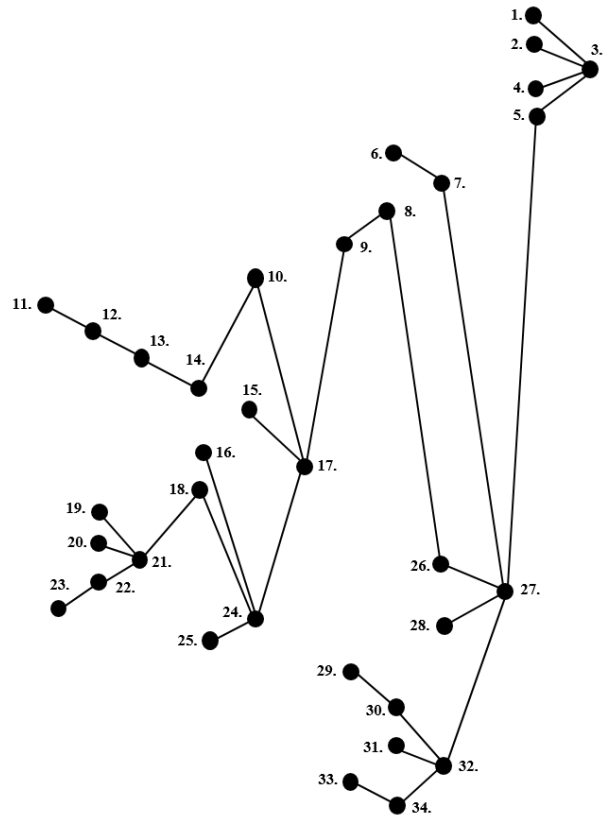


Рис. 1. Дерево зависимостей (Ф. М. Достоевский, фрагмент романа «Идиот»)

ный общий подграф $mcs(G_i, G_j)$. Очевидно, что чем больше похожи G_i и G_j , тем больше их максимальный общий подграф.

В табл. 1 представлен ряд мер (см. [1] и [7]), которых объединяет понятие максимального общего подграфа (здесь $|G|$ – число вершин графа G). Рассмотрим подробнее расстояние:

$$d(G, F) = 1 - \min_{i=1, \dots, k} \left(\frac{|mcs'(g_{\min(i, m)}, f_i)|}{i} \right).$$

Здесь для двух помеченных графов $G = (V, E_1)$, $F = (V_2, E_2)$ максимальным общим индуцированным помеченным подграфом $mcs'(G, F)$ (MCILS, maximum common induced labeled subgraph) графов G и F назовем граф $mcs'(G, F) = (V, E)$ с максимальным числом вершин, такой, что $V \subset V_1$, $V \subset V_2$ и $mcs'(G, F)$ изоморфен подграфам графов G и F , индуцированным подмножеством вершин V .

При подсчете d учитывается упорядоченность (нумерация) вершин (пусть для определенности число вершин в графах m меньше k , где $|G|=m$, $|F|=k$). Здесь граф g_i является подграфом G , который содержит вершины с номерами от 1 до i и все ребра G , инцидентные этим вершинам (аналогично определяются графы f_i). Эта функция

Табл. 1

Меры на основе максимального общего подграфа.

№	Расстояние между графами	Нормализовано для отрезка [0, 1]	Рассмотрена в работах
1	$d_1(G_1, G_2) = G_1 + G_2 - 2 mcs(G_1, G_2) $	Нет	Н. Bunke
2	$d_2(G_1, G_2) = 1 - \frac{ mcs(G_1, G_2) }{\max(G_1 , G_2)}$	Да	Н. Bunke, К. Shearer
3	$d_3(G_1, G_2) = 1 - \frac{ mcs(G_1, G_2) }{ G_1 + G_2 - mcs(G_1, G_2) }$	Да	W. Wallis, P. Shoubridge и др.
4	$d_4(G, F) = 1 - \min_{i=1, \dots, k} \left(\frac{ mcs'(g_{\min(i, m)}, f_i) }{i} \right)$	Да	Н. Д. Москин

удовлетворяет всем свойствам метрики (неотрицательность, тождественность, симметричность, неравенство треугольника).

В качестве примера рассмотрим три дерева, изображенные на рис. 2. Для графов G_1 и G_2 максимальный общий подграф изоморфен первому графу G_1 (аналогично для пары G_1 и G_3). Для пары G_2 и G_3 максимальный общий подграф также будет изоморфен графу G_1 . Однако меры принимают различные значения (табл. 2).

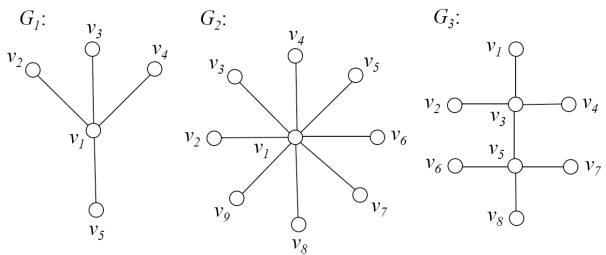


Рис. 2. Три дерева

Для того чтобы в табл. 2 подсчитать расстояния d_4 требуется поставить в соответствие каждой вершине ее порядковый номер. Предположим, номер вершины v_i совпадает с i . Тогда построим цепочки порождающих графов для G_i (в качестве примера на рис. 3 показана такая цепочка для G_1) и найдем число вершин в соответствующих максимальных общих подграфах. Затем подставив в формулу для подсчета d_4 , найдем значения метрики.

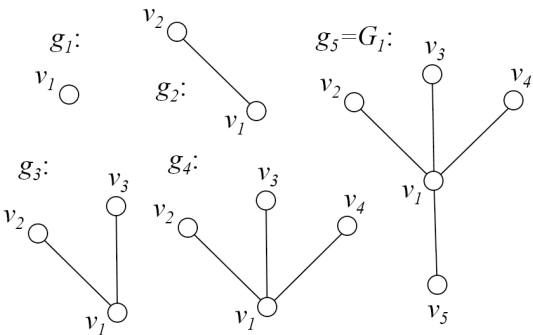


Рис. 3. Подграфы g_i графа G_1

Табл. 2

Меры d_1 - d_4 на основе максимального общего подграфа

	G_1 и G_2	G_2 и G_3	G_1 и G_3
d_1	$d_1(G_1, G_2) = 4$	$d_1(G_2, G_3) = 5$	$d_1(G_1, G_3) = 3$
d_2	$d_2(G_1, G_2) = \frac{4}{9} \approx 0,444$	$d_2(G_2, G_3) = \frac{1}{3} \approx 0,333$	$d_2(G_1, G_3) = \frac{3}{8} = 0,375$
d_3	$d_3(G_1, G_2) = \frac{4}{9} \approx 0,444$	$d_3(G_2, G_3) = \frac{5}{11} \approx 0,455$	$d_3(G_1, G_3) = \frac{3}{8} = 0,375$
d_4	$d_4(G_1, G_2) = \frac{4}{9} \approx 0,444$	$d_4(G_2, G_3) = 0,5$	$d_4(G_1, G_3) = \frac{5}{8} = 0,625$

Далее исследуем данную метрику более подробно на конечных деревьях с не более чем заданным числом вершин.

2. Алгоритмы подсчета расстояния на множестве графов

Опишем два алгоритма, которые позволяют рассчитать функцию распределения значений метрики на конечных деревьях с не более чем заданным числом вершин.

Первый алгоритм является точным, в нем перебираются все пары деревьев с не более чем N вершинами. Для каждой пары рассчитывается значение метрики и строится функция распределения. Второй алгоритм – приближенный, в нем рассчитывается эмпирическая функция распределения путем перебора пар псевдослучайных деревьев и применением эвристического метода для расчета метрики.

Первый алгоритм (полный перебор всех пар деревьев). Деревья представляются при помощи кода Прюфера [8], который взаимно однозначно сопоставляет деревьям с n помеченными вершинами размещения с повторениями из n по $(n-2)$. Известен метод, который для каждого кода дерева строит множество его ребер.

Таким образом, для генерации всех деревьев с не более чем N вершинами необходимо для каждого $n = 1, \dots, N$ перебирать все размещения с повторениями из n по $(n-2)$ и для каждого размещения получать множество ребер дерева. Обозначим такой алгоритм генерации через $Generate(N)$. Каждый вызов такого алгоритма возвращает множество ребер очередного дерева. Рассмотрим алгоритм, в котором сгенерированные множества ребер всех деревьев запоминаются в массиве *tree*:

```

Generate_Trees(N)
Вход: число вершин в дереве N
Выход: множество деревьев tree и их число M
M = 0
foreach E in Generate(N)
    M = M + 1
    tree[M] = E
return tree, M

```

Далее для каждой пары различных деревьев ищутся максимальные общие индуцированные помеченные подграфы порождающих подграфов. Пусть $T = (V, E)$ - помеченное дерево, $V = \{1, \dots, N\}$.

Порождающим подграфом $T(i)$ назовем подграф дерева T , индуцированный подмножеством вершин $\{1, \dots, i\}$.

Пусть $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ – два помеченных графа. Назовем модулярным пересечением двух помеченных графов $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ граф $G = (V, E)$ такой, что

$$V = V_1 \cap V_2 \text{ и } (v_1, v_2) \in E \Leftrightarrow (v_1, v_2) \in E_1 \text{ и } (v_1, v_2) \in E_2 \\ \text{или } (v_1, v_2) \notin E_1 \text{ и } (v_1, v_2) \notin E_2.$$

Для поиска MCILS при помощи алгоритма Брона – Кербоша [9] ищется максимальная клика в модулярном пересечении графов. Обозначим алгоритм поиска максимальной клики через $Bron_Kerbosh(G)$. На входе граф G , на выходе – число вершин в максимальной клике.

Приведем псевдокод алгоритма построения модулярного пересечения подграфов из двух цепочек порождающих подграфов.

```

MPG(E1, E2, n)
Вход: множества ребер первого и второго дерева E1, E2, максимальный используемый номер вершин n
Выход: множество ребер модулярного произведения E
E = 0
for v1 = 1 to n
    for v2 = 1 to n
        if (v1, v2) in E1 = (v1, v2) in E2
then
    add(E, (v1, v2))

```

Приведем псевдокод алгоритма расчета метрики для двух деревьев.

```

Metric(E1, E2)
Вход: множества ребер первого и второго деревьев E1, E2
Выход: значение метрики
m1 = число вершин первого графа
m2 = число вершин второго графа
k = max(m1, m2)
d = 1
for i = 1 to k
    G = MPG(E1, E2, i)
    c = Bron_Kerbosh(G)
    d = min(d, c / i)
return 1-d

```

Подсчет статистики для значений метрики осуществляется следующим образом (M – число всех деревьев, полученных процедурой $Generate_Trees$):

```

for  $i = 1$  to  $M - 1$ 
  for  $j = i + 1$  to  $M$ 
     $d = \text{Metric}(\text{tree}[i], \text{tree}[j])$ 
     $\text{count}[d] = \text{count}[d] + 1$ 

```

Здесь $\text{count}[d]$ равно числу пар деревьев, значение метрики между которыми равно d .

Второй алгоритм (эвристический). Так как полный перебор работает очень долго, предлагается следующий приближенный метод сбора статистики с помощью имитационного моделирования. Перебирается заданное число пар псевдослучайных деревьев. Каждое дерево получается из псевдослучайно полученного кода Прюфера. При этом строится псевдослучайное размещение с повторениями (массив из $n-2$ элементов заполняется псевдослучайными целыми числами из интервала от 1 до n). Для полученной пары деревьев применяется приближенный алгоритм расчета метрики. Здесь при поиске максимальной клики в модулярном пересечении применяется следующий эвристический алгоритм. На первом шаге ищется вершина u с максимальной степенью и формируется множество $U = \{u\}$. Далее на каждом шаге среди вершин $V \setminus U$ ищется вершина с максимальной степенью, смежная со всеми вершинами множества U . Если такой вершины нет, то алгоритм завершает работу. Если такая вершина найдена, то она добавляется в множество U и делается переход на следующий шаг. Отметим, что задача о клике относится к классу NP-полных задач. Как и для других NP-полных задач, эффективного алгоритма для поиска клики достаточно большого размера на данный момент не найдено.

3. Анализ статистики расстояний

Рассмотрим возможные значения расстояний для деревьев с заданным числом вершин n . В табл. 3,4 приведены значения расстояний между возможными парами деревьев. Из табл. 3 видно, что при малых n значение метрики не превосходит 0,5. Однако при больших n ситуация меняется. Как видно из табл. 4, среднее значение метрики слабо растет и стабилизируется в точке 0,587. 90%-ный диапазон значений метрики сужается с ростом n и является более островершинным по сравнению с нормальным законом. Коэффициент асимметрии – отрицательный и приближается к нулю с ростом n . Как видно из рис. 4, распределение значений расстояний является мультимодальным. С ростом n количество мод растет. Попытка аппроксимации распределе-

Табл. 3

Распределение значений метрики для пар деревьев при малых значениях числа вершин n

Значение метрики d	Доля графов пар графов, со значением метрики не больше, чем d
$n = 3$	
0	0,333
0,5	1
$n = 4$	
0	0,125
0,25	0,208
0,33	0,467
0,5	1
$n = 5$	
0	0,036
0,2	0,068
0,25	0,214
0,33	0,485
0,4	0,5
0,5	1
$n = 6$	
0	0,007
0,167	0,018
0,2	0,064
0,25	0,211
0,33	0,517
0,4	0,532
0,5	1

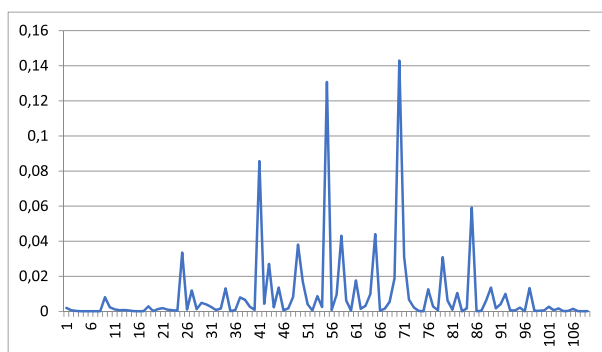
ния расстояний с помощью нормального закона при $n=100$ закончилась неудачей. Соответствующая гипотеза была отвергнута с помощью критерия Пирсона на уровне значимости 0,1.

Последние 4 столбца табл. 4 позволяют оценить редкие расстояния. Значения, приведенные в этих столбцах, можно назвать критическими. Если будем считать, что близких объектов не более 10% среди всех, то используя столбец 5, получим при $n=10$, что маленьким расстоянием будем считать любое расстояние меньшее 0,453. Можно заметить, что критическое значение длины маленького расстояния растет с ростом n , от 0,453 ($n=10$) до 0,561 ($n=100$). Аналогично определяются аномально большие расстояния. Расстояния между критическими значениями (при $n=10$ от 0,453 до 0,625) можно считать часто встречающимися.

Табл. 4

Распределение значений метрики для пар деревьев при разных значениях числа вершин n

1. К-во вершин	2. Среднее значение	3. Коэффициент асимметрии	4. Длина малых расстояний с общим количеством меньшим 5 % меньше	5. Длина малых расстояний с общим количеством меньшим 10 % меньше	6. Длина больших расстояний с общим количеством меньшим 10 % больше	7. Длина больших расстояний с общим количеством меньшим 5 % больше
10	0,562	-0,317	0,440	0,453	0,625	0,652
15	0,575	-0,255	0,469	0,495	0,638	0,649
20	0,579	-0,230	0,486	0,509	0,630	0,648
30	0,583	-0,187	0,530	0,533	0,632	0,633
40	0,585	-0,139	0,531	0,549	0,623	0,627
50	0,586	-0,133	0,540	0,553	0,619	0,624
75	0,587	-0,103	0,548	0,560	0,613	0,621
100	0,587	-0,085	0,556	0,561	0,610	0,616

Рис. 4. Распределение частот расстояний для $n=50$

В качестве примера в табл. 5 представлена матрица расстояний между графами фрагментов фольклорных и литературных текстов (описание в [5 и 6]). Первый фрагмент Ф.М. Достоевского (см. рисунок во введении, число вершин $n=34$), автор второго фрагмента М.Ю.Лермонтов ($n=19$), третьего фрагмента – А. С. Пушкин ($n=21$), четвертый фрагмент взят из Голубиной книги ($n=29$), пятый – из былины «Илья Муромец» ($n=20$). Шестой и седьмой фрагменты принадлежат Ф.

М. Достоевскому ($n=22, 27, 41$ соответственно). Анализируя табл. 5, можно заметить, что в ней самое маленькое расстояние 0,429, а самое большое 0,732. Это согласуется с проведенными экспериментами. Для более детального объяснения различий между графами требуется помощь специалистов-филологов.

Заключение

В основе предложенной математической модели лежит идея генерации графов с упорядоченными вершинами, между которыми находится расстояние. Были реализованы два алгоритма: эвристический и полный перебор с заданными условиями (при этом деревья представлялись при помощи кода Прюфера). При малом числе вершин n значение метрики не превосходит 0,5. При больших n среднее значение метрики слабо растет и стабилизируется в точке 0,587. Гипотеза о соответствии распределения нормальному закону при $n=100$ была отвергнута на уровне значимости 0,1. В дальнейшем данный подход можно применить при исследовании других метрик и подмножеств графов.

Табл. 5

Матрица расстояний между графами фрагментов фольклорных и литературных текстов.

	1	2	3	4	5	6	7	8
1	0	0,676	0,667	0,588	0,706	0,667	0,618	0,585
2	0,676	0	0,667	0,552	0,5	0,667	0,63	0,732
3	0,667	0,667	0	0,552	0,429	0,5	0,556	0,683
4	0,588	0,552	0,552	0	0,655	0,552	0,667	0,61
5	0,706	0,5	0,429	0,655	0	0,6	0,571	0,707
6	0,667	0,667	0,5	0,552	0,6	0	0,556	0,732
7	0,618	0,63	0,556	0,667	0,571	0,556	0	0,667
8	0,585	0,732	0,683	0,61	0,707	0,732	0,667	0

Литература

1. *Москин Н.Д.* Теоретико-графовые модели, методы и программные средства интеллектуального анализа текстовой информации на примере фольклорных и литературных произведений. Дис. ... докт. техн. наук. Петрозаводск. 2022. 370 с.
2. *Варфоломеев А.Г., Кириков П.В., Рогов А.А.* Вероятностный подход к сравнению расстояний между подмножествами конечного множества // Ученые записки Петрозаводского государственного университета. 2010. №8(113). С. 83-88.
3. *Сидоров Ю.В., Кириков П.В., Рогов А.А.* Сравнение дендрограмм с равным числом вершин // Ученые записки Петрозаводского государственного университета. Серия: Естественные и технические науки. 2011. № 8 (121). С. 108-110.
4. *Rogov A.A., Varfolomeyev A.G., Timonin A.O., Proenca K.A.* A probabilistic approach to comparing the distances between partitions of a set // Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes. 2018. Vol. 14, № 1. P. 14-19. DOI: 10.21638/11701/spbu10.2018.102.
5. *Гладкий А.В.* Синтаксические структуры естественного языка. М.: ЛКИ. 2007. 152 с.
6. *Севбо И.П.* Графическое представление синтаксических структур и стилистическая диагностика. Киев: Наукова Думка. 1981. 192 с.
7. *Москин Н.Д.* Метрика для сравнения графов с упорядоченными вершинами на основе максимального общего подграфа // Прикладная дискретная математика. 2021. № 52. С. 105-113. DOI: 10.17223/20710410/52/7.
8. *Prüfer H.* Neuer Beweis eines Satzes über Permutationen (нем.) // Archiv für Mathematik und Physik. 1918. Bd. 27. P. 742–744.
9. *Bron C., Kerbosh J.* Algorithm 457 – Finding all cliques of an undirected graph // Communications of the ACM. 1973. Vol. 16. P. 575–577.

Рогов Александр Александрович. Петрозаводский государственный университет, г. Петрозаводск, Россия. Заведующий кафедрой. Доктор технических наук, профессор. Область научных интересов: математическое моделирование, прикладная статистика, математические методы распознавания образов, математические методы анализа литературных текстов. E-mail: rogov@petsu.ru

Москин Николай Дмитриевич. Петрозаводский государственный университет, г. Петрозаводск, Россия. Профессор. Доктор технических наук, доцент. Область научных интересов: цифровые гуманитарные науки, теоретико-графовые модели, интеллектуальный анализ данных, компьютерная лингвистика, мультимедиа-технологии, компьютерная графика. E-mail: moskin@petsu.ru (ответственный за переписку)

Воронов Роман Владимирович. Петрозаводский государственный университет, г. Петрозаводск, Россия. Профессор. Доктор технических наук, доцент. Область научных интересов: математическое моделирование, задачи оптимизации, комбинаторные задачи на графах, математические методы и модели систем локального позиционирования мобильных объектов. E-mail: rvoronov@petsu.ru

Кулаков Кирилл Александрович. Петрозаводский государственный университет, г. Петрозаводск, Россия. Доцент. Кандидат физико-математических наук, доцент. Область научных интересов: интеллектуальные пространства, сетевые технологии, мобильные приложения, электронный туризм, мониторинг промышленного оборудования, робототехника, анализ естественных языков. E-mail: kulakov@cs.karelia.ru

Research of the distribution of distances between graphs with ordered vertices

A.A. Rogov, N.D. Moskin, R.V. Voronov, K.A. Kulakov

Petrozavodsk State University, Petrozavodsk, Russia

Abstract. The article develops an approach based on a probabilistic model of generating objects with a distance between them. The distribution of distances between graphs with ordered vertices based on the maximum common is considered. One of the possible applications of such distances is the task of stylistic diagnostics of texts. Two algorithms for calculating distances on a set of graphs are presented. One of them consists of generating and exhaustively enumerating all pairs of trees, the second is heuristic. This is an approximate method of collecting statistics, where a given number of pairs of pseudo-random trees are iterated, since a complete search can take a long time. Using these algorithms, distance matrices between trees with a small and large number of vertices n were found. The experimental results showed that for small n the metric value does not exceed 0,5. For large n the average value of the metric grows slightly and stabilizes at the point 0,587. The hypothesis that the distribution corresponds to the normal law for $n=100$ was rejected using the Pearson test at a significance level of 0,1.

Keywords: *graph, comparison, metric, maximum common subgraph, vertex ordering.*

DOI: 10.14357/20790279240307 **EDN:** NCISUO

References

1. Moskin N.D. Teoretiko-grafovye modeli, metody i programmnye sredstva intellektual'nogo analiza tekstovoj informacii na primere fol'klornyh i literaturnyh proizvedenij [Graph-theoretical models, methods and software for the intellectual analysis of textual information on the example of folklore and literary works]. D.Sc. Diss. Petrozavodsk. 2022. 370 p. (In Russ.)
2. Varfolomeev A.G., Kirikov P.V., Rogov A.A. Veroyatnostnyj podhod k sravneniyu rasstoyanij mezhdru podmnozhestvami konechnogo mnozhestva [A probabilistic approach to comparing distances between subsets of a finite set]. Uchenye zapiski Petrozavodskogo gosudarstvennogo universiteta [Scientific notes of Petrozavodsk State University]. 2010. 8(113). P. 83-88. (In Russ.)
3. Sidorov Y.V., Kirikov P.V., Rogov A.A. Sravnenie dendrogramm s ravnym chislom verшин [Comparison of dendrograms with an equal number of vertices]. Uchenye zapiski Petrozavodskogo gosudarstvennogo universiteta [Scientific notes of Petrozavodsk State University]. 2011. 8(121). P. 108-110. (In Russ.)
4. Rogov A.A., Varfolomeyev A.G., Timonin A.O., Proenca K.A. A probabilistic approach to comparing the distances between partitions of a set. Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes. 2018. 14(1): 14-19. DOI: 10.21638/11701/spbu10.2018.102.
5. Gladkij A.V. Sintaksicheskie struktury estestvennogo yazyka [Syntactic structures of natural language]. Moscow, LKI Publ. 2007. 152 p. (In Russ.)
6. Sevbo I.P. Graficheskoe predstavlenie sintaksicheskikh struktur i stilisticheskaya diagnostika [Graphic representation of syntactic structures and stylistic diagnostics]. Kiev: Naukova Dumka. 1981. 192 p. (In Russ.)
7. Moskin N.D. Metrika dlya sravneniya grafov s uporyadochennymi vershinami na osnove maksimuma obshchego podgrafa [Metric for comparing graphs with ordered vertices based on maximum common subgraph]. Prikladnaya diskretnaya matematika [Applied discrete mathematics]. 2021. 52. P. 105-113. DOI: 10.17223/20710410/52/7. (In Russ.)
8. Prüfer H. Neuer Beweis eines Satzes über Permutationen (нем.) // Archiv für Mathematik und Physik. 1918. 27: 742–744.
9. Bron C., Kerbosh J. Algorithm 457 – Finding all cliques of an undirected graph // Communications of the ACM. 1973. 16. P. 575–577.

Rogov Aleksandr A. Head of Department, Doctor of Technical Sciences, Professor, Petrozavodsk State University, 33 Lenin str., Petrozavodsk, 185910, Russia. Research interests: mathematical modeling, applied statistics, mathematical methods of pattern recognition, mathematical methods of analysis of literary texts. E-mail: rogov@petsu.ru

Moskin Nikolai D. Professor, Doctor of Technical Sciences, Associate Professor, Petrozavodsk State University, 33 Lenin str., Petrozavodsk, 185910, Russia. Research interests: digital humanities, graph-theoretical models, data mining, computational linguistics, multimedia technologies, computer graphics. E-mail: moskin@petsu.ru

Voronov Roman V. Professor, Doctor of Technical Sciences, Associate Professor, Petrozavodsk State University, 33 Lenin str., Petrozavodsk, 185910, Russia. Research interests: mathematical modeling, optimization problems, combinatorial graph problems, mathematical methods and models of local positioning systems for mobile objects. E-mail: rvoronov@petsu.ru

Kulakov Kirill A. Associate Professor, PhD in Physics and Mathematics, Associate Professor, Petrozavodsk State University, 33 Lenin str., Petrozavodsk, 185910, Russia. Research interests: intelligent spaces, network technologies, mobile applications, electronic tourism, monitoring of industrial equipment, robotics, natural language analysis. E-mail: kulakov@cs.karelia.ru

Распознавание образов

Оценка обобщающей способности алгоритма вложенных контуров при анализе маммограмм*

И.А. Егошин

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Марийский государственный университет», г. Йошкар-Ола, Россия

Аннотация. В работе представлен алгоритм вложенных контуров, предназначенный для обнаружения патологических изменений, которые могут соответствовать раку молочной железы, на рентгеновских маммографических изображениях. Приведены результаты оценки его обобщающей способности. Данный алгоритм был протестирован на обширной выборке маммографических изображений всех возможных вариантов изменений, соответствующих верифицированному раку молочной железы, включая нечетко видимые и невидимые. Общая точность обнаружения алгоритма составила 90,73% – для пленочных и 96,82% – для цифровых маммограмм. Приведен сравнительный анализ использования данного алгоритма и других современных методик обнаружения изменений на маммограммах с использованием общедоступных баз данных (INbreast и CBIS-DDSM). Показана более высокая точность предложенного алгоритма. Высокая эффективность обнаружения патологических изменений, вне зависимости от различия характеристик маммограмм, полученных на разных системах, свидетельствует о высокой обобщающей способности предложенного алгоритма.

Ключевые слова: рак молочной железы, маммография, обнаружение, линии уровня, обобщающая способность.

DOI: 10.14357/20790279240308 **EDN:** NXRKLK

Введение

Рак молочной железы (РМЖ) – это онкологическое заболевание, являющееся главной причиной смерти у женщин [1], поэтому актуальными являются любые подходы к раннему его выявлению, основным из которых на сегодняшний день является маммография [2]. Однако маммографические изображения характеризуются очень большой вариабельностью характеристик как фона, так и патологических изменений, которые могут

соответствовать РМЖ. Достаточно часто РМЖ визуально не определяется на интенсивном фоне, что является основной причиной пропуска данной патологии, особенно на ранней стадии, врачами-рентгенологами [3].

Существуют различные автоматические системы медицинской диагностики, позволяющие обнаруживать широкий круг патологических изменений, применение которых в клинической практике перспективно для первичной обработки маммографических изображений [4].

Методы цифровой обработки изображений, реализованные в данных системах, можно разделить на три группы: методы, основанные на ис-

* Основные результаты разделов «Материалы и методы» и «Результаты» получены И.А. Егошиным при поддержке гранта Российского научного фонда № 24-21-00031, <https://rscf.ru/project/24-21-00031/>

пользовании адаптивного порога [5], методы, основанные на анализе окружающих областей [6] и модели глубокого обучения [7,8].

Однако все они имеют ряд недостатков при анализе маммограммических изображений. Например, широкая вариабельность как размеров патологических изменений, так и их пространственной локализации, затрудняет выбор размера окна обработки для расчета адаптивного порога. Результат использования методов, основанных на анализе окружающих областей, может быть неустойчив, когда текстурные признаки нормальных структур и тканей молочной железы (МЖ) схожи с таковыми патологического очага, что достаточно характерно для маммограмм. Сети глубокого обучения на сегодняшний день оказались одним из перспективных способов сегментации изображений из-за способности автоматически извлекать признаки изображения и обрабатывать сложные данные. Однако к их недостаткам можно отнести то, что результат прогнозирования может быть недостаточно устойчив из-за уровня обобщения модели, когда невозможно обобщать новые ситуации или контексты, которые не были представлены в данных, использовавшихся для обучения.

Кроме того, важной проблемой использования рассмотренных методов цифровой обработки для анализа маммографических изображений и внедрения их в различные компьютерные системы компьютерной диагностики РМЖ является то, что характеристики самих маммограмм также зависят от аппаратуры, на которой производится их получение, и могут варьироваться в широких пределах. Данная особенность требует дополнительной адаптивности (обобщаемости) при внедрении подобных методов в клиническую практику.

Ранее нами был разработан алгоритм вложенных контуров (ABK), который продемонстрировал хорошие результаты при идентификации широкого круга изменений, соответствующих РМЖ, в том числе плохо видимых и невидимых при визуальном их анализе [9]. Поэтому целью работы было оценить его обобщающую способность при обработке маммографических изображений, полученных из разных источников и имеющих разные характеристики.

1. Материалы и методы

Материалы. При выполнении маммографии стандартом является получение двух изображений каждой МЖ (соответственно, четырех — на пациентку), позволяющих отобразить различные отделы МЖ, поскольку полный охват всего объема

паренхимы МЖ при получении только одного изображения невозможен. Для разработки и оценки эффективности алгоритма обнаружения изменений МЖ использовался собственный набор из 356 пленочных и 252 цифровых маммограмм (всего 608 маммограмм), полученных только из частной базы данных (публичные базы данных при разработке не использовались). Присутствующие типы изменений МЖ представлены в табл. 1.

Табл. 1
Рассмотренные типы изменений МЖ

Тип изменения	Количество случаев
Звездчатые образования	322
Спикулированные с небольшим и плотным центром и длинными спикулами	14
Изменения неправильной формы с нечеткими контурами	58
Дольчатые и круглые с полностью или частично четкими контурами	60
Частично визуализируемые	22
Асимметрия плотности	46
Нечетко видимые или невидимые	16
Изменение структуры паренхимы	8
Не визуализирующиеся образования на фоне плотной паренхимы	38
Образования округлой или овальной формы	16
Тень, ассоциированная с микрокальцинатами	8
Всего	608

Для оценки обобщающей способности предложенный ABK тестировался также на цифровых маммограммах, полученных из публичных баз данных (БД) CBIS-DDSM и INbreast, которые не участвовали при разработке алгоритма:

- 257 маммограмм из БД CBIS-DDSM [10] (цифровые маммограммы в формате DICOM);
- 116 маммограмм из БД INbreast [11] (цифровые маммограммы в формате DICOM).

В табл. 2 представлены присутствующие типы изменений МЖ на маммограммах, полученных из публичных БД.

Табл. 2
Рассмотренные типы изменений МЖ в БД CBIS-DDSM и INbreast

Тип изменения	Количество в CBIS-DDSM	Количество в INbreast
Звездчатые образования	70	30
Образования с нечетким контуром	80	34
Образования с ровным контуром	18	14

Асимметрии	6	4
Частично срезанные образования	10	—
Образования менее 10 мм	73	34
Всего	257	116

Недостатком рассмотренных публичных БД является то, что в них практически отсутствуют плохо видимые и невидимые случаи, представляющие собой основную диагностическую проблему.

Алгоритм обнаружения патологических изменений на маммограммах. Маммограммы представляли собой изображения с глубиной цвета 8 бит и серой шкалой яркости, где пиксели изображения с наибольшей яркостью соответствуют более плотной ткани молочной железы (рис. 1, а), то есть каждый i -й пиксель изображения представлен своими координатами x_i, y_i и яркостью $B = I(x_i, y_i)$.

Этапы алгоритма.

Этап 1. Строятся линии уровня для всего яркостного диапазона маммограммы (рис. 1, б), с постоянным шагом градации k :

$$\begin{cases} 1, & I(x, y) = B \geq kn \\ 0, & I(x, y) = B < kn \end{cases}$$

где $0 < kn < 255$, $n = 1, \dots, 256/k$, $k = \text{const}$, $B = I(x, y)$ – яркость пикселя изображения, представленного своими координатами (x, y) .

С помощью алгоритма жука (алгоритм обхода бинаризованных пикселей) на каждом шаге гра-

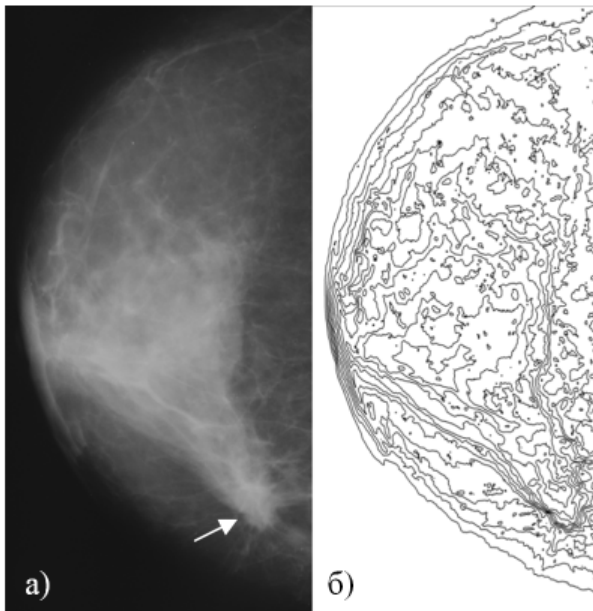


Рис. 1. Маммографическое изображение (а) и его линии уровня (б). Стрелка указывает на образование, соответствующее РМЖ

дации n обходятся все полученные необнуленные пиксели изображения и сохраняются контуры линий уровня $U_{n,i}$, где n – текущий шаг по градации яркости, $i = 1, \dots, M$, M – количество линий уровня, найденных на n -м шаге градации.

Этап 2. Из набора линий уровня последовательно выделяются серии вложенных друг в друга. Если все пиксели рассматриваемой линии уровня находятся внутри области другой линии, полученной на следующем шаге градации, то такая линия является вложенной:

$$\begin{cases} \text{вложенный,} & \frac{S(U_{n,i} \cap U_{n+1,i})}{S(U_{n,i})} = 1 \\ \text{не вложенный,} & \frac{S(U_{n,i} \cap U_{n+1,i})}{S(U_{n,i})} < 1 \end{cases},$$

где: $S(U_{n,i})$ – площадь (в пикселях), занимаемая линиями уровня U , полученными на n -м шаге градации: $S(U_{n,i} \cap U_{n+1,i})$ – площадь (в пикселях) области пересечения линий уровня U полученных на n -м и $n+1$ -м шаге градации яркости.

Этап 3. Рассчитывается набор из геометрических характеристик линий уровня, который был определен по экспериментальным данным, с наименьшим коэффициентом корреляции и меньшей дисперсией:

- 1) длина линии уровня (периметр) P , равная количеству пикселей линии уровня N ;
- 2) компактность C (отношение квадрата периметра к площади линии уровня: $C = \frac{P^2}{S}$);

- 3) координаты центра масс линии уровня:

$$x_c = \frac{1}{N} \sum_{j=1}^N x_j, \quad y_c = \frac{1}{N} \sum_{j=1}^N y_j, \quad \text{где } x_j, y_j \text{ – координаты } j\text{-ой точки линии уровня, а } N \text{ – количество точек линии уровня;}$$

- 4) стандартное отклонение расстояний от центра масс линии уровня (x_c, y_c) до каждой его точки:

$$\sigma = \sqrt{\frac{1}{N} \sum_{j=1}^N (r_j - \bar{r})^2}, \quad \text{где } \bar{r} = \frac{\sum_{j=1}^N r_j}{N},$$

$$r_j = \sqrt{(x_j - x_c)^2 + (y_j - y_c)^2}.$$

- 5) эксцентриситет линии уровня, определяемый по формуле: $E = \frac{2\sqrt{(0,5A_{\max})^2 - (0,5A_{\min})^2}}{A_{\max}}$,

$$\text{где } A_{\max} = 2\sqrt{2}\sqrt{U_x + U_y + C},$$

$$A_{\min} = 2\sqrt{2}\sqrt{U_x + U_y - C},$$

$$C = \sqrt{(U_x - U_y)^2 + 4U_{xy}^2},$$

$$U_x = \frac{1}{12} + \frac{1}{N} \sum_{j=1}^N (x_j - x_c)^2,$$

$$U_y = \frac{1}{12} + \frac{1}{N} \sum_{j=1}^N (y_j - y_c)^2,$$

$$U_{xy} = \frac{1}{N} \sum_{j=1}^N (x_j - x_c)(y_j - y_c).$$

Определены нормированные центральные моменты, инвариантные относительно переноса

и изменения масштаба [12]: $\eta_{p,q} = \frac{\mu_{p,q}}{m_{0,0}^{(p+q)/2+1}},$

где: $\mu_{p,q} = \sum_{j=1}^N (x_j - x_{avg})^p (y_j - y_{avg})^q,$

$x_{avg} = m_{1,0}/m_{0,0}, y_{avg} = m_{0,1}/m_{0,0}, m_{1,0} = \sum_{j=1}^N x_j,$

$m_{0,1} = \sum_{j=1}^N y_j;$ $m_{0,0}$ – нулевой момент равен длине

линии уровня N , p и q – порядок возведения в степень соответствующего параметра при суммировании.

Для оценки степени периодичности линии уровня проведен спектральный анализ величин r_j :

$S_k = |F(r_j)|^2$, где F – преобразование Фурье, k –

номер гармоники спектра ($k = 1, \dots, N/2$). В качестве параметра, характеризующего сложность линии уровня, используется энтропия нормированных

гармоник [13]: $H = - \sum_{k=1}^{N/2} \bar{S}_k \ln \bar{S}_k$, где $\bar{S}_k = \frac{S_k}{\sum_{k=1}^{N/2} S_k}$ – доля мощности.

Этап 4. Определяются меры подобия двух вложенных линий уровня A и B , выделенных на разных градациях яркости n для разных геометрических характеристик w , полученных на Этапе 3. Используются четыре меры, включающие разно-

сти значений, евклидово расстояние и отношение периметров:

1) сумма абсолютных разностей

$$K_1(A, B) = \sum_{i=1}^M |w_i^A - w_i^B| \text{ и ее относительное значение } K_2(A, B) = \sum_{i=1}^M \left| \frac{w_i^A - w_i^B}{w_i^A} \right|, \text{ где } M -$$

количество моментов (для нормированных центральных моментов);

2) отношение $K_3(A, B) = \frac{w_A}{w_B}$ (для периметра линий уровня);

3) евклидово расстояние между центрами масс линий уровня

$$K_4(A, B) = \sqrt{(x_c^A - x_c^B)^2 + (y_c^A - y_c^B)^2};$$

4) абсолютные разности $K_5(A, B) = |w^A - w^B|$ (по

всем остальным характеристикам линии уровня).

Критические точки для мер подобия были определены на основе обучающего набора мер подобия, полученных для маммограмм без патологических изменений и с изменениями МЖ. Если контур линии уровня отличается от вложенного, полученного на предыдущем шаге градации $n-1$, делается вывод, что рассматриваемый контур линии уровня представляет собой неподозрительную область и должен быть отброшен. Чем больше найдено подобных серий вложенных контуров линий уровня, тем больше вероятность того, что эта область является подозрительной.

На рис. 2 представлена упрощенная блок-схема предложенного алгоритма. Однако аналогичные статистические характеристики могут иметь и контуры фиброзных (связок Купера) и/или сосудистых структур, что дает ложноположительные результаты для данного алгоритма. Например, наложенные тяжи, как правило, имеют вытянутую форму (рис. 3,б), а подозрительные образования более округлые (рис. 3,а), что позволяет охарактеризовать их контуры по параметру удлинения (эксцентриситету).

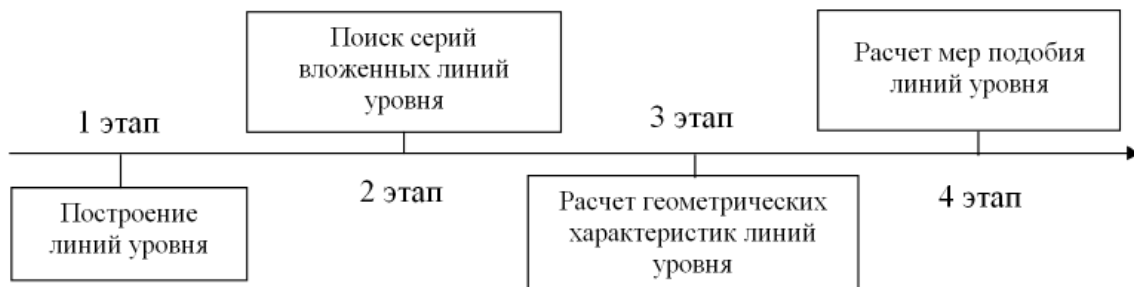


Рис. 2. Блок-схема алгоритма обнаружения изменений МЖ

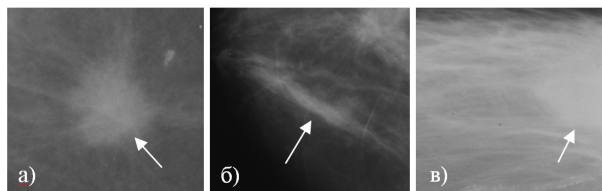


Рис. 3. Подозрительное изменение (а), фиброзный тяж (б) и частично видимое изменение (в)

Были проанализированы экспериментально полученные наборы этого параметра для подозрительных образований, фиброзных тяжей и частично видимых на маммограмме образований, которые также могут иметь удлиненную форму (рис. 3, в). В табл. 3 представлены полученные результаты. Для выборки из 110 эксцентриситетов контура были найдены его среднее значение и медиана.

Средние значения для фиброзных тяжей значительно превышают значения для подозрительного изменения и частично видимого изменения, что позволяет его отделить.

Табл. 3

Эксцентрисичность контуров

Вид Значение	Подозри- тельное изменение	Частично видимое изменение	Волок- нистые ткани
Среднее	0,582	0,793	0,929
Медиана	0,570	0,809	0,938

2. Результаты

Анализ характеристик маммографических изображений, полученных из различных источников. Аналоговые маммограммы. Источником варибельности характеристик аналоговых изображений являются параметры рентгеновского излучения, генерируемого рентгеновской трубкой и системой фильтров, особенности восприятия рентгеновской пленки и вариация параметров ее последующей обработки. Особенностью аналоговых маммограмм является отсутствие фильтров постобработки изображений. На рис. 4 представлены две пленочные оцифрованные маммограммы и построенные для них гистограммы яркости пикселей изображения МЖ.

Цифровые маммографы позволяют получить сразу цифровое изображение стандартизированного формата DICOM с разрешением 600 DPI, однако в сами программы съема заложены различные методы предобработки изображений, которые основаны на нелинейных фильтрах и могут устанавливаться во время наладки оборудования. Кроме того, они могут меняться в процессе его эксплуатации. Также оборудование разных производи-

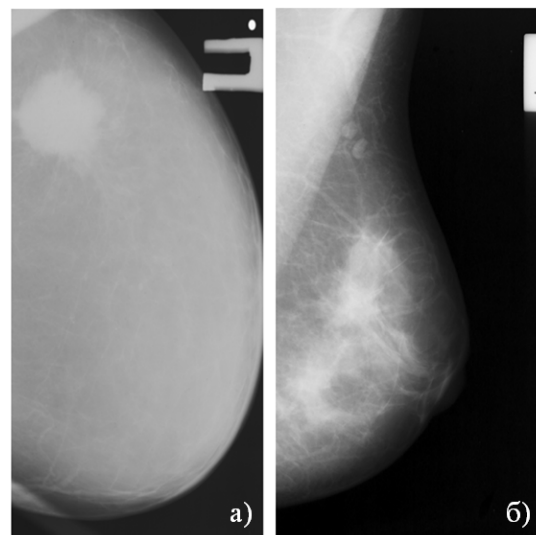
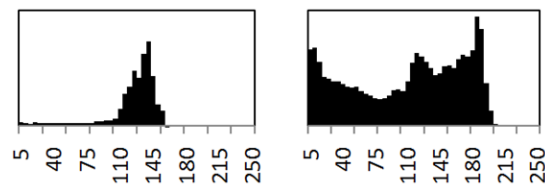


Рис. 4. Оцифрованная пленочная маммограмма с разрешением 300 DPI, полученная из частной базы данных (а); оцифрованная маммограмма, полученная из публичной базы данных MIAS [14], разрешение – 72 DPI (б)

телей может использовать разные матрицы детектора, которые имеют разную чувствительность и оптическую плотность, что также сказывается на характеристиках получаемого изображения. Отдельной проблемой является работа системы автоматической установки экспозиции в зависимости от степени поглощения рентгеновского пучка тканями исследуемой области. Это приводит к установке различных режимов работы рентгеновской трубки и, как следствие, является значимой причиной варибельности получаемых изображений. На рис. 5 представлены три цифровые маммограммы, которые соответствуют одному и тому же типу плотности МЖ, полученные на разных цифровых маммографах, и их гистограммы яркости пикселей изображения.

На рис. 5 видно, что для маммограммы, полученной на системе Siemens Mammomat Fusion, основной фон МЖ расположен в диапазоне низкой яркости от 5 до 165 градаций, медиана гистограммы – 85. Для БД CBIS-DDSM основной фон МЖ расположен в центре диапазона градаций яркости от 15 до 225, медиана – 125. БД INbreast – сильно сжата в диапазоне градаций, МЖ ущемляется в диапазон от 105 до 205 градаций яркости, медиана – 165.

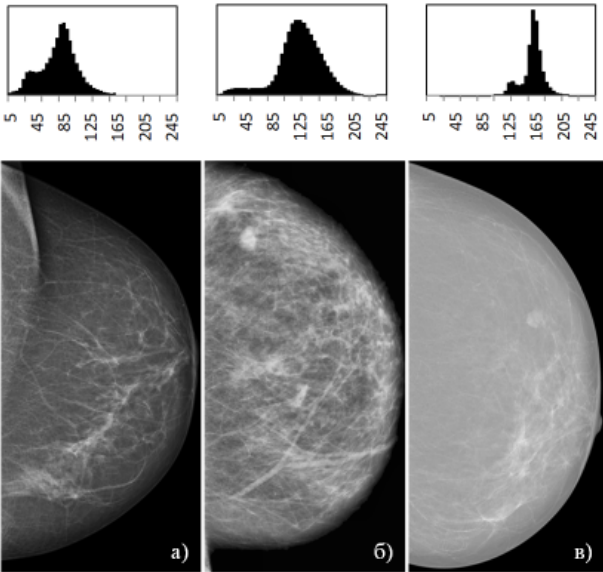


Рис. 5. Цифровая маммограмма, полученная на системе Siemens Mammomat Fusion [Германия] (а); цифровая маммограмма полученная из публичной БД CBIS-DDSM (б); цифровая маммограмма полученная из публичной БД INbreast (в)

Таким образом, из рассмотренных выше примеров можно сделать вывод, что при разработке систем автоматического обнаружения патологических изменений на маммограммах, необходимо использовать адаптивные методы, которые могли бы обобщать новые контексты характеристик изображений, не влияя при этом на качество обнаружения самих патологических изменений.

Оценка эффективности АВК. В табл. 4 представлены результаты обнаружения изменений на пленочных маммограммах с помощью АВК. Общая точность обнаружения АВК в исследуемой выборке составила 90,73 % (323 случая из 356). Средняя частота ложных меток составила 1,3 на изображение – для маммограмм жировой степени плотности и 1,8 – для маммограмм высокой степени плотности.

В табл. 5 представлены результаты обнаружения изменений на цифровых маммограммах с помощью АВК. Всего было обработано 252 цифровые маммограммы (126 пациенток). Все найденные случаи соответствовали маммограммам в какой-либо одной проекции, тогда как изменения на изображениях во второй проекции были обнаружены. Общая точность обнаружения АВК в исследуемой выборке цифровых маммограмм составила 96,82 % (244 случая из 252).

На рис. 6 представлены маммограммы на которых АВК обнаружил изменения, пропущенные врачом-рентгенологом при визуальном анализе. Врачи-рентгенологи, которым были продемонстрированы маркировки предложенного АВК, изменили свое решение о случае РМЖ на основании результатов работы алгоритма в 6 из 100 случаев, включенных в анализ.

Оценка обобщающей способности и сравнительный анализ. В табл. 6 представлены результаты обработки маммограмм из БД CBIS-DDSM, где присутствовало по два изображения (проекции) на каждую МЖ. Был зафиксирован единственный

Табл. 4

Результаты обнаружения изменений АВК на пленочных маммограммах

Тип изменения	Результат обнаружения
Звездчатые образования	205 из 220 (93,18 %)
Спикулизованные с небольшим и плотным центром и длинными спикулами	12 из 14 (85,71 %)
Изменения неправильной формы с нечеткими контурами	51 из 58 (87,93 %)
Дольчатые и круглые с полностью или частично четкими контурами	12 из 12 (100 %)
Частично визуализируемые	17 из 18 (94,4 %)
Асимметрия плотности	13 из 18 (72,2 %)
Нечетко видимые или невидимые	7 из 16 (43,8 %)

Табл. 5

Результаты обнаружения изменений АВК на цифровых маммограммах

Тип изменения	Результат обнаружения
Звездчатые образования	102 из 102 (100 %)
Образования округлой или овальной формы	16 из 16 (100 %)
Асимметрии плотности паренхимы	28 из 28 (100 %)
Не визуализирующиеся образования на фоне плотной паренхимы	32 из 38 (84 %)
Частично визуализирующиеся образования	4 из 4 (100 %)
Образования дольчатой формы	46 из 48 (95 %)
Изменение структуры паренхимы	8 из 8 (100 %)
Тень, ассоциированная с микрокальцинатами	8 из 8 (100 %)

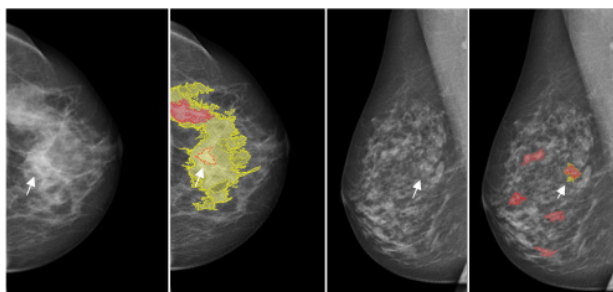


Рис. 6. Примеры маммограмм на которых разработанный метод АВК обнаружил изменения (белая стрелка), пропущенные врачом-рентгенологом

случай, когда изменение не было промаркировано на обоих изображениях. Во всех остальных случаях изменения не были промаркированы только на одном из изображений, в то время как на другом изображении пары они были промаркированы, что, как представляется, нельзя расценивать как истинно отрицательный результат работы алгоритма.

В табл. 7 представлены результаты обработки маммограмм из БД INbreast. В базе INbreast изменения, соответствующие РМЖ, не были промаркированы на пяти изображениях, однако случаев, когда они не были промаркированы на обоих изображениях одной МЖ, не было.

Результаты, приведенные в табл. 6 и 7, демонстрируют, что, несмотря на различие характеристик изображений, полученных из разных БД,

Табл. 6
Результаты работы АВК на маммограммах из БД CBIS-DDSM

Тип изменения	Результат обнаружения
Звездчатые образования	69 из 70 (98 %)
Образования с нечетким контуром	78 из 80 (97 %)
Образования с ровным контуром	17 из 18 (94 %)
Асимметрии	6 из 6 (100 %)
Частично срезанные образования	10 из 10 (100 %)
Образования менее 10 мм	70 из 73 (95 %)
Всего	250 из 257 (97 %)

Табл. 7
Результаты работы АВК на маммограммах из БД INbreast

Тип изменения	Результат обнаружения
Звездчатые образования	30 из 30 (100 %)
Образования с нечетким контуром	32 из 34 (94 %)
Образования с ровным контуром	13 из 14 (92 %)
Асимметрии	4 из 4 (100 %)
В том числе: образования менее 10 мм	32 из 34 (94 %)
Всего	111 из 116 (95,6 %)

предложенный АВК показывает высокую точность в обнаружении патологических изменений на них.

Дополнительно проведено сравнение предложенного АВК с другими современными мето-

Сравнение АВК со современными методиками

Табл. 8

База данных	Методы и источники	Год	Точность (%)
CBIS-DDSM	Shams S. и др. [15], Deep GeneRAtive Multi-task сверточная нейронная сеть.	2018	89 %
CBIS-DDSM	Tsochatzidis L. и др. [16], контролируемый подход к поиску объектов на изображении на основе его контента.	2019	81 %
CBIS-DDSM	Falconi L. и др. [17], сеть глубокого обучения VGG16.	2020	84,4 %
CBIS-DDSM	Ansar W. и др. [18], сеть глубокого обучения на основе MobileNet.	2020	74,5 %
CBIS-DDSM	Zhang H. и др. [19], классификация с кросс-модальным семантическим анализом.	2020	87,05 %
CBIS-DDSM	Himanish S.D. и др. [20], классификатор Xception с тонкой настройкой.	2022	89,2 %
CBIS-DDSM	Li H. [7], DualCoreNet.	2022	93 %
CBIS-DDSM	Предложенный алгоритм	2022	97 %
CBIS-DDSM	Jafari Z. и др. [8], CNN с выбором признаков.	2023	96%
INbreast	Dhungel N. и др. [21], каскад методов глубокого обучения.	2017	90 %
INbreast	Carneiro G. и др. [22], сверточная нейронная сеть ConvNet.	2017	90 %
INbreast	Shi P. и др. [23], сверточная нейронная сеть.	2019	83,6 %
INbreast	Zhang H. и др. [19], классификация с кросс-модальным семантическим анализом.	2020	87,93 %
INbreast	El Houbi E.M. и др. [24], сверточная нейронная сеть.	2021	94,52 %
INbreast	Himanish S.D. и др. [20], классификатор Xception с тонкой настройкой.	2022	95,1 %
INbreast	Предложенный алгоритм	2022	95,6 %
INbreast	Jafari Z. и др. [8], CNN с выбором признаков.	2023	94,5 %

диками обнаружения патологических изменений на маммограммах полученных из публичных баз данных.

Из табл. 8 видно, что предложенный АВК превосходит современные методики обнаружения изменений МЖ, несмотря на различия характеристик изображений, на которых он разрабатывался и тестировался.

Заключение

Проведенные исследования демонстрируют высокую чувствительность разработанного АВК не только в случаях типичных и четко видимых образований, но и в случаях атипичных и невидимых патологических изменений. В то же время АВК потенциально способен обнаружить образования на маммографических изображениях, характеристики которых могут варьироваться в широких пределах, что показывает высокую обобщающую способность алгоритма. С этой точки зрения представляется разумным использовать АВК в качестве части автоматической системы 2-го или 3-го анализа, чтобы отметить подозрительные зоны для их последующего детального исследования. Этот подход, вероятно, может повысить выявляемость РМЖ и, следовательно, исходы лечения в таких случаях. В то же время АВК нечувствителен к скоплениям микрокальцинатов, что требует включения дополнительного алгоритма в конструкцию системы автоматизированного анализа маммограмм. В качестве такого алгоритма может быть использована методика, ранее предложенная в работе [25]. В перспективе целесообразно разработать методы фильтрации ложноположительных результатов и, следовательно, повышения точности АВК.

Литература

1. Ferlay J., Soerjomataram I., Ervik M., Dikshit R., Eser S., Mathers C., Rebelo M., Parkin D.M., Forman D., Bray F. GLOBOCAN 2012 v1.1, Cancer Incidence and Mortality Worldwide: IARC CancerBase. No. 11. Lyon, France: International Agency for Research on Cancer, 2014. Available from: <http://globocan.iarc.fr/> [Accessed 16 January 2024].
2. United States Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement, *Annals of Internal Medicine*. 2009. 151. P. 716-726.
3. Пасынков Д.В., Егошин И.А., Колчев А.А., Ключкин И.В., Пасынкова О.О. Эффективность системы компьютерного анализа маммограмм в диагностике вариантов рака молочной железы, трудно выявляемых при скрининговой маммографии // *Российский электронный журнал лучевой диагностики*. 2019. Т.9. №2. С. 107-118. DOI:10.21569/2222-7415-2019-9-2-107-118
4. Васильев Ю.А., Тыров И.А., Владзимирский А.В., Арзамасов К.М., Шулькин И.М., Кожихина Д.Д., Пестренин Л.Д. Двойной просмотр результатов маммографии с применением технологий искусственного интеллекта: новая модель организации массовых профилактических исследований // *Digital Diagnostics*. 2023. Т.4. №2. С. 93-104. DOI:10.17816/DD321423
5. Singh N., Veenadhari S. Breast cancer segmentation using global thresholding and region merging. *International Journal of Computer Sciences and Engineering*. 2018. 6(12). P. 292-297. DOI:10.26438/ijese/v6i12.292297
6. Podgornova Y.A., Sadykov S.S. Comparative analysis of segmentation algorithms for the allocation of microcalcifications on mammograms. *Information Technology and Nanotechnology*. 2019. 2391. P. 121-127. DOI:10.18287/1613-0073-2019-2391-121-127
7. Li H., Chen D., Nailon W.H., Davies M.E., Laurenson D.I. Dual convolutional neural networks for breast mass segmentation and diagnosis in mammography. *IEEE Trans Med Imaging*. 2022. 41(1). P. 3-13. DOI:10.48550/arXiv.2008.02957
8. Jafari Z., Karami E. Breast Cancer Detection in Mammography Images: A CNN-Based Approach with Feature Selection. *Information*. 2023. 14(7). P. 1-14. DOI:10.3390/info14070410
9. Egoshin I., Pasyunkov D., Kolchev A., Kliouchkin I., Pasyunkova O. A segmentation approach for mammographic images and its clinical value. *IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems, COMCAS* 2017; 2018. P. 1-6. DOI:10.1109/COMCAS.2017.8244764
10. Lee R.S., Gimenez F., Hoogi A., Miyake K.K., Gorovoy M., Rubin D.L. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci Data*. 2017. 4. P. 170177. DOI:10.1038/sdata.2017.177
11. Moreira I.C., Amaral I., Domingues I., Cardoso A., Cardoso M.J., Cardoso J.S. INbreast: toward a full-field digital mammographic database. *Acad Radiol*. 2012. 19(2). P. 236-248. DOI: 10.1016/j.acra.2011.09.014
12. Flusser J., Zitova B., Suk T. Moments and Moment Invariants in Pattern Recognition. New York: John Wiley and Sons. 2009. 312 p.

13. *Останин С.А., Шайдук А.М., Козлов Д.Ю., Коновалов В.К., Леонов С.Л., Федоров В.В., Шойхет Я.Н., Колмогоров В.Г., Лобанов М.Н.* Энтропийный метод оценки сложности контура медицинских изображений // Известия Алтайского государственного университета. 2013. Т.1/2. №77. С. 177-180. DOI:10.14258/izvasu(2013)1.2-36
14. *Suckling J., Parker J., Dance D.* The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica. International Congress Series. 1994. 1069. P. 375-378.
15. *Shams S., Platania R., Zhang J., Kim J., Lee K., Park S.J.* Deep generative breast cancer screening and diagnosis. International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain. 2018. P. 859-867. DOI: 10.1007/978-3-030-00934-2_95
16. *Tsochatzidis L., Costaridou L., Pratikakis I.* Deep learning for breast cancer diagnosis from mammograms—A comparative study. J. Imaging. 2019. 5(3). P. 1-11. DOI:10.3390/jimaging5030037
17. *Falconi L.G., Perez M., Aguilar W.G., Conci A.* Transfer learning and fine tuning in breast mammogram abnormalities classification on CBIS-DDSM database. Adv. Sci. Technol. Eng. Syst. J. 2020. 5. P. 154-165. DOI:10.25046/aj050220
18. *Ansar W., Shahid A.R., Raza B., Dar A.H.* Breast cancer detection and localization using mobilenet based transfer learning for mammograms. International symposium on intelligent computing systems, Sharjah, United Arab Emirates. 2020. P. 11-21. DOI:10.1007/978-3-030-43364-2_2
19. *Zhang H., Wu R., Yuan T., Jiang Z., Huang S., Wu J., Hua J., Niu Z., Ji D.* DE-ada*: A novel model for breast mass classification using cross-modal pathological semantic mining and organic integration of multi-feature fusions. Inf. Sci. 2020. 539. P. 461-486. DOI:10.1016/j.ins.2020.05.080
20. *Himanish D., Akalpita D., Anupal N., Saurav M., Kangkana B., Zhongming Z.* Breast cancer detection: Shallow convolutional neural network against deep convolutional neural networks based approach. Frontiers in Genetics. 2023. 13. P. 1097207. DOI:10.3389/fgene.2022.1097207
21. *Dhungel N., Carneiro G., Bradley A.P.* A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med. image Anal. 2017. 37. P. 114-128. DOI:10.1016/j.media.2017.01.009
22. *Carneiro G., Nascimento J., Bradley A.P.* Automated analysis of unregistered multi-view mammograms with deep learning. IEEE Trans. Med. imaging. 2017. 36(11). P. 2355-2365. DOI:10.1109/TMI.2017.2751523
23. *Shi P., Wu C., Zhong J., Wang H.* Deep learning from small dataset for BI-RADS density classification of mammography images. In 2019 10th International Conference on Information Technology in Medicine and Education (ITME), Qingdao, China. 2019. P. 102-109. DOI:10.1109/ITME.2019.00034
24. *El Houbby E.M., Yassin N.I.* Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks. Biomed. Signal Process. 2021. 70. P. 102954. DOI:10.1016/j.bspc.2021.102954
25. *Пасынков Д.В., Егошин И.А., Колчев А.А., Романьчева Е.А., Ключкин И.В., Пасынкова О.О.* Обнаружение и классификация скоплений микрокальцинатов на маммографических изображениях // Медицинская техника. 2024. №1. С. 29-32.

Егошин Иван Александрович. Морийский государственный университет, г. Йошкар-Ола, Россия. Младший научный сотрудник. Область научных интересов: цифровая обработка изображений, методы глубокого обучения, математическое моделирование. E-mail: jungl91@mail.ru

Evaluation of an generalization ability of the nested contours algorithm in the mammograms analysis

I.A. Egoshin

Federal State Budgetary Educational Institution of Higher Education “Mari State University”, Yoshkar-Ola, Russia

Abstract. The work presents an nested contours algorithm designed for detecting pathological changes that may correspond to breast cancer on X-ray mammographic images, and provides the results of evaluating its generalization ability. This algorithm was tested on a large dataset of mammographic images with all possible variations of changes corresponding to verified breast cancer, including faintly visible and invisible ones. The overall detection accuracy of the algorithm was 90.73% for film and 96.82% for digital mammograms. A comparative analysis of using this algorithm and other modern methods of change detection on mammograms, with publicly available databases (INbreast and CBIS-DDSM), is also provided. The higher accuracy of the proposed algorithm is demonstrated. The high efficiency of detecting pathological changes, regardless of the differences in mammogram characteristics obtained from different systems, indicates the high generalization ability of the proposed algorithm.

Keywords: *breast cancer, mammography, detection, level lines, generalization ability.*

DOI: 10.14357/20790279240308 **EDN:** NXRKLK

References

1. Ferlay J., Soerjomataram I., Ervik M., Dikshit R., Eser S., Mathers C., Rebelo M., Parkin D.M., Forman D., Bray F. GLOBOCAN 2012 v1.1, Cancer Incidence and Mortality Worldwide: IARC CancerBase. No. 11. Lyon, France: International Agency for Research on Cancer, 2014. Available from: <http://globocan.iarc.fr/> [Accessed 16 January 2024].
2. United States Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement, *Annals of Internal Medicine*. 2009; 151: 716-726.
3. Pasyukov D.V., Egoshin I.A., Kolchev A.A., Kliouchkin I.V., Pasyukova O.O. The value of computer aided detection system in breast cancer difficult to detect at screening mammography. *REJR*. 2019; 9(2): 107-118 (In Russ). doi:10.21569/2222-7415-2019-9-2-107-118
4. Vasilev Yu.A., Tyrov I.A., Vladzimirskyy A.V., Arzamasov K.M., Shulkin I.M., Kozhikhina D.D., Pestrenin L.D. Double-reading mammograms using artificial intelligence technologies: A new model of mass preventive examination organization. *Digital Diagnostics*. 2023; 4(2): 93-104 (In Russ). doi:10.17816/DD321423
5. Singh N., Veenadhari S. Breast cancer segmentation using global thresholding and region merging. *International Journal of Computer Sciences and Engineering*. 2018; 6(12): 292-297. doi:10.26438/ijcse/v6i12.292297
6. Podgornova Y.A., Sadykov S.S. Comparative analysis of segmentation algorithms for the allocation of microcalcifications on mammograms. *Information Technology and Nanotechnology*. 2019; 2391: 121-127. doi:10.18287/1613-0073-2019-2391-121-127
7. Li H., Chen D., Nailon W.H., Davies M.E., Laurenson D.I. Dual convolutional neural networks for breast mass segmentation and diagnosis in mammography. *IEEE Trans Med Imaging*. 2022; 41(1): 3-13. doi:10.48550/arXiv.2008.02957
8. Jafari Z., Karami E. Breast Cancer Detection in Mammography Images: A CNN-Based Approach with Feature Selection. *Information*. 2023; 14(7): 1-14. doi:10.3390/info14070410
9. Egoshin I., Pasyukov D., Kolchev A., Kliouchkin I., Pasyukova O. A segmentation approach for mammographic images and its clinical value. *IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems, COMCAS* 2017; 2018. p. 1-6. doi:10.1109/COMCAS.2017.8244764
10. Lee R.S., Gimenez F., Hoogi A., Miyake K.K., Gorovoy M., Rubin D.L. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci Data*. 2017; 4:170-177. doi:10.1038/sdata.2017.177
11. Moreira I.C., Amaral I., Domingues I., Cardoso A., Cardoso M.J., Cardoso J.S. INbreast: toward a full-field digital mammographic database. *Acad Radiol*. 2012; 19(2): 236-248. doi: 10.1016/j.acra.2011.09.014
12. Flusser J., Zitova B., Suk T. Moments and Moment Invariants in Pattern Recognition. New York: John Wiley and Sons; 2009. 312 p.

13. *Ostanin S.A., Shayduk A.M., Kozlov D.Yu., Konovalov V.K., Leonov S.L., Fedorov V.V., Shoikhet Ya.N., Kolmogorov V.G., Lobanov M.N.* Entropy method for estimating the complexity of edge medical images. The News of Altai State University. 2013; 1/2(77): 177-180 (In Russ). doi:10.14258/izvasu(2013)1.2-36
14. *Suckling J., Parker J., Dance D.* The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica. International Congress Series. 1994; 1069: 375-378.
15. *Shams S., Platania R., Zhang J., Kim J., Lee K., Park S.J.* Deep generative breast cancer screening and diagnosis. International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain; 2018. p. 859-867. doi: 10.1007/978-3-030-00934-2_95
16. *Tsochatzidis L., Costaridou L., Pratikakis I.* Deep learning for breast cancer diagnosis from mammograms—A comparative study. J. Imaging. 2019; 5(3): 1-11. doi:10.3390/jimaging5030037
17. *Falconi L.G., Perez M., Aguilar W.G., Conci A.* Transfer learning and fine tuning in breast mammogram abnormalities classification on CBIS-DDSM database. Adv. Sci. Technol. Eng. Syst. J. 2020; 5: 154-165. doi:10.25046/aj050220
18. *Ansar W., Shahid A.R., Raza B., Dar A.H.* Breast cancer detection and localization using mobilenet based transfer learning for mammograms. International symposium on intelligent computing systems, Sharjah, United Arab Emirates; 2020. p. 11-21. doi:10.1007/978-3-030-43364-2_2
19. *Zhang H., Wu R., Yuan T., Jiang Z., Huang S., Wu J., Hua J., Niu Z., Ji D.* DE-ada*: A novel model for breast mass classification using cross-modal pathological semantic mining and organic integration of multi-feature fusions. Inf. Sci. 2020; 539: 461-486. doi:10.1016/j.ins.2020.05.080
20. *Himanish D., Akalpita D., Anupal N., Saurav M., Kangkana B., Zhongming Z.* Breast cancer detection: Shallow convolutional neural network against deep convolutional neural networks based approach. Frontiers in Genetics. 2023; 13: 1097207. doi:10.3389/fgene.2022.1097207
21. *Dhungel N., Carneiro G., Bradley A.P.* A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med. image Anal. 2017; 37: 114-128. doi:10.1016/j.media.2017.01.009
22. *Carneiro G., Nascimento J., Bradley A.P.* Automated analysis of unregistered multi-view mammograms with deep learning. IEEE Trans. Med. imaging. 2017; 36(11): 2355-2365. doi:10.1109/TMI.2017.2751523
23. *Shi P., Wu C., Zhong J., Wang H.* Deep learning from small dataset for BI-RADS density classification of mammography images. In 2019 10th International Conference on Information Technology in Medicine and Education (ITME), Qingdao, China; 2019. p. 102-109. doi:10.1109/ITME.2019.00034
24. *El Houbby E.M., Yassin N.I.* Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks. Biomed. Signal Process. 2021; 70: 102954. doi:10.1016/j.bspc.2021.102954
25. *Pasynkov D.V., Egoshin I.A., Kolchev A.A., Romanycheva E.A., Kliouchkin I.V., Pasynkova O.O.* Detection and Classification of Microcalcifications in Mammographic Images. Meditsinskaya tekhnika. 2024; 1: 29-32 (In Russ).

Egoshin Ivan A. Junior researcher, Federal State Budgetary Educational Institution of Higher Education “Mari State University”, 1 Lenin square, Yoshkar-Ola, Russia. E-mail: jungl91@mail.ru

Управление рисками и безопасностью

Enhancing Kubernetes Security: The Crucial Role of DevSecOps

GHADEER DARWESH, JAAFAH HAMMOUD, A.A. VOROBEOVA

ITMO University, Saint Petersburg, Russia

Abstract. This article highlights the significance of integrating DevSecOps (Development, security and Operations) practices into the research on detecting common attacks in Kubernetes environments. As Kubernetes gains rapid traction as a prominent container orchestration platform, the security challenges associated with containerized applications have grown in magnitude. However, traditional security methodologies often struggle to keep pace with the dynamic and fast-evolving nature of containerized environments, leaving potential vulnerabilities for malicious actors to exploit. By emphasizing the importance of DevSecOps, this article aims to underscore its role in improving the security posture of Kubernetes deployments and promoting a proactive approach to safeguarding containerized applications. The article also discusses key considerations and benefits of implementing DevSecOps in the context of Kubernetes security research.

Ключевые слова: *DevSecOps, Kubernetes security, DevOps, security practices.*

DOI: 10.14357/20790279240309 **EDN:** SDRKHO

Introduction

In recent years, Kubernetes has emerged as the leading container orchestration platform, revolutionizing the deployment and management of containerized applications. As Kubernetes adoption continues to soar, so does the need for robust security measures to protect these environments from emerging threats and vulnerabilities. In response, the concept of DevSecOps (Development, security and Operations) has gained prominence as a vital approach to integrating security practices throughout the entire software development lifecycle.

Kubernetes offers a highly scalable and flexible infrastructure for orchestrating containers, enabling organizations to efficiently manage complex distributed systems [1]. Its popularity stems from its ability to automate application deployment, scaling, and man-

agement, while providing features such as load balancing, service discovery, and self-healing. However, as organizations increasingly rely on Kubernetes to power critical applications, ensuring the security and integrity of these deployments becomes paramount.

The dynamic nature of containerized environments introduces unique security challenges. Misconfigurations, vulnerabilities in container images, unauthorized access, and attacks on the Kubernetes control plane are just a few examples of the risks that must be addressed. Therefore, security considerations should be an integral part of Kubernetes deployments, right from the initial development stages through to production.

DevSecOps, a methodology that combines development, operations, and security, aims to embed security practices throughout the software development

lifecycle. It emphasizes the collaboration and communication between development, operations, and security teams, enabling a proactive approach to security rather than a reactive one. By integrating security early and continuously into the development pipeline, organizations can identify and address security vulnerabilities at each stage, reducing the likelihood of security breaches and minimizing their impact [2].

In the context of Kubernetes, DevSecOps plays a crucial role in securing deployments and protecting containerized applications [3]. It involves implementing security measures as code, automating security processes, and incorporating security-focused metrics collection and analysis. DevSecOps helps organizations detect and prevent attacks, ensure compliance with industry regulations, and establish a culture of security awareness and accountability.

This article explores the growing adoption of Kubernetes, the increasing importance of security in Kubernetes environments, and the role of DevSecOps in securing Kubernetes deployments. By emphasizing the need for a proactive security approach and the integration of security practices throughout the development lifecycle, organizations can enhance the overall security posture of their Kubernetes deployments and safeguard their containerized applications against emerging threats.

1. Related work

The related work in the field of DevSecOps and Kubernetes security research spans a wide range of topics, including cultural adoption, practical implementation, emerging trends, and challenges faced during DevSecOps adoption. The following studies provide valuable insights for researchers, practitioners, and organizations seeking to enhance the security posture of their Kubernetes deployments. By leveraging the findings from existing research, researchers can contribute to the advancement of secure and resilient security solutions that safeguard Kubernetes environments and ensure the overall success of DevSecOps practices.

Mary and Ricardo in “A Systematic Literature-Review of DevSecOps” [4] provide a comprehensive analysis of the cultural aspects of DevSecOps adoption. The study highlights the significance of fostering a security-first mindset within organizations and promoting a collaborative and shared responsibility for security among development, operations, and security teams. By synthesizing findings from a wide range of sources, the article emphasizes the role of culture in successful DevSecOps implementation and advocates for continuous learning and knowledge-sharing to build a security-aware organizational culture.

On the practical implementation side, Rahul B. S [5] developing and updating of the latest security methods that tend to give adequate, proficient and productive results in a secured manner. Therefore, to demonstrate the DevSecOps we intend to use the open source tools which can be freely downloaded and used to demonstrate the data security. In order to run the frequent process of integrating and testing for the security of data, we require more resource and time respectively. To overcome this problem here we implementing and integrating security as part of the pipeline process and hence can achieve faster response time and fend off attacks beforehand and accordingly create a secure environment and more protected system.”,”author”:[{“dropping-particle”:"",“family”:”Rahul”,“given”:”S”,“non-dropping-particle”:"",“parse-names”:false,”suffix”:""},{“container-title”:”International Journal of Advance Research”,“id”:"ITEM-1”,“issued”:{“-date-parts”:[["2019"]]},“title”:”Implementation of DevSecOps using Open-Source tools”,“type”:"article-journal"},“uris”:[“http://www.mendeley.com/documents/?uuid=6b3df69e-c99a-3f14-9ff8-06c0e1d95062”]},“mendeley”:{“formattedCitation”:[5],“plain-TextFormattedCitation”:[5],“previouslyFormattedCitation”:[5],“properties”:{“noteIndex”:0},“schema”:"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"} offers valuable insights into integrating security practices within DevOps pipelines, with a specific focus on utilizing open-source solutions in Kubernetes environments. The research presents case studies of organizations that have effectively adopted DevSecOps and leveraged open-source tools and technologies to enhance security. This study serves as a practical guide for organizations seeking to embrace DevSecOps while capitalizing on open-source security solutions, providing real-world examples of successful implementation strategies.

Additionally, Runfeng Mao [6] provides a unique perspective by gathering insights from diverse sources such as blogs, whitepapers, and other non-peer-reviewed publications. By exploring emerging trends and challenges in DevSecOps adoption from a variety of sources, this work offers preliminary understandings of the current state of DevSecOps practices. The findings from this research can inform the identification of emerging themes and areas of interest in the field.

Furthermore, “DevSecOps: A Multivocal Literature Review” by Havard Myrbakken and Ricardo Colombo-Palacios [7]. synthesizes multiple perspectives from academic and practitioner sources to present a comprehensive view of DevSecOps practices and their impact on security. This multidimensional review highlights the significance of fostering a collaborative DevSecOps culture across organizations and emphasizes the role of

communication and coordination among development, operations, and security teams. The article underscores the value of integrating security practices throughout the software development lifecycle and emphasizes the need for continuous improvement and adaptation in dynamic and evolving environments.

Finally, “Challenges and Solutions When Adopting DevSecOps: A Systematic Review” by Roshan N. Rajapakse delves into the barriers faced by organizations in adopting DevSecOps practices [8] this trend has presented the challenge of ensuring secure software delivery while maintaining the agility of DevOps. The efforts to integrate security in DevOps have resulted in the DevSecOps paradigm, which is gaining significant interest from both industry and academia. However, the adoption of DevSecOps in practice is proving to be a challenge. Objective: This study aims to systemize the knowledge about the challenges faced by practitioners when adopting DevSecOps and the proposed solutions reported in the literature. We also aim to identify the areas that need further research in the future. Method: We conducted a Systematic Literature Review of 54 peer-reviewed studies. The thematic analysis method was applied to analyze the extracted data. Results: We identified 21 challenges related to adopting DevSecOps, 31 specific solutions, and the mapping between these findings. We also determined key gap areas in this domain by holistically evaluating the available solutions against the challenges. The results of the study were classified into four themes: People, Practices, Tools, and Infrastructure. Our findings demonstrate that tool-related challenges and solutions were the most frequently reported, driven by the need for automation in this paradigm. Shift-left security and continuous security assessment were two key practices recommended for DevSecOps. People-related factors were considered critical for successful DevSecOps adoption but less studied. Conclusions: We highlight the need for developer-centered application security testing tools that target the continuous practices in DevSecOps. More research is needed on how the traditionally manual security practices can be automated to suit rapid software deployment cycles. Finally, achieving a suitable balance between the speed of delivery and security is a significant issue practitioners face in the DevSecOps paradigm.”, “author”: [{“dropping-particle”: “”, “family”: “Rajapakse”, “given”: “Roshan N.”, “non-dropping-particle”: “”, “parse-names”: false, “suffix”: “”}], {“dropping-particle”: “”, “family”: “Zahedi”, “given”: “Mansoorah”, “non-dropping-particle”: “”, “parse-names”: false, “suffix”: “”}], {“dropping-particle”: “”, “family”: “Babar”, “given”: “M. Ali”, “non-dropping-particle”: “”, “parse-names”: false, “suffix”: “”}], {“dropping-particle”: “”, “family”: “Shen”, “given”: “Haifeng”, “non-dropping-particle”: “”, “parse-names”: false, “suffix”: “”}], “container-title”: “Information and Software Technology”, “id”: “ITEM-1”, “issued”: {“date-parts”: [[“2022”, “1”, “1”]]}, “page”: “106700”, “publisher”: “Elsevier”, “title”: “Challenges and solutions when adopting DevSecOps: A systematic review”, “type”: “article-journal”, “volume”: “141”, “uris”: [“http://www.mendeley.com/documents/?uuid=71d090d2-d3bb-3a23-a6c9-98c2e0eadaf4”]], “mendeley”: {“formattedCitation”: “[8]”, “plainTextFormattedCitation”: “[8]”, “previouslyFormattedCitation”: “[8]”, “properties”: {“noteIndex”: 0}, “schema”: “https://github.com/citation-style-language/schema/raw/master/csl-citation.json”}. The study explores the challenges and limitations that may hinder successful integration of security in the DevOps workflow. By identifying common challenges, the research provides valuable recommendations and mitigation strategies to overcome these obstacles effectively, facilitating the adoption of DevSecOps and enhancing security practices in Kubernetes environments.

2. The Need for DevSecOps in Kubernetes Security Research

Traditional security approaches often struggle to keep pace with the dynamic and distributed nature of Kubernetes environments. Several challenges arise when attempting to secure Kubernetes deployments using conventional security methods:

- **Complexity and Scale:** Kubernetes environments are highly complex, comprising numerous interconnected components, including pods, nodes, clusters, and the control plane [9] we see that most of the cloud-based applications and services often consist of hundreds of micro-services; however, the traditional monolithic pattern is no longer suitable for today’s development life-cycle. This is due to the difficulties of maintenance, scale, load balance, and many other factors associated with it. Consequently, people switch their focus on containerization—a lightweight virtualization technology. The saving grace is that it can use machine resources more efficiently than the virtual machine (VM). Traditional security tools and practices may struggle to adequately address the scale and intricacies of these environments.
- **Rapid Deployment and Continuous Integration/Continuous Deployment (CI/CD):** Kubernetes promotes rapid application deployment through CI/CD pipelines, where new code changes are frequently built, tested, and deployed. Traditional security measures, often applied as an afterthought,

may not be able to keep up with the speed of these deployments [10] the number and variety of business requirements are increasingly complex, keeps sustained growth, the process of continuous integration delivery of information systems becomes increasingly complex, the amount of repetitive work is growing. This paper focuses on the continuous integration of specific information systems, a collaborative work scheme for continuous integrated delivery based on Jenkins and Ansible is proposed. Both theory and practice show that continuous integrated delivery cooperative systems can effectively improve the efficiency and quality of continuous integrated delivery of information systems. The effect of the optimization and upgrading of the information system is obvious.”, “author”: [{ “dropping-particle”: “”, “family”: “Petrochina”, “given”: “Wang Yiran”, “non-dropping-particle”: “”, “parse-names”: false, “suffix”: “” }, { “dropping-particle”: “”, “family”: “Petrochina”, “given”: “Zhang Tongyang”, “non-dropping-particle”: “”, “parse-names”: false, “suffix”: “” }, { “dropping-particle”: “”, “family”: “Petrochina”, “given”: “Guo Yidong”, “non-dropping-particle”: “”, “parse-names”: false, “suffix”: “” }], “container-title”: “2018 International Conference on Artificial Intelligence and Big Data, ICAIBD 2018”, “id”: “ITEM-1”, “issued”: { “date-parts”: [[“2018”, “6”, “25”]] }, “page”: “245-249”, “publisher”: “Institute of Electrical and Electronics Engineers Inc.”, “title”: “Design and implementation of continuous integration scheme based on Jenkins and Ansible”, “type”: “article-journal” }, “uris”: [“http://www.mendeley.com/documents/?uuid=2ab94d35-3d8e-3139-83c4-d618e514cbc3”] }, “mendeley”: { “formattedCitation”: “[10]”, “plainTextFormattedCitation”: “[10]”, “previouslyFormattedCitation”: “[10]”, “properties”: { “noteIndex”: 0 }, “schema”: “https://github.com/citation-style-language/schema/raw/master/csl-citation.json” } }.

- **Container Lifecycle Management:** Containers, the building blocks of Kubernetes, have unique security challenges. Ensuring the integrity and security of container images, scanning for vulnerabilities, and enforcing secure configurations across a large number of containers require specialized security considerations that may be lacking in traditional approaches.
- **Dynamic Workloads and Orchestration:** Kubernetes is designed to handle dynamic workloads, including scaling pods and allocating resources based on demand. Traditional security approaches may struggle to adapt to the continuously changing environment and may not effectively protect against attacks targeting the orchestration layer.

Integrating security early in the software development lifecycle is crucial to mitigating risks and addressing vulnerabilities effectively as shown in Figure 1. By incorporating security practices from the early stages of development, organizations can benefit in several ways:

- **Shift-Left Security:** Integrating security early in the development process, often referred to as “shifting left,” enables proactive identification and remediation of security vulnerabilities. By conducting security assessments, code reviews, and threat modeling during the development phase, potential risks can be identified and mitigated before they manifest in production environments [11].
- **Cost and Time Efficiency:** Addressing security issues early in the development lifecycle helps reduce the cost and effort associated with fixing vulnerabilities at later stages. Fixing security flaws during development is typically less time-consuming and less disruptive compared to addressing them in production environments.
- **Security by Design:** Incorporating security considerations throughout the development process allows for security to be built into the application architecture and design. By adopting secure coding practices, secure configuration management, and secure deployment strategies, organizations can create a more resilient and secure Kubernetes environment.

DevSecOps offers a comprehensive approach to addressing security gaps and vulnerabilities in Kubernetes deployments. By integrating security practices throughout the software development lifecycle, organizations can effectively tackle the unique security challenges posed by Kubernetes:

- **Automated Security Testing:** DevSecOps promotes the use of automated security testing tools and techniques, such as vulnerability scanning, static code analysis, and penetration testing. These practices help identify security weaknesses and vulnerabilities specific to Kubernetes configurations, container images, and deployed applications.
- **Continuous Monitoring and Incident Response:** DevSecOps emphasizes continuous monitoring of Kubernetes environments, including logging, auditing, and anomaly detection. By implementing robust monitoring solutions, organizations can detect and respond to security incidents promptly, minimizing the impact of potential attacks [1].
- **Infrastructure as Code (IaC):** DevSecOps encourages the use of Infrastructure as Code principles in Kubernetes deployments. By defining infrastructure and security configurations as code, organizations

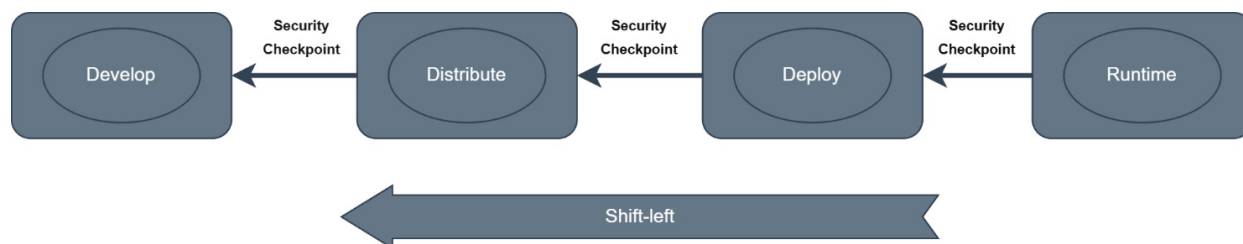


Fig. 1. Security integration in software development life cycle

can enforce consistent and secure deployment practices, reducing the risk of misconfigurations and ensuring compliance with security best practices.

- **Collaboration and Communication:** DevSecOps promotes cross-functional collaboration between development, operations, and security teams. This collaboration enables early identification and resolution of security issues, fosters a shared responsibility for security, and facilitates knowledge sharing to stay updated with the latest security threats and mitigation techniques.

By leveraging DevSecOps practices in Kubernetes security research, organizations can address security gaps, proactively mitigate vulnerabilities, and build a secure foundation for their containerized applications. This approach ensures that security is not an afterthought but an integral part of the entire software development lifecycle, enabling organizations to achieve robust and resilient Kubernetes deployments.

3. Key Principles of DevSecOps in Kubernetes Security

DevSecOps in the context of Kubernetes security research is guided by several key principles that help organizations effectively integrate security practices throughout the development lifecycle. These principles ensure collaboration, automation, and the seamless integration of security measures within Kubernetes deployments [12].

– Collaboration and Communication between Development, Operations, and Security Teams:

Foster a culture of collaboration and shared responsibility between development, operations, and security teams. This involves breaking down silos and promoting open communication channels for sharing security knowledge, insights, and best practices.

Encourage regular meetings and cross-functional discussions to align security requirements, understand operational constraints, and prioritize security considerations within the Kubernetes environment.

– Automation of Security Processes and Continuous Monitoring:

Leverage automation tools and techniques to streamline security processes within Kubernetes. This includes automating security testing, vulnerability scanning, and compliance checks to identify and address security issues in an efficient and consistent manner [13].

Implement continuous monitoring solutions that capture security-related events and activities within the Kubernetes environment. This enables real-time visibility into potential security threats, rapid incident response, and proactive mitigation of risks.

– Integration of Security Practices throughout the Entire Development Pipeline:

Integrate security considerations at each stage of the development pipeline, from initial code development to deployment and ongoing operations. This involves incorporating security checks, code reviews, and security testing into the CI/CD process. See Figure 2.

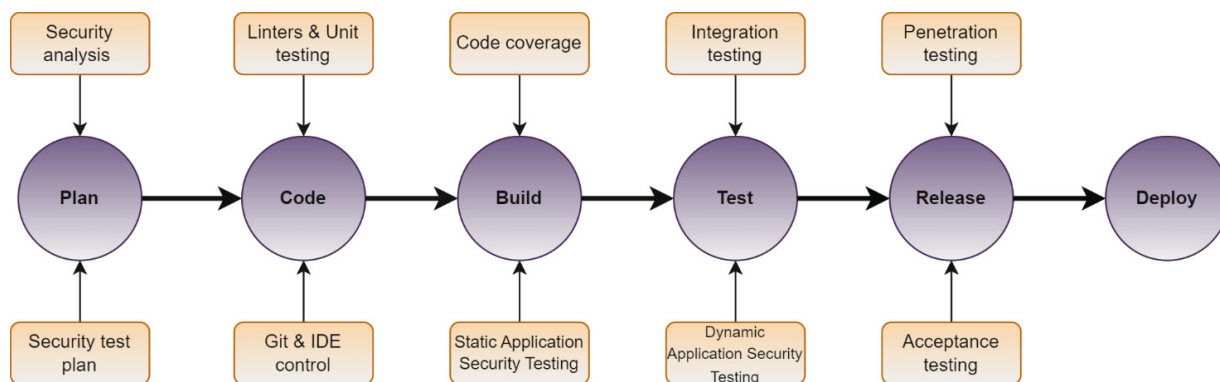


Fig. 2. DevSecOps pipeline

Implement security gates and controls that ensure the adherence to security policies and best practices at every step of the deployment and release process. This ensures that security is not an isolated activity but an integral part of the overall development workflow.

– **Implementing Security as Code to Ensure Consistency and Scalability:**

Treat security configurations and practices as code artifacts, enabling consistency, repeatability, and scalability. Define security policies, configurations, and deployment practices as code, leveraging tools like Infrastructure as Code (IaC) and configuration management tools.

Implement security-focused pipelines and version control systems to track and manage changes to security configurations and ensure their traceability and audibility.

By adhering to these key principles, organizations can establish a robust DevSecOps approach within their Kubernetes security research. This approach promotes collaboration, automates security processes, integrates security throughout the development pipeline, and treats security as a foundational element that ensures the consistency, scalability, and resilience of Kubernetes deployments.

4. Benefits of Adopting DevSecOps in Kubernetes Security Research

Early Detection and Prevention of Security Vulnerabilities and Attacks:

- By integrating security practices throughout the development lifecycle, DevSecOps enables early identification and remediation of security vulnerabilities specific to Kubernetes environments.
- Security testing, vulnerability scanning, and code reviews conducted during the development phase help identify and address security weaknesses before they are deployed into production.
- Continuous monitoring and automated security checks provide real-time visibility into potential security threats, allowing for proactive detection and prevention of attacks.

Increased Visibility and Traceability of Security-Related Events in Kubernetes:

- DevSecOps practices emphasize the implementation of robust monitoring and logging mechanisms within Kubernetes environments.
- Continuous monitoring provides detailed insights into security-related events, allowing for comprehensive visibility into the state of the Kubernetes infrastructure, containerized applications, and user activities.

- Enhanced traceability of security events enables efficient incident response, forensic analysis, and compliance auditing.

Improved Incident Response and Mitigation Capabilities:

- DevSecOps promotes a proactive approach to incident response by integrating security monitoring, automated alerting, and incident management processes within the Kubernetes environment.
- Real-time monitoring and automated incident response mechanisms enable swift detection and mitigation of security incidents, reducing the impact of potential breaches.
- Collaboration between development, operations, and security teams facilitates faster incident response, enabling coordinated efforts to resolve security issues effectively.

By adopting DevSecOps in Kubernetes security research, organizations can leverage these benefits to enhance the overall security posture of their Kubernetes deployments. Early detection and prevention of security vulnerabilities, increased visibility and traceability of security events, improved incident response capabilities, risk reduction, and compliance facilitation collectively contribute to creating a more secure and resilient Kubernetes environment.

5. Implementing DevSecOps in the Research Framework

To effectively implement DevSecOps principles in the research framework focused on Kubernetes security and attack detection, the following steps can be followed:

Integrating Security Practices within the Attack Detection Model:

- Incorporate security considerations directly into the design and implementation of the attack detection model.
- Define security-specific features and indicators that can help identify potential attacks or anomalies in the Kubernetes environment.
- Ensure that the model is trained and tested using datasets that include various attack scenarios and security-relevant patterns.

Incorporating Security-Focused Metrics Collection and Analysis:

- Define and collect security-focused metrics related to the Kubernetes environment, such as resource utilization, network traffic, container behavior, and API usage.
- Analyze these metrics to detect any abnormal patterns or deviations from expected behavior, which may indicate security threats or vulnerabilities.
- Leverage data visualization techniques to provide meaningful insights into the security posture of

the Kubernetes environment and facilitate decision-making.

Leveraging Automated Security Testing and Vulnerability Scanning Tools:

- Integrate automated security testing tools and vulnerability scanners into the research framework to identify potential weaknesses and vulnerabilities in the Kubernetes deployment.
- Use these tools to regularly scan container images, Kubernetes configurations, and the underlying infrastructure for known security issues.
- Incorporate the findings from these security tests into the research pipeline to assess the impact on the attack detection model and identify areas for improvement.

Collaboration and Communication:

- Foster collaboration between the research team, security experts, and operations personnel to ensure a holistic approach to Kubernetes security.
- Engage in regular discussions and knowledge-sharing sessions to align research objectives with security goals and best practices.
- Encourage open communication channels to promptly address security concerns, share findings, and implement security enhancements within the research framework.

By implementing these DevSecOps practices, the research framework can benefit from the integration of security into the attack detection model, the inclusion of security-focused metrics, the use of automated security testing tools, and the establishment of a feedback loop for continuous improvement. This approach enables a proactive and comprehensive approach to Kubernetes security in the research process and contributes to the development of effective security solutions for Kubernetes environments.

6. Best Practices for DevSecOps in Kubernetes Security Research

By following the best practices, organizations engaged in Kubernetes security research can enhance the effectiveness and efficiency of their DevSecOps approach. The following practices facilitate the identification and mitigation of security risks, ensure secure configuration and access controls, enable continuous monitoring and analysis, and promote regular security assessments and audits to maintain a robust security posture in Kubernetes deployments [14].

Conducting Threat Modeling and Risk Assessments Specific to Kubernetes:

- Perform threat modeling exercises to identify potential security risks and vulnerabilities specific to Kubernetes deployments. Consider factors such

as container security, network security, and access controls.

- Conduct regular risk assessments to evaluate the impact and likelihood of identified threats, prioritize them based on risk levels, and implement appropriate security measures.

Ensuring Secure Configuration and Hardening of Kubernetes Clusters:

- Implement secure configuration practices for Kubernetes clusters, including proper network segmentation, secure API server settings, and restricted access to sensitive resources.
- Follow Kubernetes hardening guides and security best practices to eliminate common security misconfigurations and reduce attack surfaces.
- Regularly review and update cluster configurations to address emerging security concerns and align with industry standards.

Implementing Identity and Access Management Controls:

- Establish strong identity and access management controls for Kubernetes environments.
- Implement robust authentication mechanisms, such as multi-factor authentication (MFA), and enforce the principle of least privilege to limit access to sensitive Kubernetes resources.
- Implement Role-Based Access Control (RBAC) policies to ensure that only authorized users have appropriate permissions to interact with Kubernetes resources [15].

Continuous Monitoring and Log Analysis for Security Events and Anomalies:

- Implement a comprehensive monitoring solution that captures and analyzes security events, logs, and metrics from Kubernetes clusters.
- Continuously monitor for suspicious activities, unauthorized access attempts, and abnormal behavior within the Kubernetes environment.
- Utilize log analysis and anomaly detection techniques to identify potential security incidents or indicators of compromise.

Regular Security Assessments and Audits:

- Conduct regular security assessments and penetration testing to evaluate the effectiveness of security controls within Kubernetes deployments.
- Perform vulnerability scanning and security testing of container images and Kubernetes configurations.
- Conduct periodic security audits to validate compliance with security policies, industry regulations, and best practices.

7. Challenges and Mitigation Strategies

Implementing DevSecOps in the research framework for Kubernetes security may encounter certain

challenges. However, with appropriate mitigation strategies, these challenges can be overcome effectively:

Overcoming Resistance to Change and Fostering a Security Culture:

Challenge: Resistance to change and lack of awareness or understanding of the importance of security practices can hinder the adoption of DevSecOps.

Mitigation: Educate and raise awareness among team members about the significance of security in Kubernetes deployments. Promote a security-first mindset and foster a culture of shared responsibility for security [16] in particular of security standards. A comprehensive overview of this scattered field is still missing and we know little about how to achieve security compliance in agile software development. Existing secondary studies (mapping studies and literature reviews. Provide training and resources to help team members understand the value and benefits of DevSecOps.

Balancing Security and Operational Requirements in Kubernetes Deployments:

Challenge: Striking a balance between implementing robust security measures and maintaining operational efficiency can be challenging.

Mitigation: Collaborate closely with operations teams to understand their requirements and constraints. Conduct risk assessments to identify critical security needs and prioritize them based on operational impact. Continuously monitor and refine security measures to optimize the balance between security and operational requirements.

Ensuring Compatibility of DevSecOps Practices with Research Objectives:

Challenge: Research objectives may have specific requirements that need to be aligned with DevSecOps practices.

Mitigation: Analyze the research objectives and identify areas where DevSecOps practices can be seamlessly integrated. Adapt DevSecOps processes to accommodate the specific needs of the research while ensuring that security considerations are not compromised. Seek guidance from domain experts and security practitioners to find the right balance.

Addressing Resource Constraints and Scalability Concerns:

Challenge: Limited resources, such as budget, personnel, or infrastructure, can pose challenges to implementing comprehensive DevSecOps practices.

Mitigation: Prioritize security activities based on risk assessments and available resources. Automate security processes, such as testing and monitoring, to reduce the manual effort required. Leverage cloud-based security services or consider outsourcing certain security tasks to optimize resource utilization. Focus

on scalability by designing the research framework in a modular and extensible manner.

It is important to note that challenges may vary based on the specific context and organization. Regularly assess the effectiveness of the mitigation strategies and adjust them as needed to overcome emerging challenges. By addressing these challenges proactively, the research framework can benefit from the enhanced security provided by DevSecOps practices, enabling more reliable and secure Kubernetes deployments.

Conclusion

In conclusion, DevSecOps plays a crucial role in ensuring the security and resilience of Kubernetes environments. This article has emphasized the importance of integrating security practices early in the software development lifecycle and highlighted the benefits and best practices of adopting DevSecOps in Kubernetes security research.

The adoption of DevSecOps brings several significant advantages. It enables early detection and prevention of security vulnerabilities and attacks, increases visibility and traceability of security events, improves incident response capabilities, reduces security risks and associated costs, and facilitates compliance with industry regulations and standards. By incorporating collaboration, automation, and security as code principles, organizations can establish a robust security posture in their Kubernetes deployments.

To fully embrace DevSecOps, it is essential for organizations and researchers to overcome challenges such as resistance to change, balancing security and operational requirements, ensuring compatibility with research objectives, and addressing resource constraints. By fostering a security culture, aligning security practices with operational needs, adapting DevSecOps to research goals, and optimizing resource utilization, these challenges can be effectively mitigated.

Looking ahead, the future of Kubernetes security research lies in exploring further advancements in DevSecOps practices. Areas for further research and development include the refinement of machine learning-based attack detection models, the integration of advanced threat intelligence and anomaly detection techniques, and the continuous evolution of automated security testing tools. Additionally, researching ways to enhance the scalability and resource efficiency of DevSecOps in Kubernetes environments will be crucial as organizations increasingly adopt container orchestration platforms.

In conclusion, embracing DevSecOps is vital to enhance the security and resilience of Kubernetes envi-

ronments. By implementing the discussed benefits and best practices, organizations can effectively protect their Kubernetes deployments, mitigate security risks, and drive innovation in the field of Kubernetes security research. By continuously improving and expanding DevSecOps practices, researchers can contribute to the advancement of secure and reliable Kubernetes ecosystems for the benefit of the entire industry.

References

1. Darwesh G., Hammoud J. and Vorobeva A.A. "A novel approach to feature collection for anomaly detection in Kubernetes environment and agent for metrics collection from Kubernetes nodes," *Sci. Tech. J. Inf. Technol. Mech. Opt.*, vol. 23, no. 3, pp. 538–546, Jun. 2023, doi: 10.17586/2226-1494-2023-23-3-538-546.
2. Gomes K. "The Importance of DevSecOps," Honor. Capstones, May 2018, Accessed: Jul. 23, 2023. [Online]. Available: <https://huskiecommons.lib.niu.edu/studentengagement-honorscapstones/1214>
3. Prates L., Faustino J., Silva M. and Pereira R. "DevSecOps metrics," *Lect. Notes Bus. Inf. Process.*, vol. 359, pp. 77–90, 2019, doi: 10.1007/978-3-030-29608-7_7/COVER.
4. Sánchez-Gordón M. and Colomo-Palacios R. "Security as Culture: A Systematic Literature Review of DevSecOps," *Proc. - 2020 IEEE/ACM 42nd Int. Conf. Softw. Eng. Work. ICSEW 2020*, pp. 266–269, Jun. 2020, doi: 10.1145/3387940.3392233.
5. Rahul S. "Implementation of DevSecOps using Open-Source tools," *Int. J. Adv. Res.*, 2019, Accessed: Jul. 22, 2023. [Online]. Available: www.IJARIIT.com
6. Mao R. et al. "Preliminary Findings about DevSecOps from Grey Literature," *Proc. - 2020 IEEE 20th Int. Conf. Softw. Qual. Reliab. Secur. QRS 2020*, pp. 450–457, Dec. 2020, doi: 10.1109/QRS51102.2020.00064.
7. Myrbakken H. and Colomo-Palacios R. "DevSecOps: A multivocal literature review," *Commun. Comput. Inf. Sci.*, vol. 770, pp. 17–29, 2017, doi: 10.1007/978-3-319-67383-7_2/COVER.
8. Rajapakse R.N., Zahedi M., Babar M.A. and Shen H. "Challenges and solutions when adopting DevSecOps: A systematic review," *Inf. Softw. Technol.*, vol. 141, p. 106700, Jan. 2022, doi: 10.1016/J.INF-SOF.2021.106700.
9. Mondal S.K., Pan R., Kabir H.M.D., Tian T. and Dai H.N. "Kubernetes in IT administration and serverless computing: An empirical study and research challenges," *J. Supercomput.*, vol. 78, no. 2, pp. 2937–2987, Feb. 2022, doi: 10.1007/s11227-021-03982-3.
10. Petrochina W.Y., Petrochina Z.T. and Petrochina G.Y. "Design and implementation of continuous integration scheme based on Jenkins and Ansible," *2018 Int. Conf. Artif. Intell. Big Data, ICA-IBD 2018*, pp. 245–249, Jun. 2018, doi: 10.1109/ICAIBD.2018.8396203.
11. Lombardi F. and Fanton A. "From DevOps to DevSecOps is not enough. CyberDevOps: an extreme shifting-left architecture to bring cybersecurity within software security lifecycle pipeline," *Softw. Qual. J.*, vol. 31, no. 2, pp. 619–654, Jun. 2023, doi: 10.1007/S11219-023-09619-3/METRICS.
12. Mohan V. and Ben Othmane L. "SecDevOps: Is it a marketing buzzword? Mapping research on security in DevOps," *Proc. - 2016 11th Int. Conf. Availability, Reliab. Secur. ARES 2016*, pp. 542–547, Dec. 2016, doi: 10.1109/ARES.2016.92.
13. Mburano B. and Si W. "Evaluation of web vulnerability scanners based on OWASP benchmark," *26th Int. Conf. Syst. Eng. ICSEng 2018 - Proc.*, Feb. 2019, doi: 10.1109/ICSENG.2018.8638176.
14. Darwesh G., Hammoud J. and Vorobeva A.A. "SECURITY IN KUBERNETES: BEST PRACTICES AND SECURITY ANALYSIS," *J. Ural Fed. Dist. Inf. Secur.*, vol. 22, no. 2, 2022, doi: 10.14529/SECUR220209.
15. Shamim S.I. "Mitigating security attacks in kubernetes manifests for security best practices violation," in *ESEC/FSE 2021 - Proceedings of the 29th ACM Joint Meeting European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, Association for Computing Machinery, Inc, Aug. 2021, pp. 1689–1690. doi: 10.1145/3468264.3473495.
16. Moyon F., Almeida P., Riofrio D., Mendez D. and Kalinowski M. "Security Compliance in Agile Software Development: A Systematic Mapping Study," *Proc. - 46th Euromicro Conf. Softw. Eng. Adv. Appl. SEAA 2020*, pp. 413–420, Aug. 2020, doi: 10.1109/SEAA51224.2020.00073.

Ghadeer Darwesh. ITMO University, Saint Petersburg, Russian Federation, PhD Student, <https://orcid.org/0000-0003-1116-9410>, ghadeerdarwesh32@gmail.com

Jaafar Hammoud. ITMO University, Saint Petersburg, Russian Federation, PhD, <https://orcid.org/0000-0002-2033-0838>, hammoudgj@gmail.com

Vorobeva Alisa A. ITMO University, Saint Petersburg, 197101, Russian Federation, PhD, Associate Professor, <https://orcid.org/0000-0001-6691-6167>, alice_w@mail.ru

Повышение безопасности Kubernetes: решающая роль DevSecOps

Гадир Дарвиш, Джафар Хаммуд, А.А. Воробьева

Университет ИТМО, г. Санкт-Петербург, Россия

Аннотация. В статье подчеркивается важность интеграции практик DevSecOps (разработка, безопасность и эксплуатация) в исследования по обнаружению распространенных атак в средах Kubernetes. По мере того, как Kubernetes быстро становится популярной платформой оркестровки контейнеров, проблемы безопасности, связанные с контейнерными приложениями, значительно возросли. Однако традиционные методологии безопасности часто с трудом успевают за динамичной и быстро развивающейся природой контейнерных сред, оставляя потенциальные уязвимости для использования злоумышленниками. Подчеркивая важность DevSecOps, статья призвана подчеркнуть его роль в повышении уровня безопасности развертываний Kubernetes и продвижении упреждающего подхода к защите контейнерных приложений. В статье также обсуждаются ключевые аспекты и преимущества внедрения DevSecOps в контексте исследований безопасности Kubernetes.

Ключевые слова: *DevSecOps, безопасность Kubernetes, DevOps, методы обеспечения безопасности.*

DOI: 10.14357/20790279240309 **EDN:** SDRKHO

Литература

1. Дарвиш Г., Хаммуд Дж., Воробьева А.А. Новый подход к сбору признаков для обнаружения аномалий в среде Kubernetes и агент для сбора метрик с узлов Kubernetes. Тех. Ж. Инф. Технол. Мех. Опт. 2023. Вып. 23. С. 538–546. DOI: 10.17586/2226-1494-2023-23-3-538-546.
2. Гомес К. «Важность DevSecOps», Честь. Capstones, май 2018 г., доступ: 23 июля 2023 г. [Онлайн]. Доступно: <https://huskiecommons.lib.niu.edu/studentengagement-honorscapstones/1214>.
3. Пратес Л., Фаустино Дж., Сильва М. и Перейра Р. «Метрики DevSecOps», Лект. Примечания Автобус. Инф. Процесс. 2019. Вып. 359. С. 77–90. DOI: 10.1007/978-3-030-29608-7_7/COVER.
4. Санчес-Гордон М. и Р. Коломо-Паласиос М. «Безопасность как культура: систематический обзор литературы по DevSecOps», Proc. - 42-я Международная конференция IEEE/ACM 2020 г. Конф. Программное обеспечение англ. Работа. ICSEW. 2020. С. 266–269. DOI: 10.1145/3387940.3392233.
5. Рахул С. «Внедрение DevSecOps с использованием инструментов с открытым исходным кодом», Int. Дж. Адв. Res., 2019, доступ: 22 июля 2023 г. [Онлайн]. Доступно: www.IJARIT.com.
6. Мао Р. и др. «Предварительные выводы о DevSecOps из серой литературы», Proc. - 20-й Международный IEEE 2020 г. Конф. Программное обеспечение Квал. Надежный. Безопасность. QRS. 2020. С. 450–457. DOI: 10.1109/QRS51102.2020.00064.
7. Мирбаккен Х. и Коломо-Паласиос Р. «DevSecOps: многоголосый обзор литературы», Commun. Вычислитель. Инф. наук. 2017. Т. 770. С. 17–29. DOI: 10.1007/978-3-319-67383-7_2/COVER.
8. Раджапаксе Р.Н., Захеда М., Бабар М.А. и Шен Х. «Проблемы и решения при внедрении DevSecOps: систематический обзор», Инф. Программное обеспечение Технол. 2022. Вып. 141. С. 106700. DOI: 10.1016/J.INFSOF.2021.106700.
9. Мондал С.К., Пан Р., Кабир Х.М.Д., Туан Т. и Дай Х.Н. «Kubernetes в ИТ-администрировании и бессерверных вычислениях: эмпирическое исследование и исследовательские задачи», J. Supercomput. 2022. Вып. 78, № 2. С. 2937–2987. DOI: 10.1007/s11227-021-03982-3.
10. Петрокина В.Ю., Петрокина З.Т. и Петричина Г.Ю. «Проектирование и реализация схемы непрерывной интеграции на основе Jenkins и Ansible», 2018 Int. Конф. Артиф. Интел. Боль-

- шие данные, ICAIBD. 2018. С. 245–249. DOI: 10.1109/ICAIBD.2018.8396203.
11. Ломбарди Ф. и Фэнтон А. «От DevOps к DevSecOps недостаточно. CyberDevOps: крайне сдвинутая влево архитектура, обеспечивающая кибербезопасность в рамках жизненного цикла безопасности программного обеспечения», Softw. Квал. Дж. 2023. Вып. 31. № 2. С. 619–654. DOI: 10.1007/S11219-023-09619-3/METRICS.
12. Мохан В. и Отман Л. Бен. «SecDevOps: это модное маркетинговое словечко? Картирование исследований по безопасности в DevOps», Учеб. - 2016 11-й Международный. Конф. Доступность, Надежность. Безопасность. ARES. 2016. С. 542–547. DOI: 10.1109/ARES.2016.92.
13. Мбурано Б. и Си В. «Оценка сканеров веб-уязвимостей на основе эталонного теста OWASP», 26-й международный конгресс. Конф. Сист. англ. ICSEng. 2018. Proc. DOI: 10.1109/ICSENG.2018.8638176.
14. Дарвиш Дж., Хаммуд Дж., Воробьева А.А. Безопасность в kubernetes: лучшие практики и анализ безопасности // Журнал Уральской федерации. Расст. Инф. Безопасность. 2022. Т. 22, № 2. DOI: 10.14529/SECUR220209.
15. Шамим С.И. «Снижение атак на безопасность в манифестах Kubernetes для нарушения лучших практик безопасности», в ESEC / FSE 2021 – Материалы 29-го совместного заседания ACM, Европейской конференции по разработке программного обеспечения и симпозиума по основам программной инженерии, Association for Computing Machinery, Inc. 2021. С. 1689–1690. DOI: 10.1145/3468264.3473495.
16. Мойон Ф., Алмейда П., Риоффрио Д., Мендес Д. и Калиновски М. «Соответствие требованиям безопасности при гибкой разработке программного обеспечения: систематическое картографическое исследование», Proc. - 46-я Евромикро Конференция. Программное обеспечение англ. Адв. Прил. SEAA. 2020. С. 413–420. DOI: 10.1109/SEAA51224.2020.00073.

Гадир Дарвиш. Университет ИТМО, г. Санкт-Петербург, Россия. Аспирант.

E-mail: ghadeerdarwesh32@gmail.com

Джафар Хаммуд. Университет ИТМО, г. Санкт-Петербург, Россия. Кандидат наук. <https://orcid.org/0000-0002-2033-0838>. E-mail: hammoudgj@gmail.com

Воробьева Алиса Андреевна. Университет ИТМО, г. Санкт-Петербург, Россия. Доцент. <https://orcid.org/0000-0001-6691-6167>. E-mail: alice_w@mail.ru

Поиск уязвимостей в смарт-контрактах на основе машинного обучения

В.С. БЕЛОУС^I, И.А. ТАРХАНОВ^{II,III}

^I Национальный исследовательский технологический университет «МИСиС», г. Москва, Россия

^{II} Федеральное государственное бюджетное учреждение высшего образования «Государственный академический университет гуманитарных наук», г. Москва, Россия

^{III} Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия

Аннотация. С ростом популярности блокчейн-проектов растет и количество децентрализованных приложений на их основе. Центральным звеном этих приложений являются смарт-контракты. Эта технология еще относительно молода и имеет ряд известных проблем с безопасностью. Статистика взлома смарт-контрактов свидетельствует об актуальности проблемы поиска уязвимостей в коде смарт-контрактов. В статье описано 3 модели машинного обучения для поиска уязвимостей в смарт-контрактах, написанных на языке Solidity. Особое внимание уделяется подготовке датасета для обучения и сравнения с известными анализаторами кода. Полученные по результатам обучения и апробации моделей метрики позволяют утверждать, что модель, состоящая из трех двунаправленных рекуррентных слоев BiGRU и трех сверточных слоев CNN эффективна при поиске уязвимостей смарт-контрактов.

Ключевые слова: уязвимость, смарт-контракт, машинное обучение, BiGRU, CNN.

DOI: 10.14357/20790279240310 **EDN:** URZLYI

Введение

Современная цифровая экономика, поддерживаемая технологией блокчейн, сталкивается с постоянным ростом интереса к смарт-контрактам – программам, автоматизирующим и обеспечивающим выполнение условий сделок на блокчейн-платформах. Смарт-контракты являются фундаментальным элементом децентрализованных приложений (далее DApps) и обеспечивают правила децентрализованных и прозрачных отношений в сетях с нулевым доверием [1]. С другой стороны распространение использования смарт-контрактов выявляет несовершенство этой технологии. Все больше появляется информации об уязвимостях [2], что поднимает вопросы безопасности в контексте цифровых децентрализованных сред.

Недостаточная проверка безопасности перед развертыванием смарт-контрактов на платформах, таких как Ethereum, приводит к уязвимостям, которые могут быть использованы злоумышленниками для проведения атак и несанкционированного доступа к финансовым средствам [3]. Потеря средств из-за уязвимостей в смарт-контрактах может иметь

серьезные последствия для участников системы, подчеркивая необходимость разработки эффективных методов обнаружения и предотвращения подобных угроз. Если рассматривать финансовые потери, то в 2016 году общий урон от уязвимостей в смарт-контрактах составил 100 миллионов долларов, в 2017 – 200 миллионов, в 2018 – 1,5 миллиарда, в 2019 и 2020 – 500 миллионов, в 2021 – 3,3 миллиарда, в 2022 – 3,8 миллиарда и около 1,5 миллиардов в 2023 году [1].

Существующие методы обнаружения уязвимостей в смарт-контрактах ограничены в контексте быстро меняющейся криптографической и кибербезопасной среды. В связи с этим, использование машинного обучения представляется перспективным подходом для повышения точности и общей эффективности обнаружения уязвимостей, учитывая динамичность и сложность кода смарт-контрактов. Множество исследований на тему поиска уязвимостей в смарт-контрактах [4] в последние годы подчеркивает актуальность этой проблемы. Целью данного исследования является разработка нового метода, основанного на машинном обуче-

нии. Это не только позволит улучшить надежность обнаружения уязвимостей в смарт-контрактах, но и способствует обеспечению устойчивости и безопасности блокчейн-систем в целом.

1. Уязвимости смарт-контрактов

Смарт-контракты – это компьютерные программы, которые размещаются и выполняются на блокчейн-сети. Каждый смарт-контракт состоит из кода, определяющего заранее установленные условия, которые, когда они выполняются, выдают определенные результаты [5]. Запускаясь на децентрализованной блокчейн-сети вместо централизованного сервера, смарт-контракты позволяют нескольким сторонам приходиться к общему результату точным, своевременным и устойчивым к вмешательству образом. Блокчейны с возможностью использования смарт-контрактов предоставляют разработчикам возможность написания произвольной логики приложения поверх децентрализованного консенсуса, который блокчейны предоставляют по умолчанию.

Смарт-контракт определяет, как пользователи могут с ним взаимодействовать, включая, кто может взаимодействовать со смарт-контрактом, в какие моменты времени и какие вводы приводят к каким выводам. Результатом являются многосторонние цифровые соглашения, которые эволюционируют от сегодняшнего вероятностного состояния, где они, вероятно, будут выполнены по желанию, до нового детерминированного состояния, где они гарантированно будут выполняться в соответствии с их кодом [6].

В данном исследовании мы сфокусируемся на анализе уязвимостей смарт-контрактов на языке программирования Solidity и блокчейна Ethereum по следующим причинам:

- Широкое распространение. Ethereum является одним из самых популярных блокчейнов, используемых для развертывания смарт-контрактов.
- Множество смарт-контрактов. На Ethereum развернуто множество смарт-контрактов, предоставляющих обширный набор данных для исследования. Это обеспечивает разнообразие в типах контрактов, исследование которых может привести к обнаружению различных уязвимостей.
- Открытость и доступность сети Ethereum, которая является открытой и децентрализованной платформой. Данные о смарт-контрактах и их транзакциях доступны публично. Это обеспечивает открытость для исследователей, желающих провести анализ и оценку безопасности.

- Активное сообщество и развитие. Ethereum обладает активным сообществом разработчиков и исследователей, что обеспечивает поддержку и обмен опытом в области безопасности смарт-контрактов.

Согласно статье [7] наиболее распространенными и опасными уязвимостями в смарт-контрактах являются:

- арифметическое переполнение,
- некорректное управление доступом,
- неотмеченная отправка,
- рекурсия.

Исследователи используют различные методологии для анализа смарт-контрактов, многие из которых общедоступны под лицензиями с открытым исходным кодом. Существует пять общих методологий для анализа смарт-контрактов, таких как формальная верификация [8], символьное выполнение [9], фаззинг [10], промежуточное представление [11] и машинное обучение [12].

Машинное обучение достигает обнадеживающих результатов в области безопасности программ [4]. По сравнению с вышеупомянутыми методами и методологиями, оно объединяет в себе статический анализ и динамическое обнаружение для решения проблем высокого уровня ложных отрицательных результатов статического анализа и низкого покрытия кода динамического анализа.

2. Математическая постановка задачи поиска уязвимостей

2.1. Математическое определение уязвимости байт-кода смарт-контракта

Пусть дан байткод смарт-контракта C , представленный как последовательность инструкций $I_1, I_2, \dots, I_n$, где каждая инструкция I_i может содержать операции чтения/записи в хранилище, математические операции, вызовы других контрактов и т. д. Пусть также имеется множество возможных уязвимостей V , которые могут присутствовать в контракте.

Задача состоит в определении, содержит ли байткод C какую-либо уязвимость из множества V . Это можно сформулировать следующим образом:

- определение функции $f: C \rightarrow \{0,1\}$, где $f(C) = 1$, если байткод C содержит уязвимость, и $f(C) = 0$ в противном случае;
- пусть $V = \{v_1, v_2, \dots, v_m\}$ – множество из m возможных уязвимостей. Каждая уязвимость v_i может быть описана некоторым предикатом $P_i(C)$, который верен, если уязвимость v_i присутствует в байткоде C , и ложен в противном случае;

– тогда задача поиска уязвимостей в байткоде смарт-контракта сводится к проверке следующего условия:

$$f(C) = \bigvee_{i=1}^m P_i(C), \quad (1)$$

где $\vee$ обозначает логическое ИЛИ. То есть, байткод C содержит уязвимость, если хотя бы один из предикатов $P_i(C)$ истинен.

Подход к решению задачи включает анализ каждого предиката $P_i(C)$ для определения того, верен ли он для данного байткода C или нет.

2.2. Постановка задачи машинного обучения

При обнаружении обозначенных ранее уязвимостей в смарт-контрактах необходимо ставить машинное обучение как задачу многозначной классификации [13].

Пусть есть множество объектов X , множество классов C и обучающая выборка представлена парой (x_i, Y_i) , где x_i – объект, Y_i – множество меток, к которым принадлежит объект x_i . Таким образом, каждый объект может быть связан с несколькими метками.

Задачу многозначной классификации можно рассматривать как предсказание множества меток для новых объектов. Модель многозначной классификации обучается на обучающей выборке с использованием функции потерь, которая учитывает совпадение предсказанных и истинных меток.

Подходом к решению задачи многозначной классификации является использование бинарных классификаторов для каждой метки. Так, если у нас есть K классов, то для каждой метки c_i мы обучаем бинарный классификатор $h_i(x)$, который предсказывает, принадлежит ли объект x метке c_i или нет. Функция предсказания для многозначной классификации определена как

$$h(x) = \{c_i | h_i(x) = 1, i = 1, 2, \dots, K\}, \quad (2)$$

где $h_i(x)$ – предсказание бинарного классификатора для метки c_i , а i принадлежит диапазону от 1 до 4, включительно.

Для бинарной классификации одной метки будем использовать логистическую функцию потерь [15], которая имеет следующую формулу:

$$L_i(y_i, \hat{y}_i) = -\frac{1}{N} \sum_{j=1}^N (y_{ij} * \log(\hat{y}_{ij}) + (1 - y_{ij}) * \log(1 - \hat{y}_{ij})) \quad (3)$$

где N – количество примеров в обучающем наборе, y_{ij} – истинное значение i -ой метки для j -го примера (0 или 1), $\hat{y}_{ij}$ – предсказанная вероятность i -ой метки для j -го примера.

Следующие критерии указывают на то, что разрабатываемая модель справляется с поставленной задачей поиска уязвимостей эффективно [13]:

- должно быть превышающее процентное соотношение истинно положительный и истинно отрицательных исходов на матрице ошибок;
- значение метрик *accuracy*, *precision*, *recall* и *F1-score* должно быть не ниже 0,9;
- значение площади под ROC-кривой не ниже 0,8;
- значение общего времени поиска уязвимостей должно быть минимальным.

В ходе эксперимента мы рассмотрим 3 модели, в основании архитектур которых лежали следующие слои:

- 1) два рекуррентных слоя GRU [14],
- 2) два рекуррентных слоя GRU и два сверточных слоя CNN [15],
- 3) три двунаправленных рекуррентных слоя BiGRU и три сверточных слоя CNN.

Первый вариант является одним из самых распространенных в существующих исследованиях [16]. Два последующих проверяют стратегию объединения скорости и легкости сверточных сетей с чувствительностью к порядку рекуррентных сетей. Основная его идея заключается в использовании одномерной сверточной сети для предварительной обработки данных перед передачей их в рекуррентную сеть. Этот прием оказывается особенно выгодным, когда имеющиеся последовательности настолько длинны (с несколькими тысячами интервалов и больше), что их нереально обработать с помощью рекуррентной сети.

3. Результаты подготовки датасета и обучения модели

3.1. Подготовка датасета для обучения

Согласно статье [8], только лишь 1% всех смарт-контрактов в блокчейне Ethereum имеют открытый исходный код. Это может быть обусловлено тем, что исходные коды могут содержать конфиденциальную информацию или коммерческие секреты. Поэтому было принято решение использовать байт-код смарт-контрактов, который находится в публичном доступе на платформе блокчейн.

Формирование датасета осуществлялось из большого датасета `crypto_ethereum` Google BigQuery [17], содержащего информацию с блокчейна Ethereum. Он был создан для аналитики и исследований в области смарт-контрактов и транзакций Ethereum. Общая схема подготовки датасета представлена на рис. 1.

Сам байт-код смарт-контрактов был взят из таблицы `contracts` (рис. 2). Для загрузки понадобились только поля `bytecode` и `block_timestamp`. Последнее содержит временную метку блока, в



Рис. 1. Общая схема преобразования данных

котором был создан контракт. Метка понадобится для формирования свежих данных.

Далее был создан запрос для получения данных и их последующей выгрузки в несколько таблиц, с которыми можно непосредственно работать. Это делается по причине ограниченности ресурсов платформы Google BigQuery для выполнения запросов. Количество кортежей в проектных таблицах будет 200 тысяч, а всего таких таблиц 15 штук. Общий объем выгруженных данных составит 3 миллиона кортежей.

Перед разметкой данных была выполнена их предварительная обработка по удалению дубликатов. После предварительной обработки количество кортежей уменьшилось почти в 30 раз и составило 106 тысяч.

Далее были сопоставлены индексы из исходного датасета с индексами из размеченного текстового файла с помощью статического анализатора Conkas и были добавлены метки в новый столбец. Значения в столбце vuls ранжируются 0 до 4, где с 0 по 3 метки наши уязвимости, а метка 4 обозначает, что байт-код смарт-контракта безопасен (рис. 3).

Разделение датасета на обучающую, тестовую и валидационную выборки будем осуществ-

лять в процентном соотношении 80/10/10. Общее количество кортежей на тестовой выборке составило 85179, на валидационной и тестовой – 10647.

Распределение каждой метки класса в разделах обучения, тестирования и проверки представлено на рис. 4. Для этого сделаем разбиение количества классов уязвимостей по одному для каждого байт-кода. В результате будет несколько одинаковых байт-кодов с разными уязвимостями.

Анализ графика показывает, что классы довольно несбалансированные. Чтобы бороться с этой проблемой далее используем метрики оценки, которые учитывают дисбаланс классов.

Согласно статье [18] наиболее эффективными методами обработки байткода смарт-контрактов являются:

- замена байт-кодов, имеющих одинаковую функциональность в одну общую операцию. Примером могут послужить опкоды PUSH1 – PUSH32, которым соответствуют байт-коды 0x60 – 0xf7;
- избавление от значащего байта 0x во всех байт-кодах.

Далее, необходимо провести токенизацию [19]. Иными словами разбить байткоды на серию байтов, каждый из которых содержит два символа

Field name	Type	Mode
address	STRING	REQUIRED
bytecode	STRING	NULLABLE
function_sighashes	STRING	REPEATED
is_erc20	BOOLEAN	NULLABLE
is_erc721	BOOLEAN	NULLABLE
block_timestamp	TIMESTAMP	REQUIRED
block_number	INTEGER	REQUIRED
block_hash	STRING	REQUIRED

Рис. 2. Поля таблицы с данными о смарт-контрактах

	bytecode_c	vuls
0	0x6080604052600436106100e6576000357c0100000000...	[4]
1	0x608060405234801561001057600080fd5b5060043610...	[4]
2	0x608060405260043610610174576000357c0100000000...	[1]
3	0x608060405234801561001057600080fd5b5060043610...	[4]
4	0x6080604052600436106100d0576000357c0100000000...	[4]
...	...	...
85174	0x608060405234801561001057600080fd5b5060043610...	[3]
85175	0x60806040526004361061008a5760003560e01c80638b...	[4]
85176	0x608060405234801561001057600080fd5b5060043610...	[2, 3]
85177	0x608060405234801561001057600080fd5b5060043610...	[1, 2]
85178	0x6080604052600436106101d05760003560e01c80637d...	[2]

Рис. 3. Пример кортежей обучающей выборки

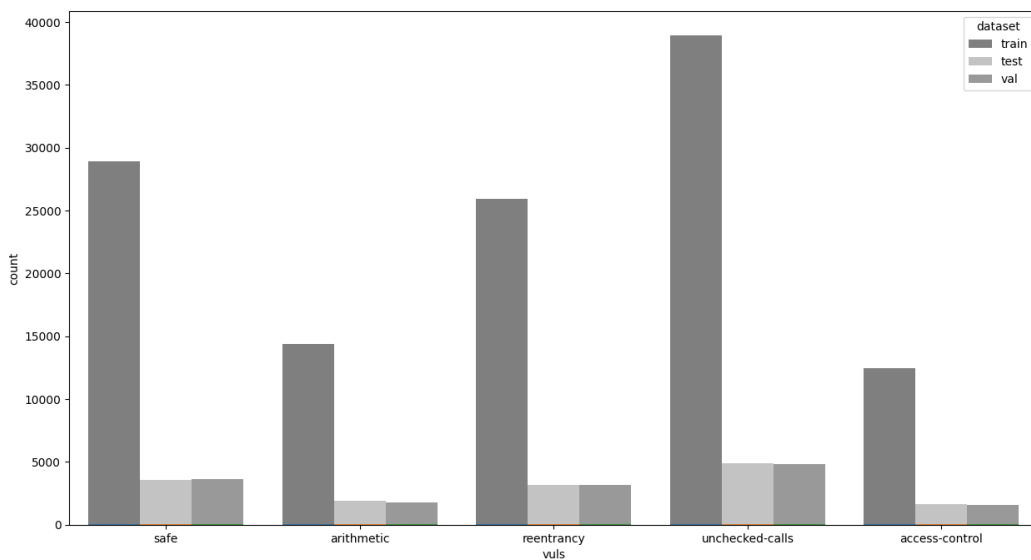


Рис. 4. Классовое распределение данных

и поможет раскрыть особенности смарт-контрактов. Токенизовав байт-код, определим длину выходной последовательности, которая покрывает 95% данных. Выбор 95-перцентиля означает, что мы игнорируем 5% данных с наибольшими длинами, что может оказаться не таким критичным для обучения модели, но позволяет избежать слишком больших значений выходной последовательности, которые могли бы привести к избыточным затратам по памяти и вычислительным ресурсам.

Далее метод `text_vectorizer` из библиотеки TensorFlow выполнит задачу векторизации, преобразуя текстовые данные в числовые последова-

тельности [20]. Следующим шагом будет преобразование обычных векторов в векторы плотных эмбедингов. За это будет отвечать соответствующий слой представленных далее моделей [21].

3.2. Обучение первой модели

Схема архитектуры данной модели представлена на рис. 5. О первых четырех блоках слоев было рассказано выше. Теперь поговорим о роли слоев Dropout и Dense. Слой прореживания или Dropout – один из наиболее эффективных и распространенных приемов регуляризации для нейронных сетей [22]. Прореживание, которое применяется

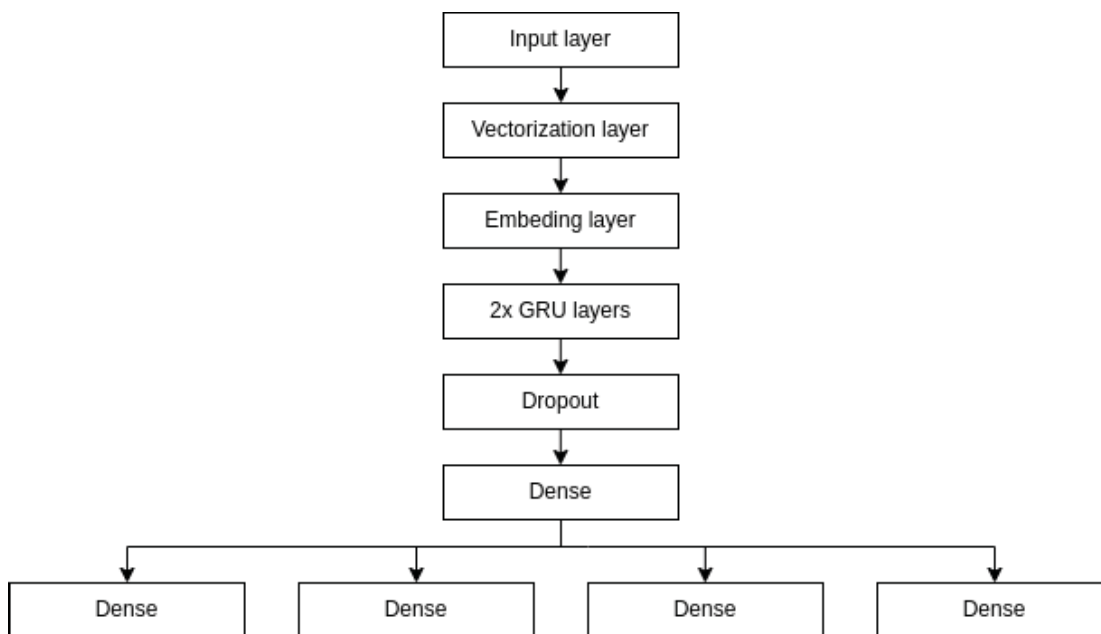


Рис. 5. Архитектура первой модели

к слою, заключается в удалении (присваивании нуля) случайно выбираемым признакам на этапе обучения.

Dense-слои, также известные как fully connected (полносвязные) слои, являются основными строительными блоками нейронных сетей. Они соединяют каждый нейрон из предыдущего слоя с каждым нейроном следующего слоя. Это позволяет модели обучаться на более сложных комбинациях признаков, которые могут быть выявлены в предыдущих слоях. В нашей архитектуре он принимает на вход выходные данные последнего слоя GRU и преобразует их в пространство той же размерности с применением функции активации ReLU [23]. Следующие за этим четыре Dense слоя указывают, что они будут выдавать значение в диапазоне от 0 до 1, что подходит для задачи бинарной классификации.

Для минимизации функции потерь в процессе обучения модели использовался оптимизатор Adam [24]. В нашем случае, данный параметр будет уменьшаться каждые пять эпох на одну десятую, если модель покажет результаты хуже, чем раньше.

Еще одним важным параметром является количество эпох обучения. Эпоха в обучении моделей представляет собой один проход через весь тренировочный набор данных. Количество эпох определяет, сколько раз весь тренировочный набор будет использоваться для обучения модели. На каждой эпохе модель проходит через все примеры в тренировочном наборе данных, вычисляет потери и обновляет веса сети в соответствии с выбранным оптимизатором и функцией потерь. Также нельзя не отметить такой параметр как размер батча. Он определяет количество образцов данных, которые обрабатываются одновременно перед выполнением одного шага обновления весов модели.

Модель будет обучена со следующими параметрами:

- количество эпох – 20,
- размер батча – 352,
- оптимизатор: Adam,
- коэффициент скорости обучения: на начальном этапе 0.001,
- коэффициент прореживания на слое Dropout – 0.2.

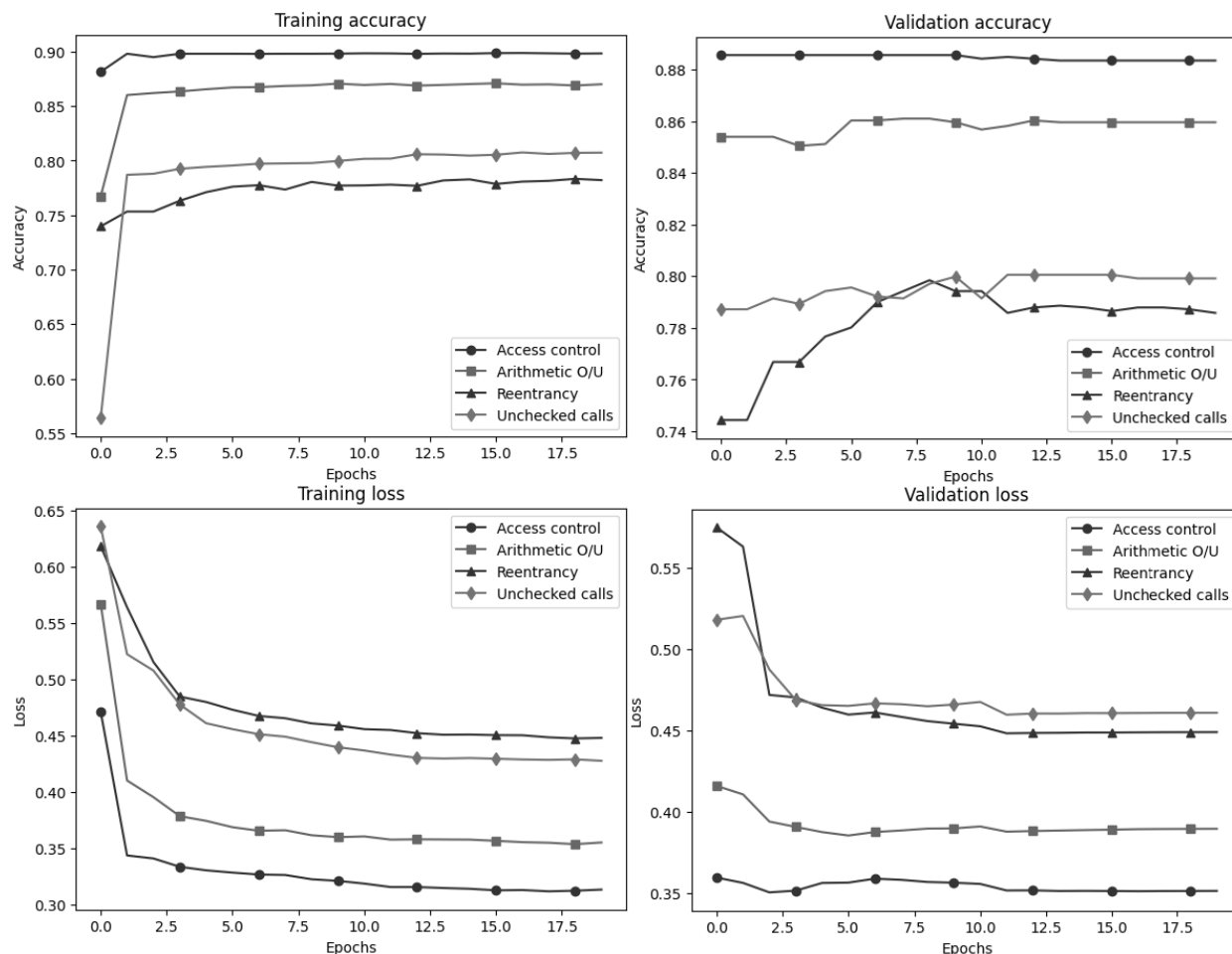


Рис. 6. Результаты обучения первой модели

Табл. 1

Метрики качества первой модели

Метрика/ Уязвимость	Access control	Arithmetic O/U	Reentrancy	Unchecked calls
Accuracy	0.90	0.87	0.77	0.79
Precision	0.81	0.87	0.75	0.75
Recall	0.90	0.87	0.77	0.79
F1-score	0.85	0.82	0.75	0.75

Результаты обучения показаны на рис. 6. Общее время обучения составило 5 дней. На одну эпоху примерно 6 часов. На тестовой выборке модель показала значения, представленные в табл. 1.

Как видно, результаты обучения низкие, поэтому было принято решение адаптировать архитектуру внедрением одномерных сверточных слоев.

3.3. Обучение второй модели

Схема архитектуры данной модели представлена на рис. 7.

Увеличение емкости сети обычно осуществляется за счет увеличения числа параметров слоя или добавления дополнительных слоев. Наложение рекуррентных слоев друг на друга – классический способ конструирования бо-

лее мощных рекуррентных сетей. Например, в настоящее время алгоритм Google Translate представляет собой стек из семи больших слоев LSTM – это огромная сеть.

Между сверточными и рекуррентными слоями был добавлен слой Reshape. Он преобразует выходные данные пулинга к форме, подходящей для входа в рекуррентные слои.

По сравнению с первой моделью было изменено только количество эпох обучения, оно стало равно 60. Результаты обучения показаны на рис. 8.

Общее время обучения данной модели составило примерно 2 дня. На одну эпоху уходило 45 минут. На тестовой выборке модель показала значения, представленные в табл. 2.

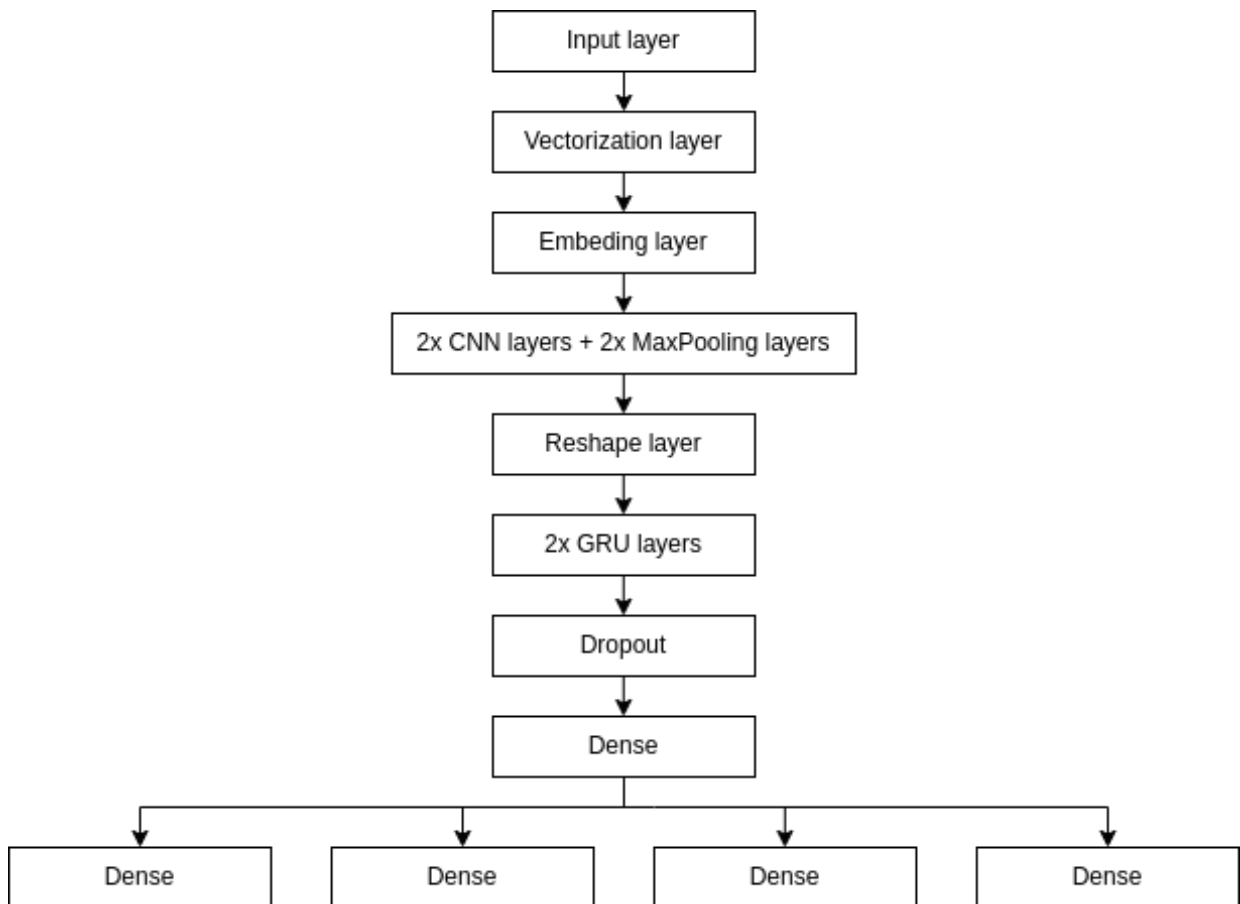


Рис. 7. Архитектура второй модели

Табл. 2

Метрики качества второй модели

Метрика/ Уязвимость	Access control	Arithmetic O/U	Reentrancy	Unchecked calls
Accuracy	0.88	0.87	0.86	0.86
Precision	0.88	0.85	0.86	0.85
Recall	0.89	0.87	0.86	0.86
F1-score	0.88	0.85	0.87	0.86

Качество модели значительно улучшилось по всем метрикам уязвимостей рекурсии и неотмеченной отправки. По арифметическому переполнению и управлению доступом улучшилась F1-мера. Но при этом, точность обучения в уязвимости управления доступом упала на 2 сотых. Это не критично, так как с несбалансированной выборкой как у нас, лучше всего ориентироваться на последние 3 метрики.

3.4. Обучение финальной модели

Схема архитектуры финальной модели представлена на рис. 9.

В финальной модели было принято решение «утяжелить» нашу модель, добавив по одному одномерному сверточному и рекуррентному слою, а также

сделать рекуррентные сети двунаправленными. Обрабатывая последовательность в двух направлениях, двунаправленная рекуррентная сеть способна выявить шаблоны, незаметные для однонаправленной сети. Основные параметры модели не отличались от предыдущей. Результаты обучения показаны на рис. 10. На тестовой выборке модель показала наилучшие результаты. Они представлены в табл. 3.

Теперь посмотрим, насколько модель ошибалась в классификации по всем уязвимостям (рис. 11). Сделать это можно с помощью матрицы ошибок.

В качестве инструментов для сравнения были выбраны два популярных анализатора, один статический Mythril[25], другой динамический sFuzz[26], а

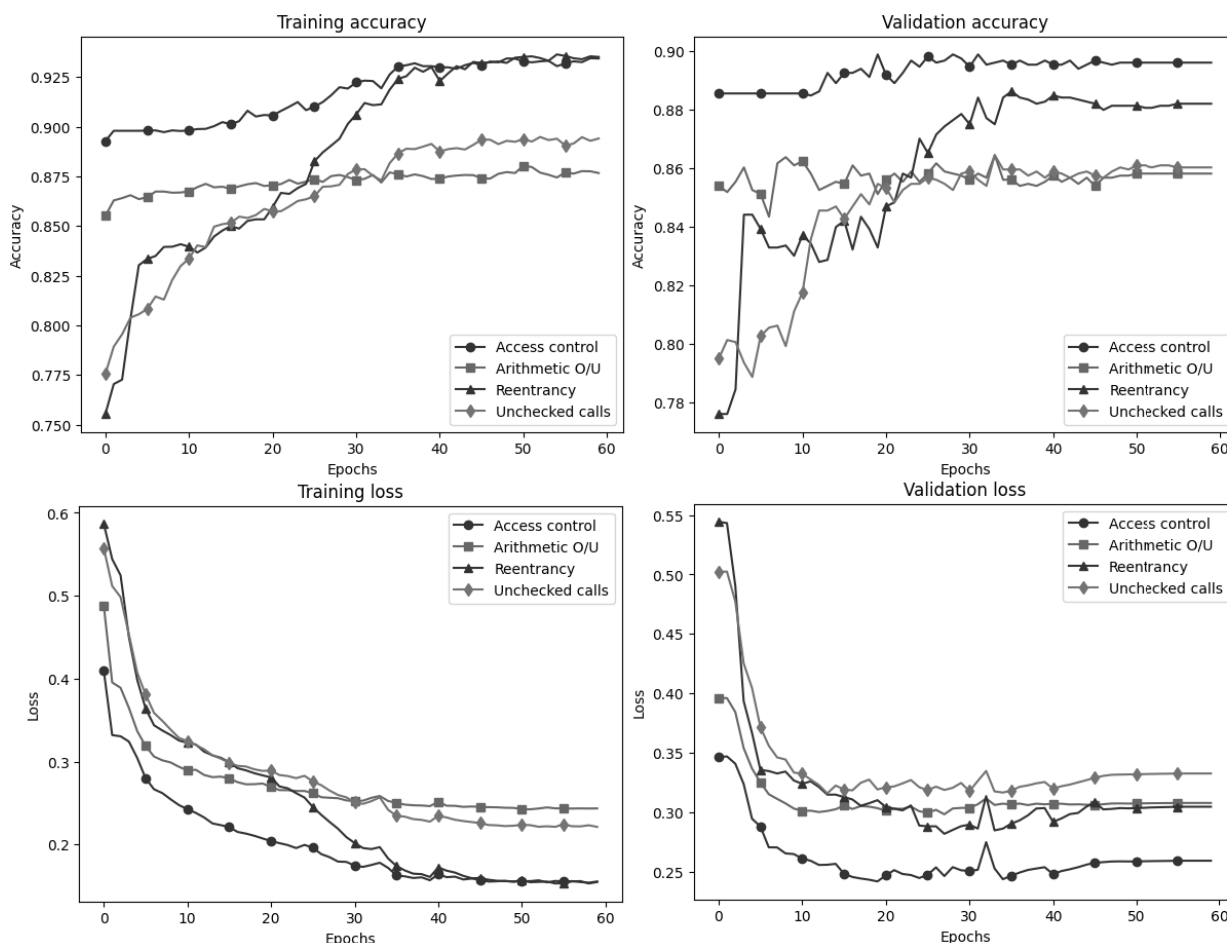


Рис. 8. Результаты обучения второй модели

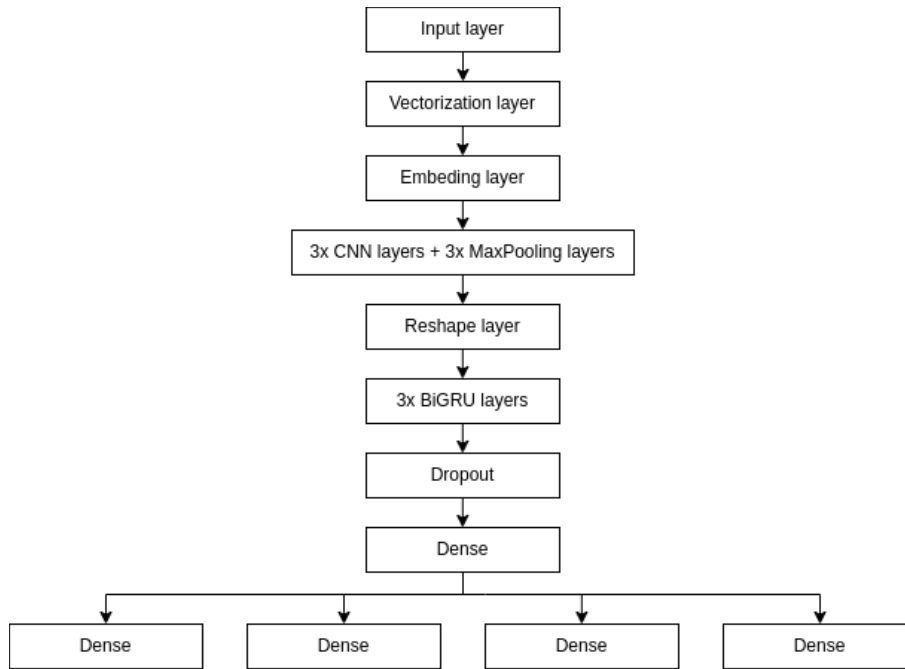


Рис. 9. Архитектура финальной модели

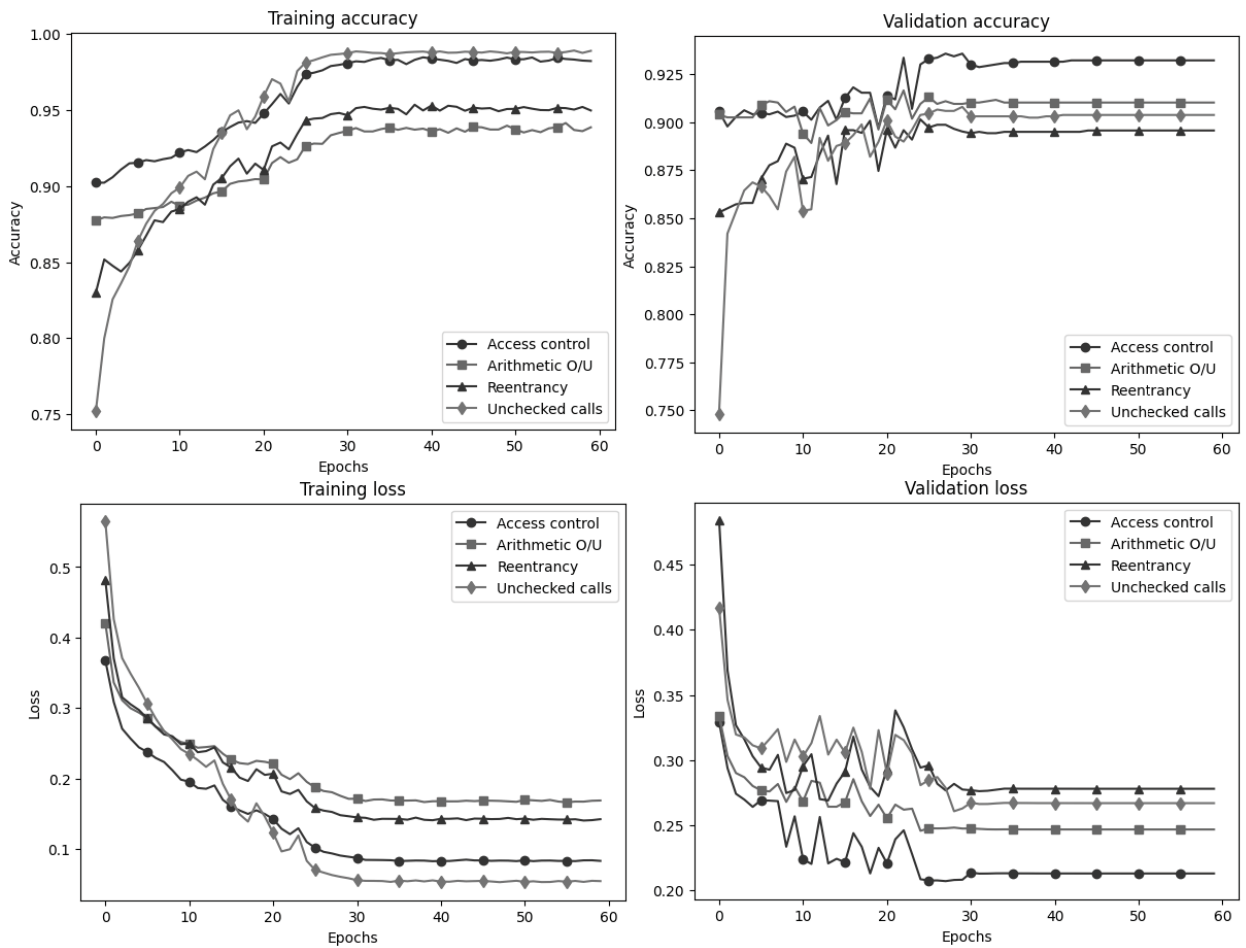


Рис. 10. Результаты обучения финальной модели

Табл. 3

Метрики качества финальной модели

Метрика/ Уязвимость	Access control	Arithmetic O/U	Reentrancy	Unchecked calls
Accuracy	0.937	0.914	0.908	0.902
Precision	0.936	0.911	0.907	0.902
Recall	0.937	0.914	0.908	0.902
F1-score	0.936	0.912	0.908	0.902

также их метрики качества из соответствующей статьи[27]. Результаты сравнения представлены в табл. 4.

Практически по всем показателем разработанный инструмент превосходит уже зарекомендовавшие себя анализаторы уязвимостей.

Заключение

В представленном исследовании был разработан метод поиска уязвимостей в смарт-контрак-

тах на основе машинного обучения и проведен ряд экспериментов, который использовал начальный датасет `crypto_ethereum`.

Для использования датасета для обучения был проведен этап подготовки данных. В результате на подготовленном датасете было обучено 3 модели. После обучения каждой модели последовательно проводилось их «утяжеление», пока не были достигнуты изначально установленные критерии эффективности ее применения.

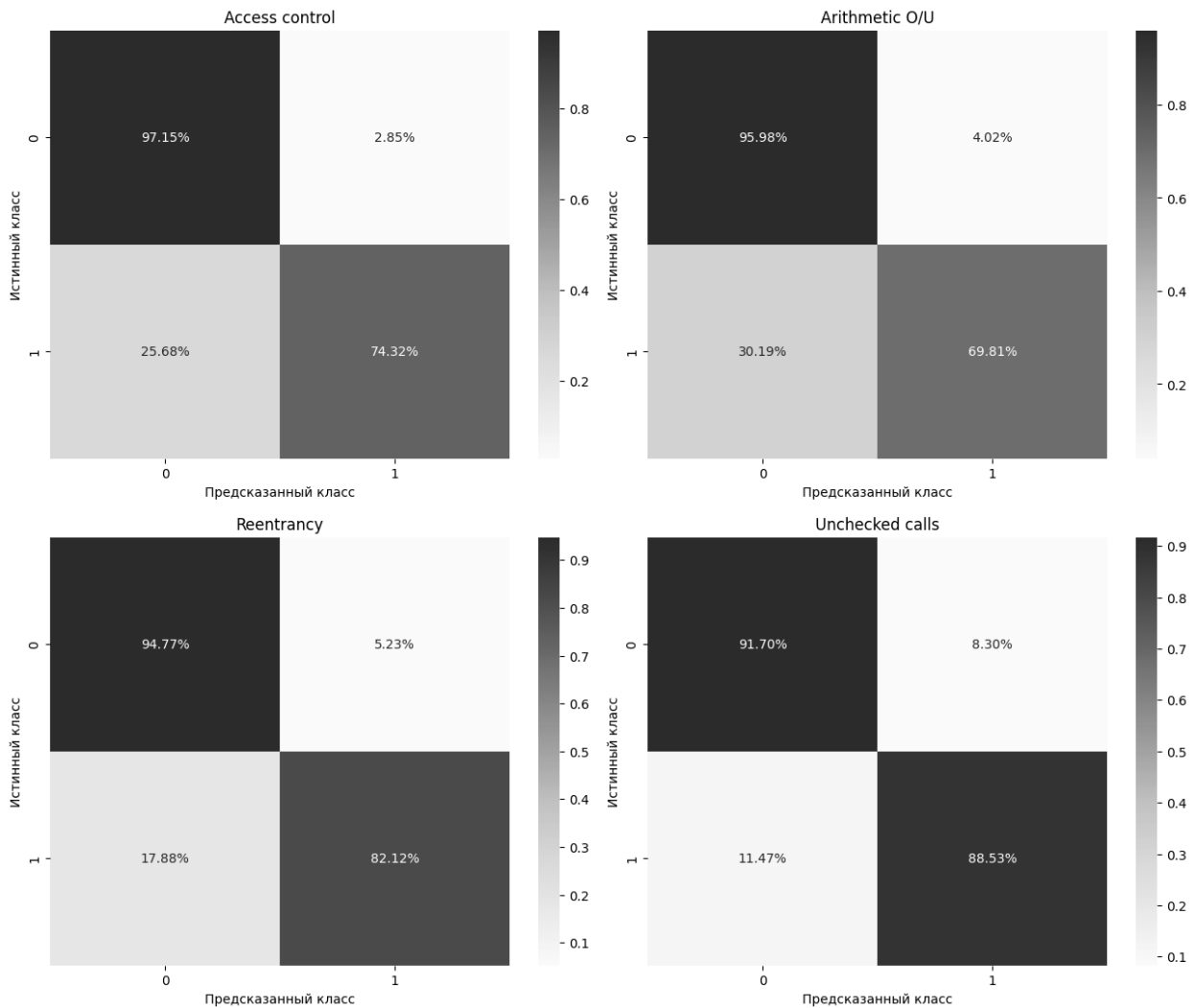


Рис. 11. Матрица ошибок финальной модели

Табл. 4

Сравнение разработанной модели с статическими инструментами

Уязвимость	Метрика/Критерий	Инструмент		
		Mythril	sFuzz	Разработанная модель
1	2	3	4	5
Reentrancy	Accuracy	70.4%	34.7%	90.84%
	Precision	100%	100%	90.75%
	Recall	67.2%	21.1%	90.84%
	F1-score	78.2%	34.7%	90.76%
Unchecked calls	Accuracy	87%	74%	90.22%
	Precision	100%	100%	90.22%
	Recall	82.7%	65.4%	90.22%
	F1-score	90.5%	79.1%	90.22%
Arithmetic O/U	Accuracy	58.5%	47.6%	91.45%
	Precision	100%	100%	91.15%
	Recall	47.7%	33.8%	91.45%
	F1-score	64.6%	50.6%	91.25%
Access control	Accuracy	77.8%	N/A%	93.75%
	Precision	100%	N/A%	93.56%
	Recall	40%	N/A %	93.75%
	F1-score	57.1%	N/A %	93.62%
Среднее время обнаружения уязвимостей		609 секунд	289 секунд	0.22 секунды

Сравнительный анализ с другими известными анализаторами уязвимостей показал эффективность машинного обучения в решении задачи классификации уязвимостей смарт-контрактов. Важным преимуществом описанного подхода является время обнаружения уязвимостей в процессе работы, что доказано экспериментально.

В будущем планируется провести подробный сравнительный анализ разработанных моделей обучения с последними исследованиями в этой области, а также апробировать его применение на реальных данных смарт-контрактов основной сети Ethereum.

Литература

1. Ray P.P. Web3: A comprehensive review on background, technologies, applications, zero-trust architectures, challenges and future directions // Internet of Things and Cyber-Physical Systems. 2023.
2. Crypto Hacks 2023: Full List of Scams and Exploits as Millions Go Missing. // ccn.com. 2024. URL: <https://www.ccn.com/education/crypto-hacks-2023-full-list-of-scams-and-exploits-as-millions-go-missing>
3. Huang Y. et al. Smart contract security: A software lifecycle perspective //IEEE Access. 2019. Vol. 7. P. 150184-150202.
4. Kiani R., Sheng V.S. Ethereum Smart Contract Vulnerability Detection and Machine Learning-Driven Solutions: A Systematic Literature Review //Electronics. 2024. Vol. 13. No 12. P. 2295.
5. Mukhopadhyay M. Ethereum Smart Contract Development: Build blockchain-based decentralized applications using solidity. Packt Publishing Ltd. 2018.
6. What Are Smart Contracts and How Do They Work? // chain.link. 2023. URL: <https://chain.link/education/smart-contracts> (дата обращения: 17.11.2023)
7. Wei Z., Sun J., Zhang Z., Zhang X., Yand X., Zhu L. Survey on Quality Assurance of Smart Contracts. // ACM Comput. Surv. 2023. URL: <https://arxiv.org/pdf/2311.00270.pdf> (дата обращения: 17.11.2023)
8. Harz D., Knottenbelt W. Towards Safer Smart Contracts: A Survey of Languages and Verification Methods. // arXiv:1809.09805. 2018. URL: <https://arxiv.org/pdf/1809.09805.pdf> (дата обращения: 17.11.2023)
9. Brousmiche K., Abdellatif T. Formal Verification of Smart Contracts Based on Users and Blockchain Behaviors Models. // 9th IFIP International Conference on New Technologies, Mobility and Security. 2018. URL: https://www.researchgate.net/publication/324175498_Formal_Verification_of_Smart_Contracts_Based_on_Users_and_Blockchain_Behaviors_Models (дата обращения: 17.11.2023)
10. He J., Balunovic M., Ambroladze N., Tsankov P., Martin T. Learning to Fuzz from Symbolic

- Execution with Application to Smart Contracts. // In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. 2019. URL: <https://dl.acm.org/doi/10.1145/3319535.3363230> (дата обращения: 17.11.2023)
11. *Click C., Paleczny M.* A simple graph-based intermediate representation // ACM Sigplan Notices. 1995. Vol. 30. No 3. P. 35-49.
 12. *Chakraborty S., Krishna R., Ding Y., Ray B.* Deep Learning Based Vulnerability Detection: Are We There Yet? // IEEE Transactions on Software Engineering. 2022. URL: https://www.researchgate.net/publication/352279734_Deep_Learning_based_Vulnerability_Detection_Are_We_There_Yet (дата обращения: 17.11.2023)
 13. Multilabel Classification: An Introduction with Python's Scikit-Learn. // KDnuggets. 2023. URL: <https://www.kdnuggets.com/2023/08/multilabel-classification-introduction-python-scikitlearn.html> (дата обращения: 17.11.2023)
 14. Gated Recurrent Unit Networks. // geeksforgeeks. 2023. URL: <https://www.geeksforgeeks.org/gated-recurrent-unit-networks/> (дата обращения: 12.01.2024)
 15. Crash Course in Convolutional Neural Networks for Machine Learning // machinelearningmastery. 2023. URL: <https://machinelearningmastery.com/crash-course-convolutional-neural-networks/> (дата обращения: 12.01.2024)
 16. *Zhuang Y. et al.* Smart contract vulnerability detection using graph neural networks // Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. 2021. P. 3283-3290.
 17. From cloud data warehouse to an AI-ready data platform. // Google BigQuery. 2024. URL: <https://cloud.google.com/bigquery?hl=ru> (дата обращения: 12.01.2024)
 18. *Sendner C., Chen H., Fereidooni H., Petzi L., König J., Stang J., Dmitrienko A.* Smarter Contracts: Detecting Vulnerabilities in Smart Contracts with Deep Transfer Learning // 2023. URL: https://www.ndss-symposium.org/wp-content/uploads/2023/02/ndss2023_s263_paper.pdf (дата обращения: 17.11.2023)
 19. *Zouhar V., Meister C., Gastaldi J., Du L., Sachan M., Cotterell R.* Tokenization and the Noiseless Channel. // 2023.acl-long.284. 2023. URL: <https://aclanthology.org/2023.acl-long.284> (дата обращения: 17.11.2023)
 20. Text Vectorization layer. // TensorFlow. 2023. URL: https://www.tensorflow.org/api_docs/python/tf/keras/layers/TextVectorization (дата обращения: 17.11.2023)
 21. Embedding layer. // TensorFlow. 2023. URL: https://www.tensorflow.org/api_docs/python/tf/keras/layers/Embedding (дата обращения: 17.11.2023)
 22. *Шолле Ф.* Глубокое обучение на Python. СПб.: Питер. 2018. 400 с.
 23. *Жерон О.* Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем. СПб.: ООО «Диалектика». 2020. 1040 с.
 24. Optimizers Adam. // TensorFlow. 2024. URL: https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam (дата обращения: 30.03.2024).
 25. Mythril. Security analysis tool for EVM bytecode // GitHub. 2024. URL: <https://github.com/Consensys/mythril> (дата обращения: 30.03.2024)
 26. sFuzz. // GitHub. 2024. URL: <https://github.com/duytai/sFuzz> (дата обращения: 30.03.2024)
 27. *Wei Z., Sun J., Zhang Z., Zhang X., Li M., Zhu L.* A Comparative Evaluation of Automated Analysis Tools for Solidity Smart Contracts. // arXiv:2310.20212v. 2023. URL: <https://arxiv.org/pdf/2310.20212> (дата обращения: 17.11.2023)

Тарханов Иван Александрович. Федеральный исследовательский центр «Информатика и управление» Российской академии наук, г. Москва, Россия. Старший научный сотрудник. Кандидат технических наук, доцент. Область научных интересов: электронный документооборот, блокчейн, информационная безопасность. E-mail: tarkhanov@isa.ru (Ответственный за переписку).

Белоус Вадим Сергеевич. Национальный исследовательский технологический университет «МИСиС», г. Москва, Россия. Магистр. Область научных интересов: машинное обучение, блокчейн. E-mail: belous.vadim@inbox.ru

Search for vulnerabilities in smart contracts based on machine learning

V.S Belous^I, I.A. Tarkhanov^{II,III}^I National University of Science and Technology "MISiS", Moscow, Russia^{II} State Academic University for Humanities, Moscow, Russia^{III} Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

Abstract. With the rising popularity of blockchain projects, the number of decentralized applications based on them is also growing. The central element of these applications are smart contracts. This technology is still relatively new and has a number of security issues. Statistics of smart contract hacking indicate the relevance of finding vulnerabilities in smart contract code problem. The article describes 3 machine learning models for searching for vulnerabilities in smart contracts written in the Solidity language. Particular attention is paid to preparing the dataset for training and comparing it with well-known code analyzers. The metrics obtained from the results of training and testing the models suggest that the model consisting of three bidirectional recurrent BiGRU layers and three convolutional CNN layers is effective in the task of searching for smart contract vulnerabilities.

Keyword: *vulnerability, smart contract, machine learning.*

DOI: 10.14357/20790279240310 **EDN:** URZLYI

References

1. Ray P.P. Web3: A comprehensive review on background, technologies, applications, zero-trust architectures, challenges and future directions // Internet of Things and Cyber-Physical Systems. 2023.
2. Crypto Hacks 2023: Full List of Scams and Exploits as Millions Go Missing. // ccn.com. 2024. URL: <https://www.ccn.com/education/crypto-hacks-2023-full-list-of-scams-and-exploits-as-millions-go-missing>
3. Huang Y. et al. Smart contract security: A software lifecycle perspective //IEEE Access. 2019. V. 7. P. 150184-150202.
4. Kiani R., Sheng V.S. Ethereum Smart Contract Vulnerability Detection and Machine Learning-Driven Solutions: A Systematic Literature Review //Electronics. 2024. V. 13. No. 12. P. 2295.
5. Mukhopadhyay M. Ethereum Smart Contract Development: Build blockchain-based decentralized applications using solidity. Packt Publishing Ltd, 2018.
6. What Are Smart Contracts and How Do They Work? // chain.link. 2023. URL: <https://chain.link/education/smart-contracts> (дата обращения: 17.11.2023)
7. Wei Z., Sun J., Zhang Z., Zhang X., Yand X., Zhu L. Survey on Quality Assurance of Smart Contracts. // ACM Comput. Surv. 2023. URL: <https://arxiv.org/pdf/2311.00270.pdf> (дата обращения: 17.11.2023)
8. Harz D., Knottenbelt W. Towards Safer Smart Contracts: A Survey of Languages and Verification Methods. // arXiv:1809.09805. 2018. URL: <https://arxiv.org/pdf/1809.09805.pdf> (дата обращения: 17.11.2023)
9. Brousmiche K., Abdellatif T. Formal Verification of Smart Contracts Based on Users and Blockchain Behaviors Models. // 9th IFIP International Conference on New Technologies, Mobility and Security. 2018. URL: https://www.researchgate.net/publication/324175498_Formal_Verification_of_Smart_Contracts_Based_on_Users_and_Blockchain_Behaviors_Models (дата обращения: 17.11.2023)
10. He J., Balunovic M., Ambroladze N., Tsankov P., Martin T. Learning to Fuzz from Symbolic Execution with Application to Smart Contracts. // In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. 2019. URL: <https://dl.acm.org/doi/10.1145/3319535.3363230> (дата обращения: 17.11.2023)
11. Click C., Paleczny M. A simple graph-based intermediate representation //ACM Sigplan Notices. 1995. V. 30. No. 3. P. 35-49.
12. Chakraborty S., Krishna R., Ding Y., Ray B. Deep Learning Based Vulnerability Detection: Are We There Yet? // IEEE Transactions on Software Engineering. 2022. URL: https://www.researchgate.net/publication/352279734_Deep_Learning_based_Vulnerability_Detection_Are_We_There_Yet (дата обращения: 17.11.2023)
13. Multilabel Classification: An Introduction with Python's Scikit-Learn. // KDnuggets. 2023. URL: <https://www.kdnuggets.com/2023/08/multilabel-classification-introduction-python-scikitlearn.html> (дата обращения: 17.11.2023)
14. Gated Recurrent Unit Networks. // geeksforgeeks. 2023. URL: <https://www.geeksforgeeks.org/>

- gated-recurrent-unit-networks/ (дата обращения: 12.01.2024)
15. Crash Course in Convolutional Neural Networks for Machine Learning // machinelearningmastery. 2023. URL: <https://machinelearningmastery.com/crash-course-convolutional-neural-networks/> (дата обращения: 12.01.2024)
 16. Zhuang Y. et al. Smart contract vulnerability detection using graph neural networks // Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. 2021. P. 3283-3290.
 17. From cloud data warehouse to an AI-ready data platform. // Google BigQuery. 2024. URL: <https://cloud.google.com/bigquery?hl=ru> (дата обращения: 12.01.2024)
 18. Sendner C., Chen H., Fereidooni H., Petzi L., König J., Stang J., Dmitrienko A. Smarter Contracts: Detecting Vulnerabilities in Smart Contracts with Deep Transfer Learning // ISBN 1-891562-83-5. 2023. URL: https://www.ndss-symposium.org/wp-content/uploads/2023/02/ndss2023_s263_paper.pdf (дата обращения: 17.11.2023)
 19. Zouhar V., Meister C., Gastaldi J., Du L., Sachan M., Cotterell R. Tokenization and the Noiseless Channel. // 2023.acl-long.284. 2023. URL: <https://aclanthology.org/2023.acl-long.284> (дата обращения: 17.11.2023)
 20. Text Vectorization layer. // TensorFlow. 2023. URL: https://www.tensorflow.org/api_docs/python/tf/keras/layers/TextVectorization (дата обращения: 17.11.2023)
 21. Embedding layer // TensorFlow. 2023. URL: https://www.tensorflow.org/api_docs/python/tf/keras/layers/Embedding (дата обращения: 17.11.2023)
 22. Sholle F. Glubokoe obuchenie na Python. SPb.: Piter. 2018. 400 p.
 23. Zheron O. Prikladnoe mashinnoe obuchenie s pomoshh'ju Scikit-Learn, Keras i TensorFlow: koncepcii, instrumenty i tehniki dlja sozdaniya intellektual'nyh sistem. SPb.: ООО «Dialektika», 2020.
 24. Optimizers Adam. // TensorFlow. 2024. URL: https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam (дата обращения: 30.03.2024).
 25. Mythril. Security analysis tool for EVM bytecode // GitHub. 2024. URL: <https://github.com/Consensys/mythril> (дата обращения: 30.03.2024)
 26. sFuzz. // GitHub. 2024. URL: <https://github.com/duytai/sFuzz> (дата обращения: 30.03.2024)
 27. Wei Z., Sun J., Zhang Z., Zhang X., Li M., Zhu L. A Comparative Evaluation of Automated Analysis Tools for Solidity Smart Contracts. // arXiv:2310.20212v. 2023. URL: <https://arxiv.org/pdf/2310.20212> (дата обращения: 17.11.2023)

Tarkhanov I.A. PhD, State Academic University for Humanities, Moscow, Russia, e-mail: tarkhanov@isa.ru

Belous V.S. Student, National University of Science and Technology “MISiS”, Moscow, Russia, e-mail: belous.vadim@inbox.ru

Trudy Instituta sistemnogo analiza Rossiyskoy akademii nauk (ISA RAN) **(Proceedings of the Institute for Systems Analysis Russian Academy of Sciences (ISA RAS))**

<http://www.isa.ru/proceedings/>

The journal was founded in 2002
The journal is published quarterly

Founder

Federal Research Center "Computer Science and Control" of Russian Academy of Sciences (FRC CSC RAS)
44/2, Vavilova str., Moscow, 119333, Russia, tel: +7(499)135-62-60,
e-mail: frccsc@frccsc.ru, <http://www.frccsc.ru>

Editor-in-Chief

Popkov Y.S., Academician of RAS, Professor,
9, Prospekt 60-letiya Oktyabrya, Moscow, 117312, Russia,
tel: +7(499)135-42-22, e-mail: popkov@isa.ru

Editorial Chairs

Popkov Y.S., Academician of RAS, Professor
Chairman of Editorial Board (Moscow, Russia)
Konstantinov I.S., Professor (Belgorod, Russia)
Scherbov S.Y., Professor (Vienna, Austria)
Smuelian B.L., Professor (Tel-Aviv, Israel)
Solovyov A.V., Professor,
Deputy Editor-in-Chief (Moscow, Russia)
van Wissen Leonardus J.G., Professor (Amsterdam, Netherlands)

Editorial Board

Aleskerov F.T., Professor (Moscow, Russia)	Livshits V.N., Professor (Moscow, Russia)
Ashimov A.A., Professor (Almaty, Kazakhstan)	Magnitsky N.A., Professor (Moscow, Russia)
Belov S.P., Professor (Belgorod, Russia)	Mizhidon A.D., Professor (Ulan-Ude, Russia)
Bulychev A.V., Ph.D., Executive Secretary (Moscow, Russia)	Orlova E.R., Professor (Moscow, Russia)
Burkov V.N., Professor (Moscow, Russia)	Petrovsky A.B., Professor (Moscow, Russia)
Gelovani V.A., Academician of RAS, Professor (Moscow, Russia)	Shvetsov A.N., Professor (Moscow, Russia)
Ilin A.V., Corresponding Members of RAS (Moscow, Russia)	Smirnov A.V., Professor (St. Petersburg, Russia)
Krut'ko V.N., Professor (Moscow, Russia)	Tishenko V.I., Ph.D. (Moscow, Russia)
Lexin V.N., Professor (Moscow, Russia)	Voytishskiy A.V., Professor (Novosibirsk, Russia)

Editorial Office

Shibaeva L.S., managing editor

Editorial address

9, Prospekt 60-letiya Oktyabrya, Moscow, 117312, Russia
E-mail: tisa@isa.ru, tel. +7 (499) 135-24-33, +7 (926) 490-87-87

Scientific journal "Trudy Instituta sistemnogo analiza Rossiyskoy akademii nauk (ISA RAN)" publishes materials on a wide range of fundamental problems of developing systems analysis methodology and its applying to solving various problems in the field of science and practice. The journal is destined for scientists, engineers and researchers working within the framework of these problems, as well as for politicians, employees of state and municipal departments, specialists of enterprises and representatives of social organizations. The rules for articles submission to the journal, as well as for their review are given at the journal's site

The journal is registered in Federal service of observing and controlling legislation in the area of mass media and protection of cultural heritage in 20 of May 2016. Digital certificate - ПИ № ФС77-65718.
Journal Subscription Index in the Ural-Press e-catalog (<http://ural-press.ru>) is 36089.

The journal is supported by the Department of Information Technologies and Computing Systems of the Russian Academy of Sciences.
The journal is included in the list of leading peer-reviewed scientific journals recommended by the Higher Attestation Commission.
Permission of the editorial board and authors is required when using data of the journal.

© Federal Research Center "Computer Science and Control" of Russian Academy of Sciences

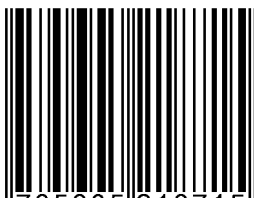
Vol. **74**
Issue **3**

ISSN 2079-0279

2024

TRUDY
INSTITUTA SISTEMNOGO ANALIZA
ROSSIYSKOY AKADEMII NAUK (ISA RAN)
(PROCEEDINGS OF THE INSTITUTE FOR SYSTEMS ANALYSIS
RUSSIAN ACADEMY OF SCIENCES (ISA RAS))

ISBN 978-5-6052187-1-5



9 785605 218715 >

Отпечатано в ООО «Поли Принт Сервис»
127015, г. Москва, ул. Бутырская, д. 86
127247, г. Москва, Дмитровское шоссе, д. 100, стр. 2
Формат 60x84/8. Зак. № В-506
Тираж 70 экз.
Тел.: +7 (495) 797-35-59