

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Факультет прикладной математики и информатики
BELARUSIAN STATE UNIVERSITY
Faculty of Applied Mathematics and Informatics

ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Материалы
XI Международного научного конгресса по информатике
(CSIST-2025)

Республика Беларусь
Минск, 29–31 октября 2025 г.

В двух частях
Часть 1

INFORMATION SYSTEMS AND TECHNOLOGIES

Proceedings
of the XI International Scientific Congress on Computer Science
(CSIST-2025)

Republic of Belarus
Minsk, October 29–31, 2025

In two parts
Part 1

Научное электронное издание

Минск, БГУ, 2025

ISBN 978-985-881-852-4 (ч. 1)
ISBN 978-985-881-851-7

© БГУ, 2025

УДК 004(06)
ББК 32.97я431

Редакционная коллегия:
академик НАН Беларуси, доктор технических наук,
профессор *С. В. Абламейко* (гл. ред.);
доктор педагогических наук,
профессор *В. В. Казаченок* (зам. гл. ред.);
кандидат физико-математических наук, доцент *Н. М. Дмитрук*;
член-корреспондент НАН Беларуси,
доктор технических наук, профессор *А. В. Тузиков*;
доктор физико-математических наук, профессор *А. Ю. Харин*

Рецензенты:
академик НАН Беларуси,
доктор физико-математических наук, профессор *Ю. С. Харин*;
член-корреспондент НАН Беларуси,
доктор технических наук, профессор *А. В. Тузиков*;
доктор физико-математических наук, профессор *Н. А. Лиходед*

Информационные системы и технологии = Information Systems and Technologies : материалы XI Междунар. науч. конгр. по информатике (CSIST-2025), Респ. Беларусь, Минск, 29–31 окт. 2025 г. В 2 ч. Ч. 1 / Белорус. гос. ун-т ; редкол.: С. В. Абламейко (гл. ред.) [и др]. – Минск : БГУ, 2025. – 1 электрон. опт. диск (CD-ROM). – Текст : электронный. – ISBN 978-985-881-852-4.

Представлены материалы международного научного конгресса, организованного Белорусским государственным университетом и Объединенным институтом проблем информатики НАН Беларуси.

Рассмотрены вопросы информационной и компьютерной безопасности, биоинформатики и приложений, интеллектуального и статистического анализа данных и принятия решений, оптимизации и надежности систем, параллельной и распределенной обработки данных.

Минимальные системные требования:
PC, Pentium 4 или выше; RAM 1 Гб; Windows XP/7/10;
Adobe Acrobat.
Оригинал-макет подготовлен в программе Microsoft Word
На русском и английском языках
В авторской редакции
Ответственный за выпуск *С. В. Шолтанюк*
Подписано к использованию 27.10.2025. Объем 10 МБ
Белорусский государственный университет.
Управление редакционно-издательской работы.
Пр. Независимости, 4, 220030, Минск.
Телефон: (017) 259-72-40.
e-mail: urir@bsu.by
<http://elib.bsu.by/>

СОДЕРЖАНИЕ

РАЗДЕЛ I

ИНФОРМАЦИОННАЯ И КОМПЬЮТЕРНАЯ БЕЗОПАСНОСТЬ

<i>Агиевич С. В.</i> О 2-транзитивности XS-схем.....	8
<i>Бережной И. Б., Садовский А. В.</i> Проактивная защита от программ-вымогателей в ОС семейства Linux.....	14
<i>Воробьев С. Ю., Ханчевский Е. А.</i> Внедрение системы менеджмента информационной безопасности в деятельность предприятия энергетической отрасли.....	25
<i>Гайдук А. Н.</i> Уточнение теста длинных серий при тестировании выходных последовательностей генераторов случайных чисел.....	34
<i>Дасько Ф. Б.</i> Автоматический анализ криптографических протоколов: построение запросов в ProVerif.....	39
<i>Дригель Р. К., Вакарь Е. И., Нестеренков С. Н.</i> Криптографические методы защиты данных в 5G-сетях: обеспечение безопасности и конфиденциальности в высокоскоростных сетях нового поколения	49
<i>Клиндухов Я. А.</i> Оценка степени разрушения статистики естественного языка при обезличивании персональных данных на основе метода изменения состава или семантики.....	55
<i>Кочин В. П., Шанцов А. В.</i> Математическая модель обеспечения информационной безопасности гетерогенной платформы виртуализации	61
<i>Матвеев Г. В., Осиневская А. А.</i> О модульном разделении секрета в специальной линейной группе	69
<i>Мицкевич М. Н.</i> Портитрование криптографических библиотек с помощью SWIG.....	75
<i>Палуха В. Ю., Мальцев М. В., Харин Ю. С., Пириштук И. К.</i> Об актуальных задачах оценки надежности криптографических генераторов	83
<i>Саевич С. Г., Петроченко И. О., Мателенок А. П.</i> Анализ аномального трафика в локальной сети при помощи протокола SNMP.....	93
<i>Соболь А. М., Кочин В. П.</i> Нормализация событий информационной безопасности операционных систем с использованием MITRE ATT&CK.....	104
<i>Третьяков И. А., Рушечников Я. И., Данилов В. В.</i> Подавление ПЭМИН в видеоинтерфейсах посредством высокочастотной ШИМ помехи	111
<i>Черезов Д. Д., Каратаева П. М., Хватов Д. А., Савкин М. Д.</i> SQL-инъекции как угроза информационной безопасности: специфика PostgreSQL и OWASP Top-10 ...	115

РАЗДЕЛ II

БИОИНФОРМАТИКА И ПРИЛОЖЕНИЯ

<i>Гончар А. В., Фурс К. В., Тузиков А. В., Андрианов А. М.</i> De novo генерация потенциальных ингибиторов белка KasA <i>Mycobacterium tuberculosis</i>	124
<i>Гузова Е. В., Гринев В. В.</i> Идентификация неканонических полипептидов лейкозных клеток человека, основанная на к-мерном анализе данных RNA-seq.....	132
<i>Есенбаева Г. Р., Исмаилова А. А., Кадиркулов К. К.</i> Использование методов глубокого обучения в интерпретации результатов бактериологических исследований.....	140
<i>Монтик Н. С., Козик И. Д., Сухаревич Д. С.</i> Применение глубокого обучения для классификации стадий сна с использованием электроэнцефалограммы.....	150
<i>Палто Е. С.</i> Разработка и сравнительный анализ моделей машинного обучения для прогнозирования риска инфаркта миокарда на основе данных массового медицинского опроса BRFSS 2022 UPDATE	158
<i>Перова И. А., Вычугжанин М. В.</i> Обработка мультимодальных данных биомедицинского сигнала с использованием метрики Хаусдорфа	167
<i>Преображенский А. П., Ряснянская В. Н., Буракова А. А., Чопоров О. Н.</i> Разработка интеллектуальной системы для анализа медицинских показателей пациентов	174
<i>Сарнацкий Д. Д., Ильюшенков И. Н., Боева Н. А., Воронюк Л. С., Яцков Н. Н., Скакун В. В., Гринев В. В.</i> Разработка R-пакета huGSVcalleR, предназначенного для идентификации и аннотации сайтов однонуклеотидного полиморфизма в геноме человека	181
<i>Трусов И. С., Гринев В. В.</i> Биоинформатический подход на основе метамоделей для улучшения прогнозирования иммуногенности опухолевых эпитопов	190
<i>Фурс К. В., Лайков Я. В., Тузиков А. В., Андрианов А. М.</i> Оценочные функции молекулярного докинга: тестирование и сравнительный анализ	199
<i>Хазиев Г. А., Зверков О. А., Рубанов Л. И., Селиверстов А. В.</i> Поиск петли в несовершенном палиндроме	204

РАЗДЕЛ III

ИНТЕЛЛЕКТУАЛЬНЫЙ И СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ, ПРИНЯТИЕ РЕШЕНИЙ

<i>Абрамович М. С., Гарматная Л. В., Дорофеев Г. С.</i> Прогнозирование эффективности кардиопротективной терапии у больных раком молочной железы	210
<i>Баркетов М. С.</i> Определение зависимости двух случайных величин на основе статистической информации	216
<i>Буданов К. М.</i> Анализ гендерных различий параметров ЭКГ на основе метода многоуровневой амплитудной дискриминации.....	220

<i>Евмененко С. Ю., Малинковский Ю. В.</i> Инвариантность стационарного распределения открытой сети обслуживания с экспоненциальным ограничением на время пребывания.....	227
<i>Иванова М. Д.</i> Распознавание именованных сущностей категории «Профессия» в русскоязычных текстах	233
<i>Лобач В. И., Савицкая В. В.</i> Прогнозирование цен опционов на основе модели Блэка-Шоулса.....	241
<i>Лобач С. В., Остапук Е. В.</i> Прогнозирование доходности финансовых активов на основе модели Васичека	251
<i>Малюгин В. И.</i> Анализ и прогнозирование по данным реального времени с использованием моделей смешанной частоты	260
<i>Ольков С. Г., Олькова О. А.</i> Методика точного измерения температуры точки росы с помощью совокупности линейных уравнений.....	266
<i>Пархименко В. А., Быков А. А.</i> Объем информации в матрицах прямых затрат разной размерности в модели Леонтьева «затраты – выпуск»	274
<i>Сафиуллин Т. Т.</i> Методология расчета высокочастотных суб-индексов ИПЦ с применением веб-скрейпинга и классификации данных	280
<i>Филиппов М. А.</i> Теоретическое исследование и экспериментальная проверка алгоритмов распознавания, основанных на мере прецедентности	289
<i>Цаплин С. Т.</i> Прогнозирование социально-экономических показателей Российской Федерации на основе гибридного подхода.....	299
<i>Шибалко С. А., Харин Ю. С.</i> О прогнозировании многомерных двоичных временных рядов	308

РАЗДЕЛ IV

ОПТИМИЗАЦИЯ И НАДЕЖНОСТЬ СИСТЕМ

<i>Акулич Е. С.</i> Применение методов оптимизации для коррекции доз инсулина для пациентов с сахарным диабетом 1 типа	318
<i>Альсевич В. В.</i> Исследование зависимости между показателями сравнительной статистики теории фирмы	326
<i>Gorokhovich V. V.</i> On characteristic properties of faces of convex sets	334
<i>Готовец М. А., Дмитрук Н. М.</i> Субоптимальные обратные связи в линейных задачах оптимального управления с возмущениями	338
<i>Дудин А. Н., Дудин С. А.</i> Система обслуживания с резервированием приборов для повторных запросов	345
<i>Дудина О. С., Дудин А. Н.</i> Анализ системы обслуживания с процессом поступления, зависящим от качества функционирования	354
<i>Калинин А. И., Лавринович Л. И.</i> Асимптотика решения задачи минимизации энергетических затрат при управлении трехтемповой сингулярно возмущенной линейной системой.....	362

<i>Kostyukova O. I., Tchemisova T. V.</i> On facial structure of the copositive and completely positive cones	368
<i>Палуха В. Ю., Пардаев А. А.</i> Оценивание параметров нелинейных регистров сдвига с обратной связью	378

РАЗДЕЛ V

ПАРАЛЛЕЛЬНАЯ И РАСПРЕДЕЛЁННАЯ ОБРАБОТКА ДАННЫХ

<i>Азаров Е. А., Марков А. Н., Барсукевич С. Н.</i> Сравнение многопоточных моделей в C++, Go и Rust в задачах параллельной обработки данных	386
<i>Баранчик М. А., Крупенич Е. Г., Мигалевич С. А.</i> Использование генеративных моделей ИИ	393
<i>Баханович С. В., Лиходед Н. А.</i> Распределенный блочный алгоритм вычисления поперечной и продольной скоростей течения крови в сосуде	399
<i>Блинов И. Н., Лядинский Т. А.</i> Асинхронность и масштабируемость: обработка данных в Android-приложениях	409
<i>Климович И. Д., Недзьведь О. В.</i> Анализ эффективности цветовых пространств CIELAB и RGB для задач сегментации изображений водорослей	416
<i>Лиходед Н. А., Малиев Э. Е.</i> OpenMP-реализация блочного алгоритма вычисления поперечной и продольной скоростей течения крови в сосуде	422
<i>Лукашевич М. М., Венгеренко В. В., Воронов А. А.</i> Сегментация изображений головы человека на основе нейронных сетей	429
<i>Лукашевич М. М., Клицунова Е.</i> Веб-приложение для рекомендации методов балансировки данных	436
<i>Сьянов Д. А.</i> Архитектура вероятностного трансформера для моделирования задержек в многоэтапных процессах	443

РАЗДЕЛ I

**ИНФОРМАЦИОННАЯ И КОМПЬЮТЕРНАЯ
БЕЗОПАСНОСТЬ**

О 2-ТРАНЗИТИВНОСТИ XS-СХЕМ

С. В. Агиевич

*НИИ прикладных проблем математики и информатики,
Белорусский государственный университет,
Минск, Беларусь, agievich@bsu.by*

XS-схемы описывают тактовые подстановки широкого класса блочных шифров. Для защиты от атаки по невозможным дифференциалам лежащая в основе шифра схема должна быть 2-транзитивной, а число тактов не меньше индекса 2-транзитивности. В работе вводится необходимое и почти достаточное условие 2-транзитивности XS-схемы, улучшается оценка сверху для индекса 2-транзитивности.

Ключевые слова: блочный шифр; тактовая подстановка; S-блок; схема; 2-транзитивность.

ON 2-TRANSITIVITY OF XS-CIRCUITS

S. V. Agievich

*Research Institute for Applied Problems of Mathematics and Informatics,
Belarusian State University,
Minsk, Belarus, agievich@bsu.by*

XS-circuits describe round permutations of a wide class of block ciphers. To protect a cipher against an impossible differential attack, the underlying circuit must be 2-transitive and the number of rounds must be at least the index of 2-transitivity. We provide a necessary and almost sufficient condition for 2-transitivity of an XS-circuit. We also improve the upper bound on the index of 2-transitivity.

Keywords: block cipher; round permutation; S-box; circuit; 2-transitivity.

1. Введение

XS-схемы, введенные в работе [1], описывают тактовые преобразования широкого класса блочных шифров. Схема размерности n задается тройкой (a, B, c) , в которой a — вектор-столбец размерности n , B — квадратная матрица порядка n , c — вектор-строка размерности n . Координаты векторов и элементы матрицы лежат в поле $\mathbb{F}_2$. Схема (a, B, c) уточняется дополнительными параметрами: полем $\mathbb{F}$ — некоторым расширением $\mathbb{F}_2$

и S -блоком S — подстановкой на $\mathbb{F}$. Уточненная схема задает следующее преобразование вектор-строк $x \in \mathbb{F}^n$:

$$(a, B, c)[S](x) = xB + S(xa)c.$$

Будем рассматривать обратимые схемы — те, для которых данное преобразование обратимо при любом выборе $\mathbb{F}$ и S . Условия обратимости даны в работе [1]. Обратные преобразования обратимой схемы (a, B, c) также описываются XS-схемой. Будем обозначать ее $(a, B, c)^{-1}$. В блочном шифре преобразования схемы (a, B, c) с разными (вообще говоря) ключезависимыми S -блоками объединяются в композиционные каскады, называясь при этом тактовыми преобразованиями или просто тактами. На вход каскада в качестве x подается открытый текст, с выхода снимается шифртекст. Преобразования t -тактового каскада с S -блоками $S_1, S_2, \dots, S_t$ обозначаются через $(a, B, c)^t[S_1, S_2, \dots, S_t]$.

Для обеспечения базовых гарантий криптографической стойкости схема (a, B, c) должна быть регулярной: во-первых, обратимой и, во-вторых, давать обратимые матрицы $A = \begin{pmatrix} a & Ba & \dots & B^{n-1}a \end{pmatrix}$ и $C = \begin{pmatrix} (cB^{n-1})^T & \dots & (cB)^T & c^T \end{pmatrix}^T$.

Расширенные гарантии связаны с защитой от тех или иных криптоаналитических атак. Так для защиты от атаки по невозможным дифференциалам (Impossible Differential Attack, IDA, см. [3]) схема должна быть 2-транзитивной, причем индекс 2-транзитивности должен быть по возможности невелик (определения будут даны позже).

Исследование 2-транзитивности было начато нами в работе [1]: было найдено необходимое условие 2-транзитивности (*сильная регулярность*) и получены оценки сверху для индекса 2-транзитивности при выполнении данного условия.

В настоящей работе мы усиливаем результаты [1]. Во-первых, предлагаем более слабое необходимое условие 2-транзитивности: *плотность профиля* схемы. Во-вторых, показываем, что для регулярной схемы при необременительных ограничениях на ее размерности плотность профиля является также достаточным условием 2-транзитивности. В-третьих, усиливаем оценки сверху для индекса 2-транзитивности, задействуя новую характеристику схем: *порог полного ранга*.

Будем использовать канонические формы регулярных схем (a, B, c) . У первой формы $c = (0, \dots, 0, 1)$, у второй $a = (1, 0, \dots, 0)^T$. У обеих форм одна и та же матрица B , она представляет собой клетку Фробениуса.

Для натурального M через $M^{[n]}$ будем обозначать n -ую факториальную степень $M : M^{[n]} = M(M-1)\dots(M-n+1)$.

2. Результаты

Определение 1. Профиль схемы (a, B, c) — это двоичная последовательность $\sigma = (\sigma_t)$, в которой

$$\sigma_t = cB^{t-1}a, \quad t = 1, 2, \dots$$

Профиль является линейной рекуррентной последовательностью (л.р.п.), порядок которой совпадает с размерностью (a, B, c) , а характеристический многочлен — с характеристическим многочленом B . Пусть схема имеет размерность n , B — клетка Фробениуса и $b = (b_1, b_2, \dots, b_n)^T$ — последний столбец B . Тогда

$$\sigma_t = \sum_{i=1}^n b_{n+1-i} \sigma_{t-i}, \quad t = n+1, n+2, \dots$$

при начальных условиях $(\sigma_1, \dots, \sigma_n) = (ca, cBa, \dots, cB^{n-1}a) = cA$.

Начальные условия не могут быть нулевыми, если (a, B, c) — регулярная схема, т. е. A и C обратимы. Действительно, в силу обратимости A противное означает, что $c = 0$. Но тогда $C = 0$, противоречие.

Особенно просто начальные условия выглядят, когда (a, B, c) записана во второй канонической форме, т.е. $a = (1, 0, \dots, 0)^T$. В этом случае A — единичная матрица и $cA = c$.

Важно, что профиль не меняется при переходе от (a, B, c) к подобной схеме $(P^{-1}a, P^{-1}BP, cP)$, где P — обратимая матрица порядка n .

Более того, регулярная схема (a, B, c) размерности n восстанавливается с точностью до подобия по начальному $2n$ -отрезку своего профиля σ .

Определение 2. Для последовательности $\sigma = (\sigma_t)$ обозначим $\text{supp}(\sigma) = \{r \in \mathbb{N} : \sigma_r \neq 0\}$ и пусть $\langle \text{supp}(\sigma) \rangle$ — множество, составленное из всевозможных сумм элементов $\text{supp}(\sigma)$. Последовательность σ называется *плотной* (dense), если найдется неотрицательное целое t_0 такое, что всякое $t \geq t_0$ лежит в $\langle \text{supp}(\sigma) \rangle$.

Лемма. Последовательность σ плотна тогда и только тогда, когда найдутся $r_1, r_2, \dots, r_k \in \text{supp}(\sigma)$ такие, что $\gcd(r_1, r_2, \dots, r_k) = 1$.

Доказательство. Если подходящий набор $(r_1, r_2, \dots, r_k)$ существует, то найдется конечное $t_0 = t_0(r_1, r_2, \dots, r_k)$ такое, что любое $t \geq t_0$ можно представить в виде суммы элементов набора. Данный факт доказан в рамках решения задачи Фробениуса (см. напр. [2]). Факт означает, что σ является плотной.

Наоборот, если подходящий набор не существует, то ненулевые элементы σ либо расположены в позициях, кратных некоторому $d > 1$, либо вообще отсутствуют. В обоих случаях σ не является плотной. Действительно, в первом случае представление t в виде суммы нельзя реализовать для t не кратных d , а во втором — ни для одного t . $\square$

Лемма означает, что если ненулевая последовательность σ не является плотной, то она разрежена с шагом $d > 1$: ненулевые элементы отстоят друг от друга на величины, кратные d . Термин «плотная» выбран нами как альтернатива «разреженная».

Тот факт, что профиль представляет собой линейную рекуррентную последовательность, упрощает проверку плотности.

Теорема 1. Пусть (a, B, c) — регулярная схема размерности n , в которой B — клетка Фробениуса, и $(b_1, b_2, \dots, b_n)^T$ — ее последний столбец. Пусть $r_1, r_2, \dots, r_k$ — номера единиц в векторах cA и $(b_n, b_{n-1}, \dots, b_1)$ при нумерации координат слева направо от 1. Профиль σ схемы (a, B, c) плотен тогда и только тогда, когда $\gcd(r_1, r_2, \dots, r_k) = 1$.

Если $c = (0, \dots, 0, 1)$, т.е. (a, B, c) записана в первой канонической форме, и $a = (a_1, a_2, \dots, a_n)^T$, то вектор cA может быть заменен на $(a_n, a_{n-1}, \dots, a_1)$.

Теорема 2. Профиль регулярной схемы (a, B, c) плотен тогда и только тогда, когда плотен профиль обратной схемы $(a, B, c)^{-1}$.

Для натуральных t и $\tau_1, \tau_2, \dots, \tau_k \leq t$ введем в рассмотрение матрицу

$$C(t; \tau_1, \tau_2, \dots, \tau_k) = \begin{pmatrix} cB^{t-\tau_1} \\ cB^{t-\tau_2} \\ \dots \\ cB^{t-\tau_k} \end{pmatrix}.$$

Определение 3. Пусть (a, B, c) — регулярная схема размерности n , σ — ее профиль. Порог полного ранга схемы — это минимальное t , для которого найдутся натуральные $t_1, t_2, \dots, t_n \leq t$ такие, что $t_i \in \langle \text{supp}(\sigma) \rangle$ и матрица $C(t; t_1, t_2, \dots, t_n)$ обратима.

Порог обозначается через $\text{frt}(a, B, c)$, от англ. “full-rank threshold”. Порог полагается равным ∞ , если подходящий набор $t_1, t_2, \dots, t_n$ не существует ни для одного конечного t .

Обратим внимание, что порог не меняется при переходе от (a, B, c) к подобной схеме $(P^{-1}a, P^{-1}BP, cP)$. Кроме этого, в определении порога максимальное из чисел $t_1, t_2, \dots, t_n$ обязательно совпадает с t .

Теорема 3. Пусть (a, B, c) — регулярная схема и σ — ее профиль. Порог $\text{frt}(a, B, c)$ конечен тогда и только тогда, когда профиль σ плотен.

Определение 4. Схема (a, B, c) размерности n над полем $\mathbb{F} = \mathbb{F}_{2^m}$ называется 2-транзитивной, если найдется натуральное t такое, что для любых двух пар (x, x') и (y, y') различных векторов $\mathbb{F}^n$ выполняется

$$\begin{aligned} y &= (a, B, c)^t [S_1, \dots, S_t](x), \\ y' &= (a, B, c)^t [S_1, \dots, S_t](x') \end{aligned}$$

при подходящем выборе $S_1, \dots, S_t \in S(\mathbb{F})$.

Минимальное t , при котором выполняется данное условие, называется индексом 2-транзитивности.

Теорема 4. Если схема (a, B, c) 2-транзитивна, то ее профиль плотен.

Теорема допускает частичное обращение.

Теорема 5. Пусть (a, B, c) — регулярная схема размерности n . Если профиль (a, B, c) плотен и $(2^m - 1)^{[n]} \geq 2^{mn} / 2$, то (a, B, c) 2-транзитивна с индексом, не превосходящем $2n + \text{frt}(a, B, c) + \text{frt}(a, B, c)^{-1}$.

Пример. В [4] предложена схема FourCell, которая задается следующей матрицей:

$$\begin{pmatrix} B & a \\ c & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Профиль схемы плотен, $\text{frt}(a, B, c) = 12$, $\text{frt}(a, B, c)^{-1} = 3$ и, следовательно, схема 2-транзитивна с индексом $\leq 2 \cdot 4 + 12 + 3 = 23$.

При этом схема не является сильно регулярной и обосновать ее 2-транзитивность на основании результатов работы [1] не удастся.

Библиографические ссылки

1. *Agievich S.* XS-circuits in block ciphers // *Mat. Vopr. Kriptogr.* 2019. Vol. 10, iss. 2. P. 7–30.
2. *Alfonsin J. L. R.* The Diophantine Frobenius Problem. Oxford University Press, 2005.
3. *Biham E., Biryukov A., Shamir A.* Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials // *Advances in Cryptology – EUROCRYPT’99 : International Conference on the Theory and Application of Cryptographic Techniques, Prague, Czech Republic, May 2–6, 1999, Proceedings / ed.: J. Stern. Berlin, Heidelberg : Springer, 1999. P. 12–23. (Lecture Notes in Computer Science, Vol. 1592.)*
4. Cryptographic properties and application of a generalized unbalanced Feistel network structure / J. Choy [et al.] // *Cryptography and Communications.* 2011. Vol. 3, iss. 3. P. 141–164.

ПРОАКТИВНАЯ ЗАЩИТА ОТ ПРОГРАММ-ВЫМОГАТЕЛЕЙ В ОС СЕМЕЙСТВА LINUX

И. Б. Бережной¹⁾, А. В. Садовский²⁾

¹⁾ НИИ прикладных проблем математики и информатики,
Белорусский государственный университет,
Минск, Беларусь, BerazhnyIB@bsu.by

²⁾ Белорусский государственный университет,
Минск, Беларусь

В статье рассматривается вопрос защиты от программ-вымогателей (ransomware) в ОС семейства Linux. Представлена общая схема функционирования таких программ и рассмотрена роль источника случайности в этой схеме. Предложен способ проактивной защиты от программ-вымогателей на основе контроля системных генераторов псевдослучайных чисел.

Ключевые слова: защита информации; программы-вымогатели; псевдослучайные числа; модификация ядра ОС Linux.

PROACTIVE PROTECTION AGAINST RANSOMWARE IN LINUX OS

I. B. Berezhnoy^{a)}, A. V. Sadovskiy^{b)}

^{a)} Research Institute for Applied Problems of Mathematics and Informatics,
Belarusian State University,
Minsk, Belarus, BerazhnyIB@bsu.by

^{b)} Belarusian State University,
Minsk, Belarus

The article considers the issue of protection against ransomware in Linux family OS. The general scheme of functioning of such programs is presented and the role of the source of randomness in this scheme is considered. A method to proactive protection against ransomware based on control of system random number generators is proposed.

Keywords: information security; ransomware; pseudo-random numbers; Linux kernel modification.

1. Введение

Программа-вымогатель (англ. ransomware, от слов ransom — выкуп и software — программное обеспечение) — тип вредоносного программного обеспечения (ВПО), который используется в целях вымогательства путем зашифровывания данных, ценных для пользователя компьютерной

системы, и последующего требования выкупа за их восстановление. Концепция программы-вымогателя, шифрующей пользовательские файлы, была представлена еще в 1996 году на конференции IEEE Security & Privacy conference [1], однако наибольшее распространение данный тип ВПО получил после 2013 года в связи с развитием криптовалют как трудноотслеживаемых способов платежей [2]. По данным компании Chainalysis суммарный ежегодный объем платежей на криптовалютные кошельки вымогателей с 2020 года превышает 500 миллионов долларов США, а за 2024 год превысил 800 миллионов долларов США [3].

Программы-вымогатели разных семейств отличаются способами проникновения, особенностями горизонтального распространения и закрепления в системе, некоторым дополнительным функционалом. Для выявления работы программ-вымогателей на конкретном компьютере системы класса MDR (Managed Detection and Response) используют ряд правил детектирования, к наиболее распространенным из которых относятся [4]:

- активная работа с реестром;
- наличие в оперативной памяти вредоносного кода;
- запуск подозрительных сервисов и программ, особенно удаленным пользователем;
- доступ к подозрительным узлам и URL;
- получение дампов памяти, связанных с сервисом безопасности LSASS;
- добавление пользователей и повышение их привилегий;
- работа с сетевыми устройствами.

Следует отметить, что применение таких правил требует предварительной настройки на конкретный образец ВПО, ориентируется на вспомогательный функционал, который может отсутствовать, и мало эффективно в случае обфускации кода ВПО и использования в ВПО уязвимостей нулевого дня.

Основные рекомендации для борьбы с программами-вымогателями, предлагаемые на текущий момент, представляют собой только общие рекомендации по борьбе с ВПО произвольного типа с повышенным акцентом на необходимость регулярного резервного копирования информации [5]. Другими словами, в качестве общей стратегии вместо способов защиты предлагается рассматривать способы минимизации последствий.

2. Схема криптографического ядра ransomware

В качестве основной компоненты всех программ-вымогателей выступает блок шифрования пользовательских данных на псевдослучайных ключах шифрования, генерируемых в ВПО. Источником таких псевдослучайных значений могут выступать как самостоятельно разработанные

хакерами генераторы псевдослучайных чисел (ГПСЧ), так и ГПСЧ, реализованные в операционной системе.

В случае самостоятельной реализации ГПСЧ в программе-вымогателе исходный код ВПО часто содержит ошибки, приводящие к понижению энтропии начальных значений ГПСЧ и, как результат, к возможности восстановления зашифрованных данных [6].

Поэтому у самых эффективных семейств программ-вымогателей криптографические ядра, непосредственно отвечающие за шифрование файлов, соответствуют следующей схеме [7]:

- 1) злоумышленник генерирует пару открытого и личного ключей для асимметричного шифрования и загружает в ВПО (криптер) открытый ключ, после чего криптер некоторым образом доставляется на компьютер-жертву;

- 2) при запуске криптер осуществляет поиск файлов, предположительно представляющих ценность для владельца компьютера, по определенным признакам, чаще всего — по расширению имен файлов;

- 3) для каждого найденного файла с помощью криптографически сильного ГПСЧ, реализованного в операционной системе, генерируется ключ симметричного шифрования, затем с помощью этого ключа файл зашифровывается (либо полностью, либо частично — в случае большого объема файла), после чего к нему дописывается блок с использованным ключом, зашифрованный на открытом ключе злоумышленника;

- 4) жертва после оплаты выкупа получает программное обеспечение для расшифрования файлов, содержащее личный ключ злоумышленника (декриптер);

- 5) декриптер для каждого файла расшифровывает на личном ключе злоумышленника дописанный блок, извлекает из него ключ симметричного шифрования и с его помощью расшифровывает исходный файл.

Указанная схема обладает следующими преимуществами для злоумышленника:

- 1) ни при каких обстоятельствах закрытый ключ злоумышленника не передается жертве до оплаты выкупа;

- 2) злоумышленнику не требуется от жертвы никаких дополнительных данных, кроме идентификатора использованной ключевой пары;

- 3) жертве от злоумышленника требуется один файл фиксированного размера вне зависимости от размера зашифрованных данных;

- 4) при соблюдении известных требований на размер ключа восстановление личного ключа по открытому для жертвы невозможно;

- 5) при симметричном шифровании разных файлов использованные ключи не могут совпадать; как следствие, существенно затруднен криптоанализ алгоритма шифрования файлов либо процесс восстановления личного ключа.

Основное тонкое место описанной схемы действий криптографического ядра программы-вымогателя — получение данных от системного ГПСЧ. Во время работы ransomware случайный ключ размером не менее 16 байтов требуется для каждого подходящего файла. А так как количество таких файлов, подходящих по шаблоны поиска, обычно исчисляется тысячами, то, соответственно, за короткий промежуток времени с ГПСЧ потребуется большое количество псевдослучайных данных. При этом конкретный поток криптоера не сможет продолжить свою работу до получения псевдослучайного ключа шифрования.

Таким образом, если имеется возможность повлиять на работу системного ГПСЧ и затруднить/приостановить его работу, то функционирование потоков процесса ransomware будет заморожено до момента шифрования пользовательских данных. В случае, если дальнейшие действия позволят остановить процесс программы-вымогателя, то такой подход может позволить полностью предотвратить шифрование пользовательских файлов (или потерять только первых N файлов, зашифрованных к моменту срабатывания блокировки), то есть осуществить проактивную защиту.

3. Функционирование ГПСЧ в ОС семейства Linux

ГПСЧ в ОС Linux использует аппаратные события, обнаруженные ядром Linux, в качестве источников шума для подачи данных на детерминированный генератор псевдослучайных чисел. Начиная с версии 5.18.1 ядра Linux ГПСЧ функционирует согласно схеме, представленной на рис. 1 [8].

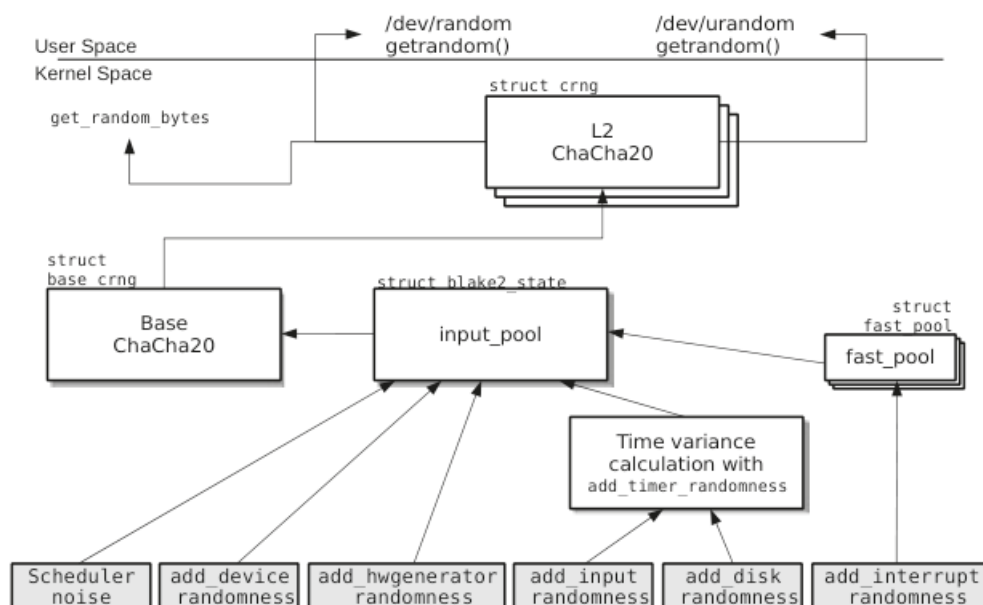


Рис. 1. Архитектура ГПСЧ в ОС Linux

ГПСЧ использует один пул энтропии — «входной пул» (англ. «input pool»). Его цель — сбор и накопление энтропии, предоставляемой различными источниками шума.

«Базовый» ChaCha20-генератор инициализируется на основе значения из «входного пула».

Набор вторичных ChaCha20-генераторов, по одному на каждый процессор, инициализируется «базовым» и предоставляет псевдослучайные данные через интерфейсы как режима пользователя (/dev/urandom, /dev/random, системный вызов getrandom()), так и режима ядра (функция get_random_bytes()).

Интерфейсы можно разделить на 2 класса:

- 1) блокирующие: /dev/random, getrandom();
- 2) неблокирующие: /dev/urandom, getrandom(GRND_INSECURE), get_random_bytes().

При выполнении запроса к блокирующему интерфейсу при недостаточной оценке энтропии input_pool интерфейс переходит в режим ожидания до накопления необходимой энтропии. Извлечение данных сопровождается снижением оценки энтропии input_pool.

Неблокирующие интерфейсы формируют запрошенное количество псевдослучайных данных независимо от того, достаточно ли энтропии у input_pool или ChaCha20-генератора. Они работают за счет периодической инициализации симметричного потокового алгоритма ChaCha20 данными из input_pool.

Вышеуказанные интерфейсы реализованы в файле random.c, являющемся частью ядра ОС Linux [9]. Соответственно, для влияния на данные интерфейсы требуется возможность внесения изменений в ядро ОС Linux.

4. Возможности модификации ядра Linux

ОС Linux — операционная система с открытым исходным кодом, а значит, предоставляет разработчикам полный контроль над системой, включая возможность модификации и пересборки любых компонентов ядра, в том числе и файла random.c.

Однако у данного подхода имеются следующие недостатки:

- 1) версия ядра должна соответствовать версии дистрибутива ОС;
- 2) трудоемкая процедура конфигурации и компиляции;
- 3) при внесении даже минимальных изменений в конкретном модуле требуется пересборка всего ядра.

Еще одним способом является использование *Linux Security API* — специального интерфейса, предназначенного для интеграции сторонних модулей безопасности, таких как SELinux, AppArmor и другие [10]. *Linux*

Security API позволяет модулям безопасности установить свои обработчики для системных вызовов, изучать контекст операции и централизованно принимать решение о её разрешении или запрете, работая при этом параллельно с системой разграничения доступа ядра ОС.

Однако, у *Linux Security API* есть пара важных ограничений:

- 1) модули безопасности не могут быть загружены динамически, являются частью ядра и требуют его пересборки;
- 2) в большинстве случаев одновременно может функционировать только один модуль безопасности.

При этом запрет на динамическую загрузку принципиальный: модуль безопасности должен быть частью ядра, чтобы обеспечивать безопасность постоянно, с момента загрузки.

Альтернативой без необходимости сборки собственного ядра является использование инструментов для отладки и анализа производительности функций ядра: FTrace и KProbe [11].

FTrace (Function Tracker) – это встроенная инфраструктура ядра Linux, предназначенная для отслеживания вызовов функций и трассировки их потоков выполнения. Изначально созданная для профилирования, она позволяет добавлять собственные хуки для экспорта и мониторинга функций ядра, что дает возможность отслеживать частоту и длительность вызовов функций, отображать графы вызовов, фильтровать интересующие функции по шаблонам и т. д.

Технология FTrace работает за счет поиска указанной сигнатуры функции в /proc/kallsyms. Обязательным условием для ее нахождения является то, что она должна быть экспортирована в коде во время сборки, либо добавлена в /proc/kallsyms напрямую.

Таким образом, FTrace предоставляет удобный функционал для внесения своих изменений в системный ГПСЧ или логирования связанных событий, таких как заполнение пула энтропии или вызов интерфейса получения псевдослучайных данных из пространства пользователя.

Главный недостаток FTrace – необходимость поддержки соответствующих FTrace опций конфигурирования при сборке ядра ОС. Тем не менее, ядра, используемые популярными дистрибутивами Linux, все необходимые опции в себе содержат, так как они не влияют на производительность и полезны при отладке. Однако полная универсальность у технологии отсутствует.

Технология KProbe, существующая и улучшающаяся с 2002 года, представляет собой специализированное API, в первую очередь предназначенное для отладки и трассирования ядра. В отличие от FTrace, эта технология позволяет устанавливать пред- и постобработчики не только на вход и возврат из функции, но и для *любой инструкции* в ядре.

Обработчики получают доступ к регистрам и могут их изменять. Таким образом, возможно получить как мониторинг, так и возможность влиять на дальнейший ход работы целевых функций ядра.

При этом, так как KProbe основывается на прерываниях и управляет регистрами процессора, возможности собственных обработчиков ограничены (нельзя выделить много памяти, заниматься вводом/выводом), а написание собственных обработчиков может иметь высокую техническую сложность (например, для получения аргументов перехватываемой функции надо их извлекать из заранее известных регистров и стека).

5. Экспериментальная проверка

Была проведена экспериментальная проверка возможности внесения изменений в работу ГПСЧ ОС Debian 12 посредством хука FTrace.

В качестве вносимого изменения рассматривалась дополнительная задержка для каждого вызова функции `get_random_bytes_user()`, которая используется в интерфейсах обращения к `/dev/random` из пространства пользователя.

Эксперимент проводится в несколько шагов.

1. Создание теста работы `/dev/random`, который представляет собой 1000 обращений на чтение 8MB из `/dev/random`. После чего выводится график и статистика времени выполнения.

2. Создание собственного обработчика, который будет вводить искусственную задержку в 50 мс. В качестве инструмента задержки использован `mdelay` (активное ожидание процессора).

3. Проведение теста до включения собственного обработчика.

4. Внедрение обработчика в ядро посредством команды `insmod`.

5. Проведение теста с активным обработчиком.

6. Сравнение результатов теста до и после включения обработчика.

На рис. 2 и 3 представлены результаты теста до включения обработчика. На рис. 4 и 5 представлены результаты теста после включения обработчика.

```
=== Benchmark Results ===
Samples: 1000
Chunk size: 8 MB
Average time: 17352.8 µs
Minimum time: 15216 µs
Maximum time: 28153 µs
Average throughput: 483.416 MB/s
```

Рис. 2. Метрики теста до встраивания обработчика

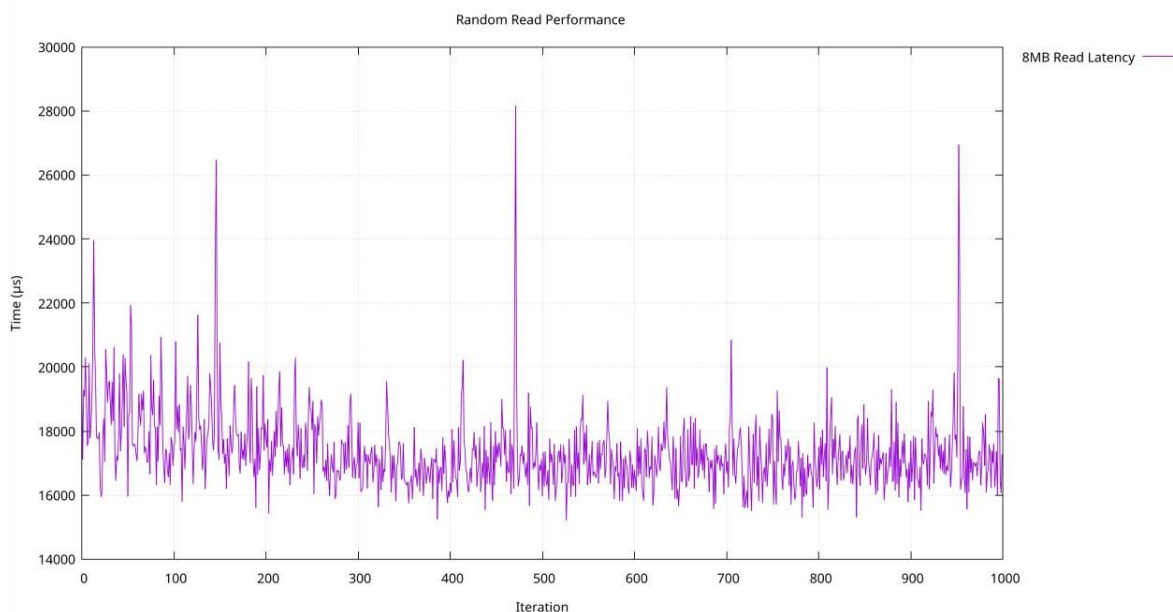


Рис. 3. Визуализация времени выполнения теста до встраивания обработчика

```

=== Benchmark Results ===
Samples: 1000
Chunk size: 8 MB
Average time: 65609.6 μs
Minimum time: 62418 μs
Maximum time: 73233 μs
Average throughput: 127.856 MB/s

```

Рис. 4. Метрики теста после встраивания обработчика

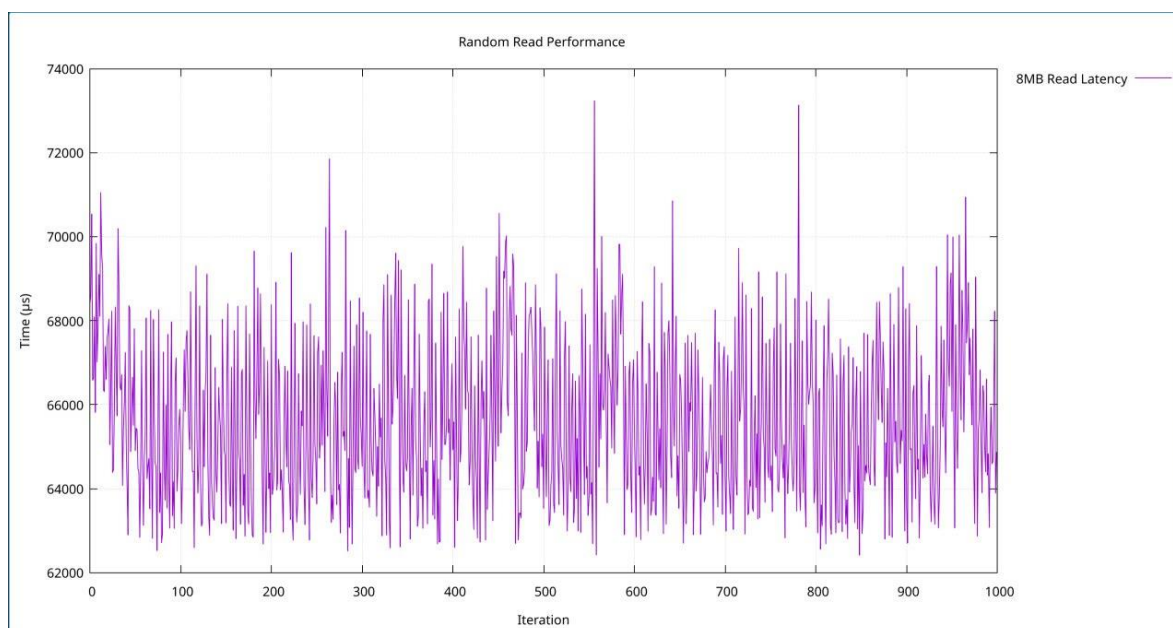


Рис. 5. Визуализация времени выполнения теста после встраивания обработчика

Среднее время выполнения чтения увеличилось с 17,353 мс до 65,610 мс, что примерно совпадает с ожидаемыми результатами (добавление 50 мс). Отличие может быть обусловлено тем, что `mdelay` использует примерную оценку прошедшего времени и, в зависимости от состояния аппаратуры, может выдавать ошибку в оценке.

Также был проведен аналогичный эксперимент по проверке возможности внесения изменений в работу ГПСЧ ОС Debian 12 посредством пред-обработчика KProbe с тем же функционалом (задержкой на 50 мс вызова функции `get_random_bytes_user()`). Представленный на рис. 6 и 7 результат эксперимента также подтвердил ожидаемое поведение: среднее время выполнения чтения из `/dev/random` увеличилось на 50 мс.

```
=== Benchmark Results ===
Samples: 1000
Chunk size: 8 MB
Average time: 65978.7 µs
Minimum time: 62278 µs
Maximum time: 75158 µs
Average throughput: 127.141 MB/s
```

Рис. 6. Метрики теста после встраивания пред-обработчика KProbe

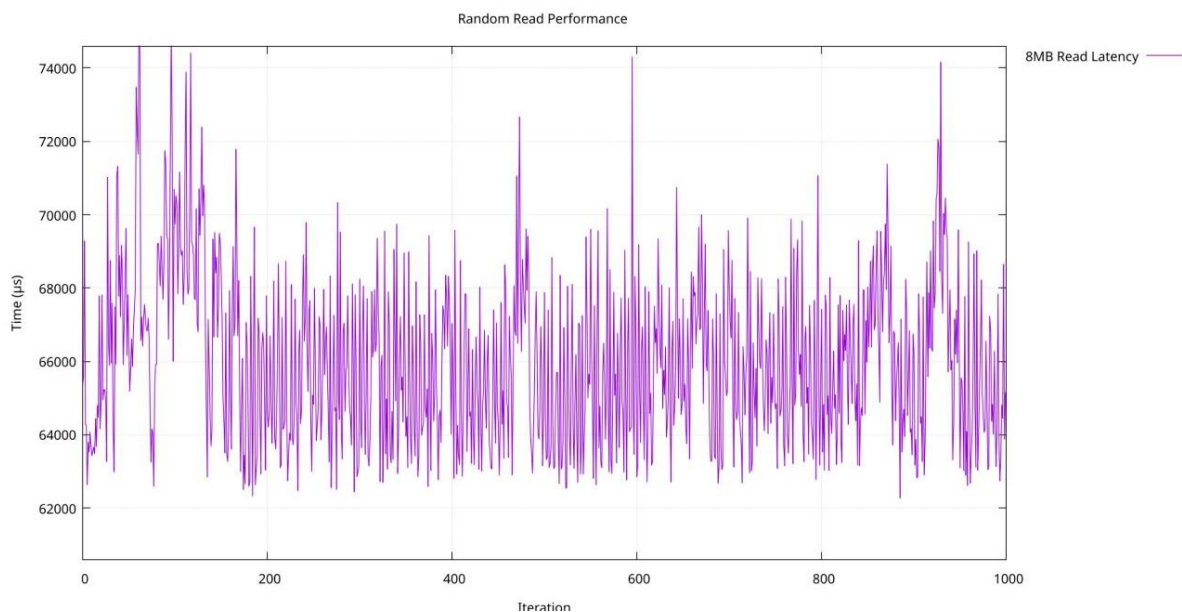


Рис. 7. Визуализация времени выполнения теста после встраивания пред-обработчика KProbe

Следует отметить, что как внедрение, так и выгрузка собственного обработчика из ядра системы требует прав суперпользователя, но при их

наличии легко автоматизируемые. Поэтому было реализовано консольное приложение с функциональностью диспетчера модулей с одновременной поддержкой обеих технологий FTrace и KProbe, который позволяет загружать и выгружать заранее подготовленные модули, тем самым управляя задержкой системного ГПСЧ (рис. 8).

```
=== Custom Kernel Module Manager ===
Detected modules in ../:
  1. kprobe_override      [LOADED]
  2. ftrace_hook_demo     [NOT LOADED]

Options:
  3. Load module
  4. Unload module
  5. Refresh list
  6. View dmesg
  7. Exit
Select option: █
```

Рис. 8. Интерфейс разработанного приложения

6. Заключение

В статье рассмотрен вопрос проактивной защиты от программ-вымогателей путем контроля системного ГПСЧ ОС Linux. Приведен обзор возможных подходов к реализации такого контроля как на основе полной пересборки ядра, так и путем изменения поведения ядра в уже собранном виде с помощью технологий FTrace и KProbe. Разработаны и экспериментально подтверждены методы внедрения задержек в работу ГПСЧ, что может повлечь замораживание работы программ-вымогателей до начала этапа шифрования пользовательских данных.

Библиографические ссылки

1. Young A., Yung M. Cryptovirology: extortion-based security threats and countermeasures // Proceedings 1996 IEEE Symposium on Security and Privacy. 1996. P. 129–140
2. Fruhlinger J. Ransomware explained: How it works and how to remove it. // CSO Online. URL: <https://www.csoonline.com/article/563507/what-is-ransomware-how-it-works-and-how-to-remove-it.html> (date of access: 09.09.2025).
3. 35% Year-over-Year Decrease in Ransomware Payments, Less than Half of Recorded Incidents Resulted in Victim Payments // Chainalysis. URL: <https://www.chainalysis.com/blog/crypto-crime-ransomware-victim-extortion-2025/> (date of access: 09.09.2025).

4. Managed detection and response in 2024 // Kaspersky Security Services. – URL: <https://securelist.com/kaspersky-managed-detection-and-response-report-2024/115635/> (date of access: 09.09.2025).
5. Ransomware protection: How to keep your data safe in 2025 // Kaspersky Resource Center. URL: <https://www.kaspersky.com/resource-center/threats/how-to-prevent-ransomware> (date of access: 09.09.2025).
6. *Nugroho Y.* Decrypting Encrypted files from Akira Ransomware (Linux/ESXI variant 2024) using a bunch of GPUs // TinyHack.com – URL: <https://tinyhack.com/2025/03/13/decrypting-encrypted-files-from-akira-ransomware-linux-esxi-variant-2024-using-a-bunch-of-gpus> (date of access: 09.09.2025).
7. Full source of the Conti Ransomware // Github. – URL: <https://github.com/gharty03/Conti-Ransomware> (date of access: 09.09.2025).
8. *Müller S.* Documentation and Analysis of the Linux Random Number Generator. Technical report // Bundesamt für Sicherheit in der Informationstechnik. URL: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/Studies/LinuxRNG/LinuxRNG_EN_V5_0.pdf (date of access: 09.09.2025).
9. Linux Kernel // Github. URL: <https://github.com/torvalds/linux/blob/master/drivers/char/random.c> (date of access: 09.09.2025).
10. Linux Security Modules: General Security Hooks for Linux // The kernel development community. URL: <https://docs.kernel.org/security/lsm.html> (date of access: 09.09.2025).
11. Перехват функций в ядре Linux с помощью ftrace // Хабр : сайт. URL: <https://habr.com/ru/articles/413241/> (дата обращения: 09.09.2025).

ВНЕДРЕНИЕ СИСТЕМЫ МЕНЕДЖМЕНТА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В ДЕЯТЕЛЬНОСТЬ ПРЕДПРИЯТИЯ ЭНЕРГЕТИЧЕСКОЙ ОТРАСЛИ

С. Ю. Воробьев, Е. А. Ханчевский

*Научно-исследовательское и проектно-изыскательское республиканское унитарное предприятие «Белэнергосетьпроект»,
Минск, Беларусь, s.varabyou@besp.by, zh@besp.by*

Цифровизация экономики ставит перед организациями энергетической отрасли задачи обеспечения надлежащего уровня информационной безопасности. В научной литературе отсутствуют исследования, посвященные внедрению системы менеджмента информационной безопасности в деятельность предприятий национальной энергетики. В статье описан процесс внедрения системы менеджмента информационной безопасности в деятельность предприятия, структурно входящего в состав ГПО «Белэнерго» Министерства энергетики Республики Беларусь.

Ключевые слова: энергетика; информационная безопасность; система менеджмента; СМИБ; ISO/IEC 27001.

IMPLEMENTATION OF AN INFORMATION SECURITY MANAGEMENT SYSTEM IN THE ACTIVITIES OF AN ENTERPRISE IN THE ENERGY INDUSTRY

S. Yu. Vorobyov, E. A. Khanchevsky

*Scientific research and design and survey republican unitary enterprise "Belenergosetproekt",
Minsk, Belarus, s.varabyou@besp.by, zh@besp.by*

The digitalization of the economy poses challenges to energy industry organizations to ensure an appropriate level of information security. There are no studies in scientific literature devoted to the implementation of an information security management system in the activities of national energy enterprises. The article describes the process of implementing an information security management system in the activities of an enterprise that is structurally part of the GPO Belenergo of the Ministry of Energy of the Republic of Belarus.

Keywords: energy; information security; management system; ISMS; ISO/IEC 27001.

1. Введение

В белорусской энергетике наравне с другими отраслями национальной экономики происходят активные процессы цифровизации, что ставит перед организациями энергетической отрасли актуальные задачи по

обеспечению информационной безопасности (далее – ИБ) и защите информации (далее – ЗИ).

Вопросам ИБ на объектах энергетики посвящены работы белорусских и российских авторов С. Ю. Воробьёва, Е. А. Ханчевского, А. И. Белоуса, И. А. Костомахи, И. Н. Колоска, Е. С. Коркиной, М. Г. Головенчик, Г. Г. Краско, Г. Г. Головенчик, В. А. Северин, Д. Хитрых [1–7]. Эффективность внедрения системы менеджмента ИБ (далее – СМИБ) в деятельность предприятий и организаций проанализирована в статьях А. В. Вилковой, А. А. Кайсарова, А. Т. Касымбека, П. А. Лончих, А. В. Суханова, М. Д. Хвистика, Л. А. Федоськиной, И. Ю. Шахалова [8–15].

Вместе с тем в настоящее время отсутствуют работы, посвященные исследованию эффективности от внедрения СМИБ в деятельность организаций энергетической отрасли.

Целью статьи является исследование проекта внедрения СМИБ в деятельность одного из предприятий, структурно входящего в ГПО «Белэнерго» Министерства энергетики Республики Беларусь (далее – Предприятие), анализ полученных от его реализации результатов, перспективы по дальнейшему укреплению состояния защищенности обрабатываемой информации в информационных системах (далее – ИС) Предприятия.

2. Основная часть

Набором общепризнанных международным сообществом требований и лучших практик, предъявляемых к СМИБ, является серия стандартов ISO/IEC 270xx, принятая Международной организацией по стандартизации ISO (International Organization for Standardization) [16]. После проведения мероприятий по переводу и терминологической адаптации Государственным комитетом по стандартизации Республики Беларусь данные технические нормативные правовые акты (далее – ТНПА) были введены в действие (по содержанию и смысловой нагрузке они полностью идентичны стандартам ISO). В национальной системе стандартизации действуют следующие из них:

- СТБ ISO/IEC 27000-2024 «Информационные технологии. Методы обеспечения безопасности. Общий обзор и словарь» – представляет собой обзор систем менеджмента информационной безопасности, которые являются предметом серии стандартов СМИБ, а также терминологическую базу связанных с ними терминов [17];

- СТБ ISO/IEC 27001-2024 «Информационная безопасность, кибербезопасность и защита конфиденциальности. Требования» (далее – СТБ 27001) – по сути является основным стандартом данной серии и

устанавливает требования к разработке, внедрению, поддержанию и постоянному улучшению СМИБ в контексте организации [18];

- СТБ ISO/IEC 27002-2024 «Информационная безопасность, кибербезопасность и защита конфиденциальности. Средства управления информационной безопасностью» (далее – СТБ 27002) предназначен для использования при разработке руководства по менеджменту ИБ в конкретных отраслях и организациях с учетом их специфики [19];

- СТБ ISO/IEC 27003-2014 «Информационные технологии. Методы обеспечения безопасности. Руководство по внедрению системы менеджмента информационной безопасности» (далее – СТБ 27003) описывает процессный подход по определению и разработке СМИБ от начала до фактического завершения проекта [20];

- СТБ ISO/IEC 27004-2014 «Информационные технологии. Методы обеспечения безопасности. Менеджмент информационной безопасности. Измерения» представляет собой руководство по разработке и применению мер измерения и проведению процесса измерения внедренной СМИБ (в том числе средств управления или их совокупности) [21];

- СТБ ISO/IEC 27005-2024 «Информационная безопасность, кибербезопасность и защита конфиденциальности. Руководство по менеджменту рисков информационной безопасности» (далее – СТБ 27005), по авторскому мнению, является не менее значимым ТНПА, чем СТБ 27001, так как описывает механизм реализации выполнения требований, касающихся действий по рассмотрению рисков ИБ и менеджмента рисков ИБ [22].

Внедрение СМИБ в деятельность Предприятия имело под собой следующие основания:

- укрепление деловой репутации на рынке как высоконадежной организации, применяющей в своей деятельности наилучшие общепризнанные практики по информационной безопасности;

- получение лицензии для осуществления деятельности в сфере технической и криптографической защиты информации по проектированию, созданию и аудиту систем информационной безопасности критически важных объектов информатизации (в соответствии с ч. 4.3 ст. Закона Республики Беларусь от 14.10.2022 № 213-З «О лицензировании» (далее – Закон о лицензировании) необходимо наличие действующего сертификата соответствия Национальной системы подтверждения соответствия Республики Беларусь на систему менеджмента информационной безопасности [23]).

Реализация проекта по внедрению СМИБ осуществлялась согласно рекомендаций СТБ 27002 и СТБ 27003: были созданы рабочая группа и план внедрения СМИБ (с определением этапов, сроков реализации проекта и ответственных лиц). Состав рабочей группы подбирался исходя из требований СТБ 27001 к организации, в которой планируется внедрение

СМИБ: контекст (основной вид деятельности), соответствие нормативным и договорным требованиям в сфере ИБ, подбор и обучения персонала, проведение мероприятий по снижению рисков ИБ (что, как, правило требует существенных материальных затрат, например, на закупку антивирусного программного обеспечения или проведения мероприятий по сканирования сети). Таким образом в состав рабочей группы вошли руководители подразделений ИБ, информационных технологий, планово-экономического и бухгалтерии, правовой и кадровой работы. Возглавил рабочую группу первый заместитель директора – главный инженер, как непосредственный куратор основных производственных процессов Предприятия.

Рабочая группа начала свою деятельность с инвентаризации информационных активов Предприятия: информации, обрабатываемой в ИС и определения основных бизнес-процессов. Были подвергнуты ревизии локальные правовые акты (далее – ЛПА), регулирующие вопросы ИБ и ЗИ, на предмет их соответствия законодательным и нормативным требованиям. Результатом работы стали не только актуализация действующих локальных актов, но разработка новелл (например, предоставление возможности работнику работать дистанционно, в том числе с использованием собственного оборудования).

Необходимо обратить внимание, что СТБ 27001 не только не противоречит нормам национального законодательства, регулирующего вопросы ИБ и ЗИ, но прямо коррелирует с ним (согласно п. 5.31 Приложения СТБ 27001 правовые, законодательные, нормативные и договорные требования, относящиеся к информационной безопасности, и подход к выполнению этих требований в организациях должен быть идентифицирован, документирован и поддерживаться в актуальном состоянии).

При актуализации и разработке ЛПА в последние имплементировались требования правовых актов законодательства в данной сфере. Так, согласно Политике информационной безопасности роль представителя руководства Предприятия по вопросам информационной безопасности была возложена на первого заместителя директора – главного инженера, что коррелирует п. 3.15 Указа Президента Республики Беларусь от 14.02.2023 № 40 «О кибербезопасности» (далее – Указ № 40), в котором руководители организаций назначают одного из своих заместителей ответственным за организацию работ по кибербезопасности [24].

С целью повышения вовлечения всех работников Предприятия в реализацию проекта внедрения СМИБ и повышения осведомленности по вопросам ИБ и ЗИ проекты перерабатываемых ЛПА направлялись в структурные подразделения для ознакомления и получения обратной связи в виде замечаний и предложений (в случае наличия и актуальности последних в проекты ЛПА вносились соответствующие правки).

Была выявлена необходимость актуализации иных ЛПА Предприятия, непосредственно не входящих в документацию СМИБ: Положения о

коммерческой тайне, Политике о защите персональных данных, должностных инструкций некоторых работников.

Необходимо отметить, что ведомственные стандарты ГПО «Белэнерго» СТП 33243.01.216-16 «Подстанции электрические напряжением 35 кВ и выше. Нормы технологического проектирования» (далее – СТП 33243.01.216) [25] и СТП 33240.20.117-18 «Цифровые подстанции. Требования к проектированию» [26] в своей структуре содержат раздел, предъявляющий требования к ИБ при проектировании энергообъектов, и, соответственно, коррелируют с п. 5.8 Приложения А СТБ 27001, который предъявляет требования по интеграции ИБ в менеджмент проектов.

Внедрение СМИБ предъявляет требования к процессам, связанным с управлением людьми: подбору работников в соответствии с их профессиональными знаниями, умениями и навыками, надлежащими морально-этическими качествами, а также постоянному повышению осведомленности работающих в сфере ИБ и ЗИ. В соответствии со ст. 11 Декрета Президента Республики Беларусь от 15.12.2014 № 5 «Об усилении требований к руководящим кадрам и работникам организаций» стадия изучения кандидатов на вакантные должности Предприятия сопровождается запросами характеристики с предыдущего места работы и сведений из единого государственного банка данных о правонарушениях (в отношении кандидатов на руководящие должности) [27]. Требование о повышении квалификации работников подразделений ЗИ предусмотрены Указом № 40 (при этом работники центров кибербезопасности должны пройти обучение в республиканском унитарном предприятии «Национальный центр обмена трафиком» (далее – НЦОТ)) [24], Указом Президента Республики Беларусь от 16.04.2013 № 196 «О некоторых мерах по совершенствованию защиты информации» [28], приказом Оперативно-аналитического центра при Президенте Республики Беларусь от 20.02.2020 № 66 «О мерах по реализации Указа Президента Республики Беларусь от 9 декабря 2019 г. № 449» [29]. Указом Президента Республики Беларусь от 28.20.2021 № 422 «О мерах по совершенствованию защиты персональных данных» владельцы критически важных объектов информатизации и собственники (владельцы) ИС, в которых обрабатывается информация, распространение и (или) предоставление которой носят ограниченный характер, обеспечивают не реже одного раза в три года обучение работников ИБ в Национальном центре защиты персональных данных (далее – НЦЗПД) [30]. В процессе внедрения СМИБ обучение на базе НЦЗПД и НЦОТ прошли не только специалисты по ИБ, но также руководители и работники иных структурных подразделений Предприятия.

Согласно парадигме обработки рисков ИБ, принятой СТБ 27005 и СТБ 34.101.70-2016 «Информационные технологии. Методы и средства безопасности. Методика оценки рисков информационной безопасности в информационных системах» [31], были проведены мероприятия по

идентификации уязвимостей активов Предприятия, выявлению потенциальных угроз, которые могут воздействовать на последние, оценке возможных последствий в стоимостном эквиваленте, и, соответственно, принятию мер по снижению рисков ИБ (в том числе были пересмотрены права предоставления доступа к ИС Предприятия, а также запланировано к приобретению и внедрению специализированное программное обеспечение: система предотвращения утечек конфиденциальной информации DLP (Data Loss Prevention).

Финальной стадией проекта согласно методологии, изложенной СТБ 27003, было проведение внутреннего аудита СМИБ на соответствие требованиям СТБ 27001. По результатам выявленных замечаний определены ответственные лица и сроки устранения.

На момент написания статьи Предприятием подана заявка на сертификацию СМИБ на соответствие требованиям СТБ 27001 в организацию, аккредитованную на проведение аудита и сертификации в Национальной системе аккредитации Республики Беларусь.

3. Заключение

Организации энергетической отрасли Республики Беларусь при осуществлении деятельности обрабатывают в ИС информацию, представляющую самостоятельный и ценный ресурс, нарушение конфиденциальности которой может привести к совокупности неблагоприятных последствий [32, с. 59]. Вызовом времени для энергетической отрасли является не только осознание необходимости в проведении комплекса мероприятий организационного, правового и технического характера по ЗИ, но и вовлеченность в последние работников всех уровней, в первую очередь, руководителей высшего звена.

Анализ описанного процесса внедрения СМИБ на Предприятии позволяет сделать следующие выводы.

1. Следование процессному подходу позволило определить цели внедрения СМИБ в деятельность Предприятия, заинтересованные стороны, участников проекта и сроки реализации.

2. В реализацию проекта был вовлечен весь персонал Предприятия: от директора до специалиста, что способствовало выделению необходимых материальных ресурсов в достаточном количестве, своевременной и качественной актуализации локальной правовой базы, внесению изменений и дополнений в должностные инструкции работников, повышению осведомленности последних по вопросам ИБ и ЗИ.

3. Предусмотренное Указами Президента Республики Беларусь [24, 28] повышение квалификации в сфере технической и криптографической ЗИ на базе НЦОТ и НЦЗПД прошли не только специалисты,

обеспечивающие ЗИ на Предприятии, но также ряд работников от заместителя директора до руководителей структурных подразделений.

4. На Предприятии ведется серьезная работа подбору и изучению кандидатов на вакантные должности, включая изучение моральных и деловых качеств, законопослушности и гражданской самосознательности.

5. СТБ 27001 коррелирует СТП 33240.20.117-18 и СТП 33243.01.216 в части менеджмента ИБ при реализации проектов.

6. Парадигма применения в СМИБ законодательных, нормативных и договорных требований позволяет гибко и адаптивно выстраивать локальную правовую базу, регулирующую вопросы ИБ и ЗИ, в том числе путем имплементации требований национального законодательства.

7. Реализацией проекта внедрения СМИБ процесс поддержания надлежащего уровня ИБ в организации не заканчивается (методологией СМИБ предусмотрены систематические мероприятия по переоценке рисков ИБ и проведения внутренних аудитов на соответствие требованиям СТБ 27001).

8. Подтверждение соответствия СМИБ требованиям СТБ 27001 в Национальной системе соответствия Республики Беларусь позволит Предприятию получить лицензию на право выполнения работ по проектированию, созданию и аудиту систем информационной критически важных объектов информатизации, а также укрепить деловую репутацию путем подтвержденного сертификатом декларирования следования нормативным требованиям ИБ.

СМИБ не является панацеей от хищения коммерческой тайны, обрабатываемой в ИС организации, или кибератаки на АЭС, но проведение мероприятий согласно методологии, основанной на рискориентированном подходе, описанной в стандарте СТБ 27001 позволит повысить защищенность самого ценного ресурса XXI века: информации.

Библиографические ссылки

1. Воробьев С. Ю., Ханчевский Е. А. Кибератаки на критически важные объекты энергетики как источник угроз национальной безопасности // Энергетическая стратегия. 2024. Т. 102, № 6. С. 33–36.

2. Белоус А. И. Кибербезопасность объектов топливно-энергетического комплекса. Концепции, методы и средства обеспечения. Москва ; Вологда : Инфра-Инженерия, 2020.

3. Костомаха И. А. Методы удалённого проникновения злоумышленника в технологические сегменты сети предприятий электроэнергетики // Энергетик. 2024. № 8. С. 19–22.

4. Колосок И. Н., Коркина Е. С. Анализ кибербезопасности объектов энергетики с учётом механизма и кинетики нежелательных процессов // Энергетик. 2024. № 2. С. 3–8.

5. Головенчик М. Г., Краско Г. Г., Головенчик Г. Г. Проблемы кибербезопасности умных городов // Наука и инновации. 2020. № 12 (214). С. 51–57.

6. *Северин В. А.* Проблемы правового регулирования и методического обеспечения защиты информации в организациях ТЭК // Проблемы экономики и юридической практики. 2022. Т. 18, № 2. С. 96–103.
7. *Хитрых Д.* О цифровой трансформации энергетической отрасли // Энергетическая политика. 2021. № 10 (164). С. 76–89.
8. *Вилкова А. В., Литвишков В. М., Швырев Б. А.* Проблемы непрерывного обучения персонала информационной безопасности // Мир науки, культуры, образования. 2019. № 4(77). С. 29–31.
9. Внедрение международного стандарта ИСО/МЭК 27001 - основа управления информационной безопасностью предприятия / А. А. Кайсарова [и др.] // Вестник науки Южного Казахстана. 2018. № 4(4). С. 103–106.
10. *Касымбек А. Т.* Польза от внедрения международного стандарта ISO/IEC 27001 // Евразийское Научное Объединение. 2015. Т. 1, № 2(2). С. 52–53.
11. *Лонцих П. А., Сафонова О. М.* Методика создания и внедрение системы менеджмента информационной безопасности на промышленном предприятии // Системы. Методы. Технологии. 2020. № 4(48). С. 80–87.
12. *Суханов А. В., Смирнов А. С., Хитов С. Б.* Управление информационной безопасностью предприятий оборонно-промышленного комплекса в контексте стандарта ISO 27001:2013 // Научно-аналитический журнал «Вестник Санкт-Петербургского университета Государственной противопожарной службы МЧС России». 2017. № 1. С. 9–16.
13. *Хвистик М. Д., Серенков П. С.* Разработка и внедрение системы менеджмента информационной безопасности в соответствии с ISO/IEC 27001:2013 // Приборостроение-2021 : Материалы 14-й Международной научно-технической конференции, Минск, 17–19 ноября 2021 года / Минск : Белорусский национальный технический университет, 2021. С. 242–243.
14. *Федоськина Л. А., Рогачев Р. А.* Ключевые тренды в сертификации систем менеджмента информационной безопасности на основе международного стандарта ISO 27001 // Цифровая трансформация: тенденции и перспективы : Сборник трудов I Международной научно-практической конференции, Москва, 21 декабря 2022 года / Под ред. Н. Л. Кетовой и М. Т. Заргарян. М. : Общество с ограниченной ответственностью «Издательство «Мир науки», 2022. – С. 199–207.
15. *Шахалов И. Ю., Райкова Н. О.* К вопросу об интегрировании систем менеджмента качества и информационной безопасности // Правовая информатика. 2014. № 2. С. 20–26.
16. ISO: the International Organization for Standardization. URL: <https://www.iso.org/home.html> (date of access: 17.07.2025).
17. Информационные технологии. Методы обеспечения безопасности. Общий обзор и словарь : СТБ ISO/IEC 27000-2024. (Введ. 25.10.2024) / Минск : БелГИСС, 2024.
18. Информационная безопасность, кибербезопасность и защита конфиденциальности. Требования : СТБ ISO/IEC 27001-2024. (Введ. 25.10.2024) / Минск : БелГИСС, 2024.
19. Информационная безопасность, кибербезопасность и защита конфиденциальности. Средства управления информационной безопасностью : СТБ ISO/IEC 27002-2024. (Введ. 25.10.2024) / Минск : БелГИСС, 2024.
20. Информационные технологии. Методы обеспечения безопасности. Руководство по внедрению системы менеджмента информационной безопасности : СТБ ISO/IEC 27003-2014. (Введ. 14.08.2014) / Минск : Научн.-исслед. ин-т техн. защиты информации, 2014.
21. Информационные технологии. Методы обеспечения безопасности. Менеджмент информационной безопасности. Измерения : СТБ ISO/IEC 27004-2014. (Введ. 14.08.2014) / Минск : Научн.-исслед. ин-т техн. защиты информации, 2014.

22. Информационная безопасность, кибербезопасность и защита конфиденциальности. Руководство по менеджменту рисков информационной безопасности : СТБ ISO/IEC 27005-2024. (Введ. 25.10.2024) / Минск : БелГИСС, 2024.

23. О лицензировании : Закон Респ. Беларусь от 14 октября 2022 г., № 213-З // ЭТАЛОН. Законодательство Республики Беларусь / Нац. центр правовой информ. Респ. Беларусь. Минск, 2025.

24. О кибербезопасности : Указ Президента Респ. Беларусь, 14 февр. 2023 г., № 40 // ЭТАЛОН. Законодательство Республики Беларусь / Нац. центр правовой информ. Респ. Беларусь. Минск, 2025.

25. Подстанции электрические напряжением 35 кВ и выше. Нормы технологического проектирования : СТП 33243.01.216-16. (Введ. 29.01.2016 ; с отменой на территории РБ СТП 09110.01.2.104-07) / Минск : Научн.-исслед. и проектно-изыск. респ. унитарн. предпр. техн. «Белэнергосетьпроект», 2016.

26. Цифровые подстанции. Требования к проектированию : СТП 33240.20.117-18. (Введ. 16.06.2018) / Минск : Научн.-исслед. и проектно-изыск. респ. унитарн. предпр. техн. «Белэнергосетьпроект», 2018.

27. Об усилении требований к руководящим кадрам и работникам организаций : Декр. Президента Респ. Беларусь, 15 дек. 2014 г., № 5 // ЭТАЛОН. Законодательство Республики Беларусь / Нац. центр правовой информ. Респ. Беларусь. Минск, 2025.

28. О некоторых мерах по совершенствованию защиты информации : Указ Президента Респ. Беларусь, 16 апр. 2013 г., № 196 : в ред. Указа Президента Респ. Беларусь от 09.12.2019 г. // ЭТАЛОН. Законодательство Республики Беларусь / Нац. центр правовой информ. Респ. Беларусь. Минск, 2025.

29. О мерах по реализации Указа Президента Республики Беларусь от 9 декабря 2019 г. № 449. Республики Беларусь : Прик. Оперативно-аналитического центра при Президенте Республики Беларусь от 20 февраля 2020 г. № 66 // ЭТАЛОН. Законодательство Республики Беларусь / Нац. центр правовой информ. Респ. Беларусь. Минск, 2025.

30. О мерах по совершенствованию защиты персональных данных : Указ Президента Респ. Беларусь, 28 окт. 2021 г., № 422 : в ред. Указа Президента Респ. Беларусь от 24.03.2025 г. // ЭТАЛОН. Законодательство Республики Беларусь / Нац. центр правовой информ. Респ. Беларусь. Минск, 2025.

31. Информационные технологии. Методы и средства безопасности. Методика оценки рисков информационной безопасности в информационных системах : СТБ 34.101.70-2016. (Введ. 12.08.2016) / Минск : Научн.-исслед. ин-т техн. защиты информации, 2016.

32. *Воробьев С. Ю., Ханчевский Е. А.* Защита информации, представляющей коммерческую ценность, в организациях энергетической отрасли // Энергетическая стратегия. 2025. Т. 104, № 2. С. 59–62.

УТОЧНЕНИЕ ТЕСТА ДЛИННЫХ СЕРИЙ ПРИ ТЕСТИРОВАНИИ ВЫХОДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ГЕНЕРАТОРОВ СЛУЧАЙНЫХ ЧИСЕЛ

А. Н. Гайдук

*Белорусский государственный университет,
Минск, Беларусь, gaidukan@bsu.by*

Тест длинных серий проверяет, соответствует ли частота максимальной длины серии из единиц (или из нулей) в двоичном фрагменте длины L наблюдаемой последовательности теоретически ожидаемому значению π_i . При больших значениях L вычисление теоретических значений π_i требует значительных вычислительных ресурсов и в опубликованных работах эти значения рассчитаны с точностью до 4 знака после запятой. В данной работе уточнены данные значения для более точного расчета статистики теста.

Ключевые слова: случайные числа; статистический тест.

REVISITING THE TEST FOR THE LONGEST RUN OF ONES IN A BLOCK IN RANDOMNESS TEST SUITE

A. N. Gaiduk

*Belarusian State University,
Minsk, Belarus, gaidukan@bsu.by*

The purpose of the test for the longest run of ones in a block is to determine whether the length of the longest run of ones within the tested sequence is consistent with the length of the longest run of ones that would be expected in a random sequence. For large values of block lengths, the calculation of theoretical values π_i requires significant computational resources, and in published works these values are calculated with an accuracy of 4 decimal places. In this paper, theoretical values π_i are refined for a more accurate calculation of the test statistics.

Keywords: random number; statistical test.

1. Введение

Эффективность современной криптографической защиты информации основана на применении надежных генераторов случайных числовых последовательностей (ГСЧП). ГСЧП используются при создании

криптографических ключей, векторов инициализации, выработке начальных значений параметров алгоритмов, а также для решения других задач. Основным требованием, предъявляемым к используемому двоичному ГСЧП, является требование о неотличимости выходной последовательности такого генератора от последовательности независимых испытаний Бернулли с вероятностью успеха $1/2$. Для проверки качества ГСЧП используются батареи (наборы) статистических тестов, задача которых – выявление отклонений от «чистой случайности». На практике широко используются такие батареи тестов как NIST, Diehard, TestU01.

Целью данной работы является уточнение теоретических значений π_i теста длинных серий, входящего в батарею NIST [1] и в методику [2]. В опубликованных работах [1, 2] эти значения рассчитаны с точностью до 4-го знака после запятой, что приводит к возникновению значительной погрешности при вычислении результирующего P -значения теста [3].

2. Тест длинных серий

Приведем описание теста длинных серий согласно работе [2]. Тест длинных серий используется для выявления соответствия частоты максимальной серии из единиц (или нулей) в L -битном фрагменте теоретически ожидаемому значению [1].

Параметр теста: длина фрагмента L .

Входные данные: последовательность $X = (x_1, \dots, x_n)$.

Шаг 1. Разбить исходную последовательность X на $m = \lceil n / L \rceil$ фрагментов длины L бит.

$$X_i = (x_{(i-1)L+1}, \dots, x_{(i-1)L+L}), \quad i = \overline{1, m}.$$

Шаг 2. По L используя табл. 2, 4 определить разбиение множества длин серий на подмножества (классы) $D_0, \dots, D_K$, и теоретические вероятности $\pi_0, \dots, \pi_K$.

Шаг 3. Для $i = 1, \dots, m$ найти длину b_i максимальной серии из единиц в фрагменте X_i .

Шаг 4. Для $j = 1, \dots, K$ вычислить частоты попадания статистик b_i в класс D_j :

$$\nu_j = \sum_{i=1}^m \mathbf{1}\{b_i \in D_j\}.$$

Шаг 5. Вычислить статистику теста:

$$S = \sum_{j=1}^m \frac{(v_j - m\pi_j)^2}{m\pi_j}.$$

Шаг 6. Вычислить и вернуть P -значение теста:

$$P = 1 - F_{\chi_K^2}(S).$$

Условия применения. Должно выполняться условие: $m \cdot \min_{0 \leq i \leq K} \pi_i \geq 10$. Откуда следует, что минимальная длина последовательности для различных значений параметра должна удовлетворять значениям из табл. 1.

Таблица 1

Минимальные значения n теста длинных серий

L	8	128	512	1000	10000
n	427	12464	50688	100000	1490000

Примечание. Если верна гипотеза H_0 и в i -ом фрагменте встречается r единиц и $L - r$ нулей, то теоретические вероятности равны [1]:

$$P\{b_i \leq q\} = \sum_{r=0}^L C_L^r P\{b_i \leq q | r\} \frac{1}{2^L},$$

$$P\{b_i \leq q | r\} = \frac{1}{C_L^r} \sum_{j=0}^U (-1)^j C_{L-r+1}^j C_{L-j(q+1)}^{L-r},$$

где $0 \leq q \leq L$, $U = \min(L - r + 1, [r / (q + 1)])$, $0 \leq r \leq L$.

При больших значениях L вычисление вероятностей по данным формулам требует значительных вычислительных и временных ресурсов, поэтому в работах [1, 2] приведены предвычисленные значения для L из множества $\{8, 128, 512, 1000, 10000\}$ (табл. 2). Однако приведенные значения указаны лишь с точностью до 4 знака после запятой (табл. 3), что согласно работе [3] приводит к возникновению значительной погрешности при вычислениях статистики теста. В настоящей работе приведены более точные значения (табл. 4).

Таблица 2

Разбиение множества значений статистики b_i на классы согласно работам [1, 2]

L	8	128	512	1000	10000
K	3	5	5	5	6
D_0	≤ 1	≤ 4	≤ 6	≤ 7	≤ 10
D_1	2	5	7	8	11
D_2	3	6	8	9	12
D_3	≥ 4	7	9	10	13
D_4	—	8	10	11	14
D_5	—	≥ 9	≥ 11	≥ 12	15
D_6	—	—	—	—	≥ 16

Таблица 3

Вероятности попадания статистики в классы согласно работам [1, 2]

L	8	128	512	1000	10000
K	3	5	5	5	6
π_0	0,2148	0,1174	0,1170	0,1307	0,0882
π_1	0,3672	0,2430	0,2460	0,2437	0,2092
π_2	0,2305	0,2493	0,2523	0,2452	0,2483
π_3	0,1875	0,1752	0,1755	0,1714	0,1933
π_4	—	0,1027	0,1015	0,1002	0,1208
π_5	—	0,1124	0,1077	0,1088	0,0675
π_6	—	—	—	—	0,0727

Таблица 4

Более «точные» вероятности попадания статистики в классы

L	8	128	512	1000	10000
K	3	5	5	5	6
π_0	0,21484375	0,1174035788	0,1299334833	0,1388551909	0,0866323111
π_1	0,3671875	0,2429559593	0,2361223453	0,2369038167	0,2082006484
π_2	0,23046875	0,2493634832	0,2418343719	0,2387912400	0,2484185819
π_3	0,1875	0,1751770603	0,1729754256	0,1700180792	0,1939127867
π_4	—	0,1027010713	0,1032697631	0,1014361440	0,1214584851
π_5	—	0,1123988471	0,1158646108	0,1139955292	0,0680110893
π_6	—	—	—	—	0,0733660975

3. Заключение

В данной работе уточнены теоретические значения π_i теста длинных серий, входящего в батарею NIST [1] и в методику [2] для более точного вычисления статистики теста.

Библиографические ссылки

1. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications / L. E. Bassham [et al.]. NIST. 2010. Special Publication 800-22rev1a. URL: <https://www.nist.gov/publications/statistical-test-suite-random-and-pseudorandom-number-generators-cryptographic> (date of access: 08.09.2025).
2. Методика испытаний средств криптографической защиты информации на соответствие требованиям СТБ 34.101.27-2022. МИ.10127.10.01. URL: <https://www.oac.gov.by/public/content/files/files/met-10127-10-01.pdf> (дата обращения: 08.09.2025).
3. *Riesinger E., Fuß J.* Testing the Tests — Opportunities for Corrections and Improvements in NIST SP 800-22r1a and its Reference Code. URL: <https://eprint.iacr.org/2025/856> (date of access: 08.09.2025).

АВТОМАТИЧЕСКИЙ АНАЛИЗ КРИПТОГРАФИЧЕСКИХ ПРОТОКОЛОВ: ПОСТРОЕНИЕ ЗАПРОСОВ В PROVERIF

Ф. Б. Дасько

*Белорусский государственный университет,
Минск, Беларусь, dasko@bsu.by*

В статье предложен подход к формированию запросов в системе автоматизированного анализа криптографических протоколов ProVerif. Представлены запросы для проверки шести ключевых свойств безопасности: конфиденциальности, аутентификации, защиты от атак повтора, подтверждения ключа, обновления ключа и защиты от «чтения назад». Результаты могут быть использованы для повышения качества анализа за счет формального подхода к построению запросов, исключающего возможные ошибки и неточности.

Ключевые слова: криптографические протоколы; автоматизированный анализ протоколов; ProVerif; запросы безопасности; свойства безопасности.

AUTOMATED ANALYSIS OF CRYPTOGRAPHIC PROTOCOLS: CONSTRUCTION OF QUERIES IN PROVERIF

F. B. Dasko

*Belarusian State University,
Minsk, Belarus, dasko@bsu.by*

We present an approach to construct queries in ProVerif, an automated cryptographic protocol analyzer. We develop query templates for verifying six fundamental security properties: confidentiality, authentication, resistance to replay attacks, key confirmation, fresh key derivation and perfect forward secrecy. Our methodology enhances the reliability of verification results through a rigorous formal approach to query formulation that eliminates potential ambiguities and errors.

Keywords: cryptographic protocols; automated protocol verification; ProVerif; security queries; security properties.

1. Введение

Формальный анализ криптографических протоколов является эффективным инструментом, позволяющим ускорить и упростить процесс проверки свойств безопасности протоколов. При этом наличие или отсутствие ожидаемых свойств доказывается строго математически.

Одним из популярных средств автоматизированного анализа является ProVerif [1]. В основе работы данного инструмента лежит понятие модели

протокола (model) – формального описания протокола на специализированном языке. Модель определяет каналы связи, криптографические примитивы, участников протокола и их действия. Ключевые шаги протокола маркируются с помощью событий (events) – специальных меток, обозначающих значимые состояния или действия в ходе выполнения протокола.

Свойства безопасности, которые необходимо проверить, формулируются в виде запросов (queries) – утверждений формальной логики над одним или несколькими событиями. ProVerif анализирует модель и по каждому запросу выносит один из трех вердиктов: `true` (доказано, что утверждение истинно для всех возможных исполнений протокола), `false` (найден контрпример), или `can not be proved` (инструмент не смог ни доказать, ни опровергнуть свойство).

Для корректного анализа свойств протокола критически важна точная формулировка запросов. Ошибка в построении запроса приведет к тому, что он не будет соответствовать проверяемому свойству безопасности, что обесценивает результаты анализа.

Цель настоящей работы – предложить систематический подход к формулированию запросов в ProVerif для проверки основных свойств безопасности: конфиденциальности, аутентификации, защиты от атак повтора (replay protection), подтверждения ключа (key confirmation), обновления ключа (fresh key derivation) и защиты от «чтения назад» (perfect forward secrecy).

В [3] мы иллюстрируем предлагаемый подход на примере протокола Station-to-Station (STS) [2]. Этот протокол выбран благодаря своей минималистичности, хорошей изученности. На примере STS легко проследить выполнение вышеперечисленных свойств.

В разделе 2 представлены используемые соглашения. Разделы 3–9 посвящены анализу свойств безопасности. В данных разделах приводятся определения исследуемых свойств, затем они формулируются в виде строгих логических высказываний, которые, в свою очередь, трансформируются в конкретные запросы ProVerif. В разделе 10 обсуждаются полученные результаты.

Мы предполагаем, что читатель знаком с синтаксисом ProVerif.

2. Соглашения

При описании событий модели, запросов и результатов анализа мы используем синтаксис ProVerif, а также определенные именованные элементы протокола для повышения читаемости запросов.

В модели используются следующие объекты:

- `pk` – открытые долговременные ключи;

- `sk` – личные долговременные ключи;
- `eph_pk` – открытые сеансовые ключи;
- `eph_sk` – личные сеансовые ключи;
- `K` – секретные ключи;
- `id` – идентификаторы сторон;
- `cert` – сертификаты;
- `ctext` – шифртексты.

При построении модели протокола STS мы учитывали, что протокол проводится в параллельном режиме, т.е. стороны действуют симметрично и одновременно. Симметричность действий сторон позволяет определить в модели для обеих сторон одну и ту же последовательность действий, формализованную как один процесс, запускаемый с различными входными параметрами.

Поскольку один процесс используется для обеих сторон протокола, то указать, что данные принадлежат определенной стороне протокола (данные стороны А или данные стороны В) не удастся. Поэтому используется указание на то, что данные созданы текущей стороной или получены извне (собственные данные и принятые данные).

Для указания принадлежности данных используются постфиксы:

- `self` – данные, определяемые текущей стороной процесса;
- `peer` – данные, определяемые второй стороной процесса;
- `t` – данные, определяемые третьей доверенной стороной.

В протоколе STS стороны используют свой идентификатор и идентификатор второй стороны, владеют личным и открытым ключом. Открытый ключ заверен третьей доверенной стороной в виде сертификата. Стороны протокола владеют ключом, позволяющим проверять сертификаты третьей доверенной стороны. Все эти данные передаются в процесс стороны как входные параметры:

```
processParty(id_self: identifier, id_peer: identifier,
sk_self: privKey, cert_self: bitstring, pk_t: pubKey).
```

3. Конфиденциальность

Свойство конфиденциальности гарантирует, что определённые данные (передаваемые как часть содержимого защищенного сообщения или сформированные на основе данных, которыми обмениваются стороны), не будут доступны нелегитимным сторонам и останутся неизвестными для злоумышленника.

В системе ProVerif для проверки доступности данных противнику применяется логический предикат `attacker(data)`. Данный предикат

принимает истинное значение в случае, если атакующий способен получить значение `data`.

Данные `data` могут быть представлены через свободное имя или через переменную. Использование свободных имен более гибкое, но требует привязки к событию протокола.

Свойство может быть формализовано так: «Если участник протокола обладает данными `data`, то они не известны атакующему».

Мы выделяем два типа событий, рекомендуемых для использования в запросах конфиденциальности: «участник протокола получил значение `data`» (обозначим `PartyKnowsData(data)`) и «участник завершил протокол со значением `data`» (обозначим `PartyFinishes(data)`). Первый тип события фиксирует знание на наиболее раннем этапе, в то время как второй – на наиболее позднем.

В зависимости от выбора события запрос конфиденциальности в ProVerif примет следующий вид:

```
query data: datatype; event(PartyKnowsData(data)) &&
attacker(data) ==> false.
```

или

```
query data: datatype; event(PartyFinishes(data)) &&
attacker(data) ==> false.
```

Запрос записан в виде импликации с посылкой `false` для того, чтобы избежать применения ключевого слова `not`, использование которого в запросах затруднено.

Необходимо отметить, что указанные запросы не эквивалентны, и запрос, использующий событие завершения протокола, является менее строгим. Более строгий запрос с фиксацией непосредственного получения данных рекомендуется применять к тем данным, которые сохраняют ценность за пределами конкретного сеанса. К такой информации относятся долговременные ключи и секретные сообщения сторон.

На примере протокола STS проверим конфиденциальность долговременного личного ключа `sk_self`, а также сеансового ключа `K`, получаемого из общего секрета. Для этого зафиксируем событие `PartyKnowsLTkey(sk_self)` в самом начале процесса стороны и `PartyFinishes(K)` в конце процесса стороны.

Результат проверки:

```
Query not (event(PartyKnowsLTkey(sk)) && attacker(sk)) is
true.
Query not event( PartyFinishes(id_self_2, id_peer_2,
eph_sk_self_2, K_2)) && attacker_p1(K_2)) is true.
```

Указание фазы `phase 0` продиктовано проверкой свойства защиты от «чтения назад», которое будет рассмотрено далее.

Следует отметить, что ослабленное свойство для долговременного ключа получает вердикт `true`, тогда как усиленное свойство для сеансовых ключей возвращает результат `false`. Это объясняется возможностью злоумышленника подменить сеансовый ключ одного из легитимных участников своим значением. В результате атакующий и легитимная сторона вычисляют одинаковый ключевой материал, на основе которого формируются идентичные секретные ключи, фиксируемые ранним событием получения знания о ключе.

4. Аутентификация стороны

Аутентификация стороны представляет собой свойство протокола, обеспечивающее для одного из участников возможность анализа полномочий второго участника, а также подтверждение факта непосредственного участия второй стороны в текущем сеансе протокола.

В соответствии с данным определением, свойство формулируется так: «Если сторона А успешно завершила протокол, то сторона В действительно принимала в нем участие». Требование участия стороны В в протоколе может интерпретироваться различным образом.

Наименее строгим требованием будет то, что сторона В вступила в протокол, тогда как наиболее строгим – что последнее сообщение, полученное стороной А, действительно было отправлено стороной В (т.е. выполнение стороной В наиболее позднего шага, требующегося стороне А для успешного завершения протокола).

Такое соотношение между требованиями выполняется в связи с тем, что если сторона выполнила некоторый шаг протокола, то она выполнила и предыдущие шаги, тогда как выполнение последующих не гарантировано.

Наиболее строгое требование мы выделим в отдельное свойство «аутентификация данных» и рассмотрим его в следующем разделе. Здесь же остановимся на наименее строгом свойстве.

Запрос на его проверку в ProVerif имеет следующий вид:

```
query event(PartyFinishes(..., id_peer)) ==>
event(PartyStarted(id_peer)).
```

Здесь `PartyFinishes(..., id_peer)` это событие в запросе означает что текущая сторона `id_self` успешно закончила протокол с (как она считает) `id_peer`. Событие `PartyStarted(id_peer)` в запросе означает что сторона `id_peer` начала протокол, событие в процессе стороны фиксирует собственный идентификатор т.е. записывается как

`PartyStarted(id_self)`. Равенство идентификаторов в запросе подчеркивает, что события выполнены разными сторонами протокола.

Результат выполнения запроса в протоколе STS:

```
Query event(PartyFinishes(id_self_2, id_peer_2,
eph_sk_self_2, K_2)) ==> event(PartyStarted(id_peer_2 )) is
true.
```

5. Аутентификация данных

Свойство аутентификации данных обеспечивает гарантию того, что полученное сообщение или его компоненты сформированы конкретным участником протокола в определенный момент времени в прошлом, и что эти данные не подверглись модификации или фальсификации. При этом гарантии уникальности и актуальности данных могут не предоставляться.

Запрос для проверки данного свойства может быть сформулирован следующим образом: «если участник проверил данные и принял их, то легитимный участник действительно отправил эти данные». Необходимо подчеркнуть, что событие принятия данных должно фиксироваться непосредственно после успешной проверки полученной информации, а не просто после получения сообщения. Без проверки обеспечение свойства аутентификации данных невозможно.

Запросы аутентификации данных в ProVerif могут быть сформулированы следующим образом:

```
query id_peer:identificator, data_peer: datatype;
event(PartyAcceptsData(id_peer, data_peer)) ==>
event(PartySendsData(id_peer, data_peer))
```

Здесь событие `PartySendsData(id_self, data_self)` фиксируется непосредственно после отправки своих данных, а событие `PartyAcceptsData(id_peer, data_peer)` фиксируется сразу после проверки полученных данных. Одинаковые идентификаторы в событиях запроса указывают на то, что события фиксируются на разных сторонах протокола.

Результат выполнения запроса в протоколе STS:

```
Query event(PartyAcceptsData(id, pk, cert, ctext)) ==>
event(PartySendsData(id, pk, cert, ctext)) is true.
```

Отметим на примере протокола STS, что корректность второго сообщения означает, что и заявленный сеансовый ключ из первого сообщения был отправлен той же легитимной стороной, которая отправила второе сообщение. То, что из более позднего события (отправка второго сообщения) следует более раннее события этой же стороны (отправка первого сообщения) с согласующимися параметрами, является тривиальным свойством,

которое выполняется независимо от протокола и его фактических свойств. В связи с этим оказывается, что построение корректного запроса в виде следствия из проверки более позднего сообщения к корректности более раннего не проще, чем проведение полноценной аналитической работы.

6. Защита от повтора

Защита от повтора означает, что данные, полученные стороной протокола, были отправлены второй стороной протокола в том же сеансе.

Следует отметить два ключевых момента относительно данного свойства:

Во-первых, в ProVerif запросы конфиденциальности сразу учитывают возможность различных атак, в том числе и атак повтора (если стороны протокола реплицируются), но отдельная проверка своевременности данных все еще желательна.

Во-вторых, данные, которые могут быть признаны актуальными, должны быть созданы только легитимным участником. В этом смысле защита от повтора представляет собой усиленное требование к аутентификации данных, т.е. данные не только сформированы легитимным участником протокола, но и в конкретном сеансе.

Запрос для проверки данного свойства может быть сформулирован следующим образом: «если участник в некотором сеансе верифицировал данные и принял их, то легитимный участник действительно отправил эти данные в рамках этого же сеанса».

Здесь свойство требует контролировать события разных сторон и их принадлежность к одному и тому же сеансу протокола.

Универсальным решением будет использование инъективных событий:

```
query inj-event(PartyAcceptsData(id, data)) ==>
inj-event(PartySendsData(id, data)).
```

Альтернативным подходом является привязка к волатильным (генерируемым в каждом сеансе протокола) данным. Сложность заключается в том, что их волатильность и согласованность для двух сторон должна быть доказана. В этом смысле для протоколов выработки общего ключа (в том числе STS) было бы естественно использовать привязку к сеансовому ключу сторон, а свойства обновления и согласованности ключа (рассмотренные далее) проверяются явно. Однако этот подход не универсален. Более универсальным будет доказать волатильность аутентифицируемых данных. Таким образом, запрос, усиливающий свойство аутентификации данных до защиты от повтора, будет следующим:

```
query event(PartyAcceptsVolatileData(data_peer, SID_self1))
&& event(PartyAcceptsVolatileData(data_peer, SID_self2)) ==>
SID_self1 = SID_self2.
```

Здесь `SID_self` представляет собой некоторые гарантированно волатильные данные, генерируемые самой стороной и применяемые для отличия одного сеанса от другого.

Результат выполнения запросов в протоколе STS:

```
Query inj-event(PartyAcceptsData(id, pk, cert, ctext)) ==>
inj-event(PartySendsData(id, pk, cert, ctext)) is true.
Query event(AttributedVolatileData(ctext, eph_sk_self1)) &&
event(AttributedVolatileData(ctext, eph_sk_self2)) ==>
eph_sk_self1 = eph_sk_self2 is true.
```

Таким образом, подтверждается свойство защиты от повтора в протоколе STS.

7. Подтверждение ключа

Подтверждение ключа означает, что один из участников протокола получает гарантию того, что второй участник действительно обладает конкретным секретным ключом (или ключевым материалом, необходимым для вычисления ключа).

Данное свойство формализуется так: «если участник А завершает протокол с общим ключом, то участник В владеет этим ключом».

Запросы подтверждения ключа в ProVerif могут быть сформулированы следующим образом:

```
Query id_peer:identifier, k1:key;
event(PartyFinishes(..., id_peer, k)) ==>
event(PartyKnowsKey(id_peer, k)).
```

Результат выполнения запроса в протоколе STS:

```
Query event(PartyFinishes(id_self_2, id_peer_2,
eph_sk_self_2, K_2)) ==>
event(PartyKnowsSessionKey(id_peer_2, K_2)) is true.
```

Таким образом, свойство подтверждения ключа подтверждается для протокола STS.

8. Обновление ключа

Протокол обеспечивает обновление ключа, если в каждом сеансе протокола используется уникальный сеансовый ключ, что обеспечивает защиту от компрометации данных, переданных в предыдущих сеансах, даже при компрометации текущего сеансового ключа.

Обновление ключа формализуется следующим образом: «если участник протокола завершил сеанс протокола N с определенным ключом, и участник завершил сеанс M с тем же ключом, то это была один и тот же сеанс, т.е. $N = M$ ». Номер сеанса может быть эквивалентно заменен произвольными волатильными данными, сгенерированными данным

участником протокола; совпадение волатильных данных гарантирует принадлежность событий к одному и тому же сеансу.

Запросы обновления ключа в ProVerif могут быть сформулированы следующим образом:

```
Query id_self: identifier, k1, k2: keys, v_data1, v_data2:
volatileData;
event(PartyFinishes(id_self, v_data1, k1)) &&
event(PartyFinishes(id_self, v_data2, k2)) ==>
v_data1 = v_data2.
```

Результат выполнения запроса в протоколе STS:

```
Query event(PartyFinishes(id_self_2, id_peer1, eph_sk_self1,
K_2)) && event(PartyFinishes(id_self_2, id_peer2,
eph_sk_self2, K_2)) ==> eph_sk_self1 = eph_sk_self2 is true.
```

Необходимо подчеркнуть значимость собственного идентификатора, необходимого для отличия штатной ситуации, где один и тот же ключ получают две стороны одного сеанса, от ситуации, когда одна сторона получает одинаковый ключ дважды.

Отметим, что использование события, фиксирующего сеансовые ключи в момент их генерации, может не отражать целевое свойство. Так, в контексте протокола STS такой запрос даст отрицательный результат. Это обусловлено тем, что два репликанта (клона) участника А могут взаимодействовать между собой, генерируя идентичный сеансовый ключ, но поскольку репликанты моделируют различные сеансы протокола, возникает ситуация, при которой одним участником в различных сеансах фиксируется идентичный сеансовый ключ.

9. Защита от «чтения назад»

Под защитой от «чтения назад» понимается свойство протокола, которое гарантирует, что компрометация долговременных ключей, используемых в протоколе, не приводит к компрометации ранее сформированных сеансовых ключей.

Для проверки данного свойства необходимо смоделировать ситуацию, где стороны протокола проводят протокол, после чего атакующему передаются значения всех долговременных ключей, и проводится запрос: может ли атакующий при этих условиях вычислить значения сеансовых ключей.

Для моделирования такой ситуации в ProVerif применяются фазы, где нулевая фаза – это нормальное проведение протокола, а в фазе 1 раскрываются все долговременные параметры всех сторон протокола. Вследствие этого при проверке конфиденциальности долговременных параметров необходимо явно указывать, что проверяется свойство именно в фазе 0 (до раскрытия значения).

В виде запроса ProVerif проверку защиты от «чтения назад» можно записать как:

```
query data: datatype; event(PartyKnowsData(data)) &&  
attacker(data) phase 1 ==> false.
```

Результаты анализа протокола STS:

```
Query not (event(PartyFinishes(id_self_2, id_peer_2,  
eph_sk_self_2, K_2)) && attacker_p1(K_2)) is true.
```

Таким образом, подтверждается, что протокол STS обладает свойством защиты от «чтения назад».

10. Заключение

В данной работе представлен подход к формированию запросов в ProVerif для проверки основных свойств безопасности криптографических протоколов. Применение запросов на примере протокола Station-to-Station продемонстрировало эффективность формализации для строгой математической проверки таких свойств, как конфиденциальность, аутентификация, защита от атак повтора, подтверждение ключа, обновление ключа и защита от «чтения назад». Результаты анализа подтвердили соответствие протокола STS заявленным свойствам безопасности.

Следует отметить, что в настоящей работе не рассматривались свойства, требующие проверки через «эквивалентность наблюдений» (observational equivalence), такие как анонимность сторон или невозможность отслеживания действий участников. Анализ этих свойств с использованием ProVerif, а также расширение предложенного подхода на более сложные протоколы представляют собой перспективные направления для дальнейших исследований.

Библиографические ссылки

1. ProVerif 2.05: Automatic Cryptographic Protocol Verifier, User Manual and Tutorial / B. Blanchet [et al.]. URL: <https://bblanche.gitlabpages.inria.fr/proverif/manual.pdf> (date of access: 22.09.2025).
2. Diffie W., van Oorschot P. C., Wiener M. J. Authentication and authenticated key exchanges // Designs, Codes and Cryptography. 1992. Vol. 2, iss. 2, P. 107–125.
3. Дасько Ф. Б. ProVerif-STS-model // GitHub. URL: <https://github.com/HassiumPotassium/ProVerif-STS-model> (date of access: 22.09.2025).
4. Blanchet, B. Modeling and verifying security protocols with the applied pi calculus and ProVerif // Foundations and Trends in Privacy and Security. 2016. Vol. 1, iss. 1–2. P. 1–135.
5. Basin D., Cremers C., Meadows C. Model checking security protocols // Handbook of Model Checking / eds.: E. M. Clarke [et al.]. Springer, Cham, 2018. P. 727–762.

КРИПТОГРАФИЧЕСКИЕ МЕТОДЫ ЗАЩИТЫ ДАННЫХ В 5G-СЕТЯХ: ОБЕСПЕЧЕНИЕ БЕЗОПАСНОСТИ И КОНФИДЕНЦИАЛЬНОСТИ В ВЫСОКОСКОРОСТНЫХ СЕТЯХ НОВОГО ПОКОЛЕНИЯ

Р. К. Дригель, Е. И. Вакарь, С. Н. Нестеренков

*Белорусский государственный университет информатики и радиоэлектроники,
Минск, Беларусь, romanbratishkingg@gmail.com*

В работе исследуются криптографические методы защиты данных в 5G-сетях, включая алгоритмы шифрования, протоколы аутентификации и механизмы обеспечения конфиденциальности. Анализируются современные угрозы, такие как перехват сигналов в радиоканале и уязвимости в архитектуре сетевых сегментов. Рассмотрены стандарты 3GPP и перспективные направления, в том числе постквантовая криптография.

Ключевые слова: 5G; криптография; информационная безопасность; аутентификация; шифрование.

CRYPTOGRAPHIC METHODS FOR DATA PROTECTION IN 5G NETWORKS: ENSURING SECURITY AND PRIVACY IN NEXT- GENERATION HIGH-SPEED NETWORKS

R. K. Dryhel, E. I. Vakar, S. N. Nesterenkov

*Belarusian State University of Informatics and Radioelectronics,
Minsk, Belarus, romanbratishkingg@gmail.com*

This study investigates cryptographic methods for data protection in 5G networks, including encryption algorithms, authentication protocols, and confidentiality mechanisms. The analysis focuses on contemporary threats, such as radio signal interception and vulnerabilities in network segment architecture. The work examines 3GPP standards and explores emerging directions, particularly post-quantum cryptography.

Keywords: 5G; cryptography; information security; authentication; encryption.

1. Введение

Сети пятого поколения (5G) обеспечивают высокую скорость передачи данных и низкие задержки, но при этом создают новые вызовы в области информационной безопасности. Криптографические методы играют

ключевую роль в защите конфиденциальности и целостности данных. Цель данной работы — анализ современных криптографических решений для 5G, включая их эффективность против актуальных угроз и перспективы развития. В статье рассматриваются стандарты 3GPP, алгоритмы шифрования и методы аутентификации, а также обсуждаются проблемы внедрения и пути их решения.

2. Угрозы безопасности в 5G-сетях

Архитектура 5G, несмотря на свои преимущества, сталкивается с серьезными вызовами в области безопасности. Высокая степень виртуализации, использование программно-определяемых технологий (SDN/NFV) и массовое подключение IoT-устройств создают новые уязвимости, которые требуют инновационных подходов к защите данных.

На физическом уровне одной из наиболее распространенных угроз остается воздействие на беспроводной канал связи. Например, технология IMSI-сниффинга, известная еще со времен 4G, эволюционировала благодаря применению программно-определяемых радиомодулей (SDR). Злоумышленники могут развертывать поддельные вышки сотовой связи (false base station), которые имитируют легитимные, что позволяет перехватывать трафик пользователей. По данным исследований 2023 года, такие атаки стали на 25% чаще по сравнению с предыдущим годом [1].

Виртуализированная инфраструктура 5G, включая сетевые срезы (Network Slicing), также уязвима. Атаки на контроллеры OpenFlow могут привести к компрометации целых сегментов сети. Например, в 2022 году была обнаружена уязвимость CVE-2022-27578, позволяющая злоумышленникам нарушать изоляцию между срезами. Это особенно критично для промышленных IoT-решений, где утечка данных может привести к катастрофическим последствиям [2].

Edge-вычисления, несмотря на свои преимущества, также подвергаются атакам. Периферийные узлы обладают меньшим уровнем защиты, чем централизованные системы. Например, в одном из кейсов 2023 года хакеры использовали уязвимости в edge-устройствах для доступа к данным медицинских IoT-датчиков.

Среди злоумышленников можно выделить три основные группы: киберпреступники, нацеленные на финансовую выгоду; хактивисты, стремящиеся нарушить работу сервисов; и государственные акторы, заинтересованные в слежке. Каждая из этих групп использует уникальные методы, что требует дифференцированного подхода к защите.

3. Криптографические методы защиты в 5G

Безопасность данных в 5G обеспечивается комбинацией проверенных и новых криптографических решений. Основу составляют протоколы аутентификации, алгоритмы шифрования и механизмы защиты идентификаторов, адаптированные под высокоскоростные сети.

Для аутентификации в 5G применяется усовершенствованный протокол 5G AKA, который устраняет многие уязвимости своих предшественников. В отличие от EPS-AKA, используемого в 4G, новый протокол включает схему SUCI, где постоянный идентификатор SUPI шифруется с помощью ECIES на основе эллиптических кривых (чаще всего Curve25519). Это снижает риск перехвата IMSI. Для IoT-устройств самым распространенным является протокол EAP-TLS с сертификатами X.509, работающий на основе взаимной аутентификации [3].

Шифрование данных в 5G строится на алгоритмах, оптимизированных для высокой производительности. В пользовательской плоскости наиболее распространенным является AES-256 в режиме GCM/GMAC, который обеспечивает и конфиденциальность, и целостность данных. Для совместимости с устаревшим оборудованием сохраняется поддержка SNOW 3G и ZUC. Интересно, что использование аппаратного ускорения (AES-NI) позволяет достигать скорости шифрования до 40 Gbps, что критично для сетей с низкими задержками [4].

Стандарты 3GPP предусматривают защиту идентификаторов через схему SUPI/SUCI.

Перспективным направлением является внедрение постквантовой криптографии. В настоящее время тестируются алгоритмы, такие как Kyber (для ключевого соглашения) и Dilithium (для цифровых подписей).

4. Реализация и стандартизация криптозащиты в 5G

Разработка и внедрение криптографических методов в 5G-сетях осуществляется в рамках международных стандартов и отраслевых соглашений. Основным регулирующим документом выступают спецификации 3GPP, которые определяют архитектуру безопасности и обязательные к применению криптоалгоритмы. В серии релизов 15-17 были закреплены ключевые требования к защите пользовательских данных, сигнализации и сетевой инфраструктуры.

Архитектура безопасности 5G построена по принципу сервис-ориентированной модели (SBA), где каждый функциональный элемент отвечает за конкретные аспекты защиты. Центральную роль играют следующие компоненты: SEAF (Security Anchor Function) обеспечивает

начальную аутентификацию устройств, AUSF (Authentication Server Function) управляет процессами верификации, а SIDF (Subscription Identifier De-concealing Function) отвечает за обработку защищенных идентификаторов. Такое разделение функций позволяет гибко масштабировать систему защиты при сохранении единых стандартов безопасности.

Сложность представляет обеспечение совместимости между оборудованием разных вендоров. Проблема проявляется при реализации криптографических протоколов, таких как 5G AKA, где даже незначительные отклонения в интерпретации стандартов могут привести к нарушениям работы сети. Для решения этой проблемы GSMA разработала программу сертификации безопасности, включающую тестирование межвендорной совместимости криптографических реализаций [5].

Сравнительная характеристика методов

Метод защиты	Алгоритмы	Уровень стойкости	Применение в 5G
Аутентификация	5G AKA, EAP-TLS	Высокий	Все соединения
Шифрование данных	AES-256, ZUC	Высокий	Пользовательская плоскость
Защита идентификации	ECIES (Curve25519)	Высокий	Начальная регистрация
Постквантовая защита	Kyber, Dilithium	Экспериментальный	Будущие реализации

Примечание. Сравнительная характеристика основных методов защиты.

5. Проблемы и ограничения современных криптографических решений

Реализация криптографической защиты в 5G-сетях сталкивается с рядом фундаментальных и практических ограничений. Основной парадокс современной защиты данных заключается в необходимости балансировать между криптостойкостью и производительностью в условиях жестких требований к задержкам.

Производительность криптографических алгоритмов остается ключевым узким местом для высокоскоростных 5G-сетей. Даже при использовании аппаратного ускорения (AES-NI, криптографических сопроцессоров) дополнительные задержки от операций шифрования/дешифрования могут достигать 10–15% от общего времени обработки пакета. Особенно остро эта проблема проявляется в edge-вычислениях, где требования к low-latency конфликтуют с необходимостью обеспечения безопасности.

Энергопотребление криптографических операций является критическим фактором для массовых IoT-устройств. Реализация полноценной защиты на датчиках с ограниченным энергобюджетом часто приводит к сокращению срока работы батареи на 30–40%.

Совместимость с устаревшим оборудованием создает дополнительные риски безопасности. Необходимость поддержки старых алгоритмов (SNOW 3G, Kasumi) для обратной совместимости с 4G-устройствами расширяет поверхность для потенциальных атак. Статистика показывает, что около 35% успешных атак на гибридные 4G/5G-сети используют уязвимости в устаревших криптографических протоколах.

Реализация постквантовой криптографии сталкивается с техническими сложностями. Алгоритмы NIST PQC (Kyber, Dilithium) требуют в 10–100 раз больше вычислительных ресурсов по сравнению с традиционной ECC-криптографией. Размер ключей и подписей увеличивается в 4–5 раз, что создает проблемы для сигнальных сообщений с жесткими ограничениями на длину пакета.

Регуляторные требования в разных странах создают дополнительные сложности для разработчиков. В Европе основным ориентиром служат рекомендации ENISA, в США – стандарты NIST, а в Китае – спецификации CCSSA. Различия в требованиях к длине ключей, допустимым алгоритмам и механизмам восстановления данных вынуждают производителей создавать региональные версии оборудования. Ещё одной проблемой является создание «бэкдоров»: в некоторых странах законодательство обязывает операторов предоставлять правоохранительным органам доступ к зашифрованным данным.

Перспективные направления преодоления всех этих ограничений включают:

- развитие квантово-безопасных lightweight-алгоритмов;
- аппаратные ускорители нового поколения;
- гибридные криптосистемы с адаптивным уровнем защиты;
- технологии конфиденциальных вычислений;
- стандартизированные API для криптографических сервисов.

6. Заключение

Исследование подтверждает, что криптографические методы в 5G обеспечивают существенное улучшение защиты по сравнению с предыдущими поколениями сетей. Однако сохраняются ключевые проблемы: высокие вычислительные затраты, энергопотребление IoT-устройств и необходимость обратной совместимости. Переход на постквантовые алгоритмы создает новые уязвимости. Решение требует оптимизации

существующих методов, разработки гибридных систем и создания специализированных аппаратных ускорителей. Эти меры позволят обеспечить безопасность 5G-сетей при сохранении их производительности.

Библиографические ссылки

1. Huawei. 5G Security: Challenges and Solutions. URL: <https://www.huawei.com/en/industry-insights/outlook/mobile-broadband/5g-security-solutions> (date of access: 05.06.2025).
2. Петренко С. А. Кибербезопасность сетей 5G: новые вызовы и подходы // Информационная безопасность. 2024. № 2. С. 22–29.
3. Иванов М. А., Сидоров А. В. Криптографические протоколы для сетей 5G: анализ и перспективы // Электросвязь. 2025. № 3. С. 15–22.
4. Zero Trust Architecture / S. Rose [et al.]. NIST. 2020. Special Publication 800-207. URL: <https://csrc.nist.gov/publications/detail/sp/800-207/final> (date of access: 05.06.2025).
5. Basin D., Dreier J., Sasse R. Automated Analysis of 5G AKA Protocol // Proceedings of the IEEE Symposium on Security and Privacy. 2023. P. 890–905.

ОЦЕНКА СТЕПЕНИ РАЗРУШЕНИЯ СТАТИСТИКИ ЕСТЕСТВЕННОГО ЯЗЫКА ПРИ ОБЕЗЛИЧИВАНИИ ПЕРСОНАЛЬНЫХ ДАННЫХ НА ОСНОВЕ МЕТОДА ИЗМЕНЕНИЯ СОСТАВА ИЛИ СЕМАНТИКИ

Я. А. Клиндухов

*Национальный детский технопарк,
Минск, Беларусь, klinduhov.science@gmail.by*

В данной работе предложена процедура обезличивания персональных данных, основанная на методе изменения состава или семантики с использованием доверительных вычислительных баз (ДВБ). Разработана структурная схема процедуры обезличивания, включающая разбиение данных на блоки и их последовательную обработку с применением различных ДВБ. Для оценки эффективности метода проведён анализ индекса совпадения Фридмана до и после обезличивания. Результаты показали снижение индекса совпадения более чем в 2,5 раза, что свидетельствует о значительном разрушении статистики естественного языка и подтверждает высокий уровень информационной безопасности. Полученные значения приближаются к характеристикам случайного текста, что подчеркивает практическую применимость метода при сохранении высокого уровня информационной безопасности персональных данных.

Ключевые слова: персональные данные; обезличивание персональных данных; метод изменения состава или семантики; индекс совпадения Фридмана.

ASSESSING THE DEGREE OF NATURAL LANGUAGE STATISTICS DESTRUCTION DURING PERSONAL DATA ANONYMIZATION BASED ON THE METHOD OF COMPOSITION OR SEMANTICS ALTERATION

Y. A. Klindukhov

*National Children's Technopark,
Minsk, Belarus, klinduhov.science@gmail.by*

This paper proposes a procedure for personal data anonymization based on the method of altering composition or semantics using trusted computing bases (TCBs). A structural scheme of the anonymization procedure has been developed, involving partitioning data into blocks and their sequential processing using various TCBs. To evaluate the method's effectiveness, an analysis of the Friedman coincidence index before and after anonymization was conducted. The results showed a reduction in the coincidence index by more than 2.5 times, indicating a significant disruption of natural language statistics and confirming a high level of information security. The obtained values approach the characteristics of random text, highlighting the method's practical applicability while maintaining a high level of personal data information security.

Keywords: personal data; personal data anonymization; method of altering composition or semantics; Friedman coincidence index.

1. Введение

В условиях цифровой трансформации и стремительного роста объемов обрабатываемых персональных данных важную роль играет процесс их обезличивания [1–3]. Под обезличиванием персональных данных будем понимать действия, в результате которых становится невозможным без использования дополнительной информации определить принадлежность персональных данных конкретному субъекту персональных данных.

Одним из таких методов является метод изменения состава или семантики, сущность которого заключается в том, что выполняют обобщение, изменение или удаление части сведений, позволяющих идентифицировать субъекта персональных данных. При этом полученные обезличенные персональные данные и правила их изменения необходимо хранить раздельно.

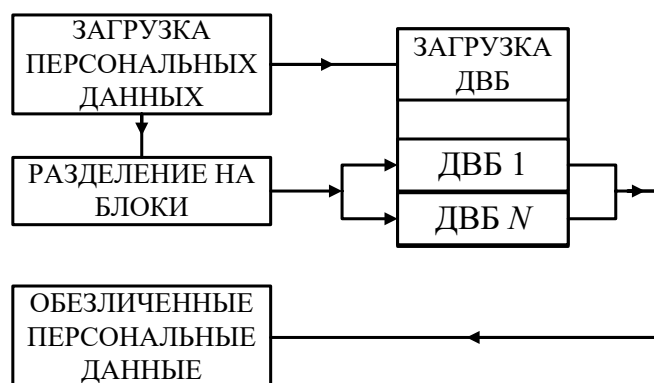
Обезличивание персональных данных указанным выше методом не требует наличие высоких вычислительных мощностей, однако характеризуется следующими недостатками. Во-первых, удаление части сведений, позволяющих идентифицировать субъекта персональных данных, без их сохранения в отдельной базе данных приводит к утрате такого свойства обезличенных персональных данных, как полнота. Во-вторых, простые замены исходных символов персональных данных обезличенными сохраняет вероятности появления соответствующих символов обезличенных персональных данных. В этом случае нарушитель информационной безопасности имеет возможность, получив доступ к обезличенным персональным данным, рассчитать вероятности появления отдельных символов и различных их сочетаний, что позволит ему раскрыть содержимое доверительной вычислительной базы. В связи с этим целью данной работы являлась оценка степени разрушения статистически естественного языка обезличенных персональных данных при обезличивании на основе метода изменения состава или семантики с использованием набора доверительных вычислительных баз.

В качестве объекта исследования использован метод изменения состава или семантики. Данный метод выбран в качестве объекта исследования, поскольку он является одним из методов обезличивания персональных данных в соответствии с требованиями Оперативно-аналитического центра при Президенте Республики Беларусь [4].

Предметом исследования являлись количественные показатели разрушения статистики естественного языка при применении предложенной структурной схемы обезличивания персональных данных на основе метода изменения состава или семантики.

2. Обезличивание персональных данных с использованием блочной замены и набора различных ДВБ

На рисунке показана структурная схема, реализующая метод изменения состава или семантики.



Общая структурная схема функционирования компьютерной программы, реализующей схему обезличивания персональных данных с применением различных ДВБ

Сущность функционирования данной схемы заключается в следующем. Персональные данные, подлежащие обезличиванию, загружают в доверительную вычислительную базу (ДВБ), которая содержит набор таблиц подстановки ДВБ 1, ДВБ 2, ..., ДВБ N. Персональные данные разбивают на блоки, каждый из которых обезличивают с помощью ДВБ 1, ДВБ 2, ..., ДВБ N, ДВБ 1, ДВБ 2, ..., ДВБ N и т.д.

Важно отметить, что содержимое ДВБ необходимо соблюдать сохранять в секрете и обновлять в соответствии с требованиями в месте эксплуатации информационной системы.

3. Оценка степени разрушения статистики естественного языка при обезличивании персональных данных на основе метода изменения состава или семантики

Для оценки уровня информационной безопасности обезличенных персональных данных определена вероятность появления каждого символа исходных персональных данных и персональных данных, обезличенных с использованием различных ДВБ. На основании полученных данных был вычислен индекс совпадения Фридмана. Данный метод криптоанализа был разработан американским криптографом и криптоаналитиком Уильямом Фридманом и опубликован в 1920 году. Выбор именно этого метода обусловлен его высокой чувствительностью к статистическим свойствам текста: индекс совпадения Фридмана позволяет количественно

оценить степень отклонения распределения символов от характеристик естественного языка и приближения к случайному тексту, что является ключевым критерием эффективности процедуры обезличивания персональных данных. Результаты математических вычислений представлены в табл. 1 и 2. Для данных вычислений использовалась следующая расчетная формула:

$$IC = \sum_{i=1}^n (p_i)^2, \quad p_i = \frac{f_i}{N}, \quad (1)$$

где IC – индекс совпадения Фридмана; p_i – вероятность появления символа в тексте; f_i – частота появления символа в тексте; N – количество символов в тексте.

Таблица 1

Индекс совпадений для исходных персональных данных

№	Символ	Количество	Вероятность появления	Вклад в индекс совпадений
1	а	120002	0,1015	0,0103
2	(	20	0,0169	0,0003
3	)	20	0,0169	0,0003
4		30	0,0254	0,0006
5	-	60	0,0507	0,0026
...	...	...	...	...
50	я	6259	0,0053	0,0000
				0,0398

Примечание. Индекс совпадения исходных персональных данных составил 0,0398.

Таблица 2

Индекс совпадений персональных данных после процедуры обезличивания

№	Символ	Количество	Вероятность появления	Вклад в индекс совпадений
1	\$	1961	0,0017	0,0000
2	+	5968	0,0050	0,0000
3	/	1189	0,0101	0,0001
4	5	949	0,0008	0,0000

Окончание табл. 2

№	Символ	Количество	Вероятность появления	Вклад в индекс совпадений
5	7	20746	0,0175	0,0003
...	...	...	...	...
120	è	1	0,0000008	0,0000
				0,0147

Примечание. Индекс совпадения персональных данных, обезличенных с применением различных ДВБ, составил 0,0147.

Полученные данные представлены в виде табл. 3.

Таблица 3

Сравнение полученных результатов

Тип данных	Индекс совпадений
Исходные	0,0398
Сложная замена	0,0147
Естественный язык (русский)	0,053
Случайный текст ($1/n$, где $n = 120$)	0,0083

Примечание. Полученные результаты показали снижение индекса совпадения более чем в 2,5 раза по сравнению с исходными персональными данными.

4. Заключение

Разработанная структурная схема процедуры обезличивания персональных данных, основанная на использовании различных ДВБ, показала свою эффективность и практическую применимость.

Результатом исследования стало количественное подтверждение эффективности метода через оценку степени разрушения статистики естественного языка персональных данных. Применение индекса совпадения Фридмана в качестве критерия оценки оказалось оправданным, поскольку данный показатель обладает высокой чувствительностью к статистическим свойствам текста и позволяет объективно измерять степень приближения обезличенных персональных данных к характеристикам случайного текста.

Полученные экспериментальные результаты демонстрируют снижение индекса совпадения Фридмана с 0,0398 для исходных данных до

0,0147 после применения процедуры обезличивания, что составляет уменьшение более чем в 2,5 раза. Это свидетельствует о существенном разрушении статистики естественного языка. Показательным является сравнение с эталонными значениями: полученный показатель (0,0147) значительно отличается от характерного для русского языка значения (0,053) и приближается к теоретическому значению для случайного текста с равномерным распределением (0,0083).

Библиографические ссылки

1. *Ворона В. А.* Биометрическая идентификация личности. М. : Горячая линия-Телеком, 2023.
2. *Коллинз М.* Защита сетей. Подход на основе анализа данных. М. : ДМК Пресс, 2020.
3. *Остапенко Г. А.* Информационные операции и атаки в социотехнических системах: организационно-правовые аспекты противодействия. М. : Горячая линия-Телеком, 2020.
4. Об изменении приказов Оперативно-аналитического центра при Президенте Республики Беларусь от 28 марта 2014 г. № 26 и от 20 февраля 2020 г. № 66 : приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 10 дек. 2024 г. № 259. URL: <https://www.oac.gov.by/public/content/files/files/law/prikaz-oac/2024-259.pdf> (дата обращения: 24.02.2025).

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ГЕТЕРОГЕННОЙ ПЛАТФОРМЫ ВИРТУАЛИЗАЦИИ

В. П. Кочин¹⁾, А. В. Шанцов²⁾

¹⁾ Белорусский государственный университет,
г. Минск, Республика Беларусь, kochyn@bsu.by

²⁾ Белорусский государственный университет,
г. Минск, Республика Беларусь, downseason@mail.ru

Определена необходимость применения гетерогенных платформ виртуализации. Выделены основные преимущества в использовании гетерогенных платформ виртуализации. Рассмотрена проблематика обеспечения требуемого уровня защищенности ресурсов гетерогенной платформы. Определена необходимость применения формализованных подходов к процессам обеспечения информационной безопасности. Предложен алгоритм и математическая модель обеспечения информационной безопасности гетерогенных платформ виртуализации.

Ключевые слова: информационная безопасность; облачные вычисления; гетерогенная платформа виртуализации.

MATHEMATICAL MODEL OF ENSURING INFORMATION SECURITY OF A HETEROGENEOUS VIRTUALIZATION PLATFORM

V. P. Kochyn^{a)}, A. V. Shantsov^{b)}

^{a)} Belarusian State University,
Minsk, Republic of Belarus, kochyn@bsu.by

^{b)} Belarusian State University,
Minsk, Republic of Belarus, downseason@mail.ru

This paper justifies the necessity of employing heterogeneous virtualization platforms, highlighting their principal advantages. It examines the challenges associated with ensuring the required security level for the resources within such heterogeneous environments. The study underscores the need for formalized approaches to information security management. In response to these challenges, we propose a novel algorithm and develop a mathematical model designed to ensure the information security of heterogeneous virtualization platforms.

Keywords: information security; cloud computing; heterogeneous virtualization platform.

1. Введение

Технологии виртуализации прочно вошли в основу современной вычислительной инфраструктуры, обеспечивая эффективное использование ресурсов, снижение затрат и высокую отказоустойчивость. Вместе с тем, увеличение масштабов и сложности виртуализированных сред обуславливает потребность в формализации управленческих процессов, в особенности – механизмов обеспечения информационной безопасности (ИБ) в облачной среде. Пренебрежение регламентами значительно увеличивает риски неоптимального расходования ресурсов, снижения производительности и возникновения угроз ИБ [1].

Внедрение гетерогенных платформ виртуализации открывает широкие возможности для повышения гибкости, отказоустойчивости и общей эффективности управления ИТ-инфраструктурами. Однако для реализации всех преимуществ подобных платформ необходима тщательная формализация процессов управления, среди которых важную роль играет формализация процессов обеспечения защищенности виртуальных машин (ВМ) и платформы в целом.

2. Преимущества применения гетерогенных платформ виртуализации

Внедрение гетерогенных платформ виртуализации позволяет достичь значительного повышения гибкости, устойчивости и безопасности обработки ресурсов. Гетерогенная платформа виртуализации включает в себя две и более подсистемы виртуализации, построенные на различных гипервизорах, объединенных в единый вычислительный кластер [2]. Использование гетерогенных платформ имеет ряд существенных преимуществ, в том числе в аспекте ИБ.

1. Повышение гибкости и совместимости. Использование разнородных гипервизоров, архитектур и аппаратных решений в рамках единой инфраструктуры обеспечивает:

интеграцию унаследованных и современных информационных систем с учётом требований безопасности;

адаптацию под специфические задачи и сервисы за счёт выбора наиболее релевантных и безопасных технологических решений для каждого конкретного случая.

2. Минимизация рисков зависимости от единственного вендора (vendor lock-in). Эксплуатация множества платформ виртуализации (таких как VMware ESXi, Microsoft Hyper-V, KVM, Xen, OpenStack) способствует:

снижению зависимости от конкретного поставщика, что уменьшает риски, связанные с уязвимостями в продуктах одного вендора;

повышению гибкости при изменениях в лицензионной политике или стратегии вендора;

обеспечению возможности миграции между платформами без необходимости полной реконфигурации инфраструктуры.

3. Повышение устойчивости и отказоустойчивости. Гетерогенная архитектура снижает системные риски, связанные с отказами на уровне гипервизора или системного ПО. Так, уязвимость или сбой одной платформы не приводит к полному отказу системы, поскольку иная подсистема остаётся работоспособной. Кроме того, появляется возможность изоляции критичных подсистем на различных гипервизорах, что повышает общий уровень безопасности архитектуры и снижает риски распространения атак.

4. Поддержка безопасных гибридных и мультиоблачных стратегий. Гетерогенные платформы являются фундаментом для построения сложных сред, объединяющих частные и публичные облака, а также периферийные вычисления. Это позволяет:

размещать рабочие нагрузки (ВМ, контейнеры) в оптимальной среде в зависимости от требований к безопасности, производительности, а также требований нормативных правовых актов;

осуществлять централизованное управление и оркестрацию с помощью кросс-платформенных инструментов (Kubernetes, Terraform, Ansible, OpenStack), абстрагирующихся от особенностей нижележащих гипервизоров.

5. Усиление защиты за счёт разнородности среды. Разнообразие программных и аппаратных платформ затрудняет для злоумышленников проведение масштабных атак, поскольку требует разработки различных эксплуатационных методов для каждого типа гипервизора, что в свою очередь позволяет:

осуществить изоляцию критически важных ресурсов на отдельных гипервизорах с уникальными настройками безопасности;

снизить риски «горизонтального перемещения» злоумышленников в случае компрометации одного из компонентов среды;

реализовать принцип комплексной защиты [3], где каждая платформа может быть настроена согласно определенных требований к обеспечению ИБ.

3. Математическая модель обеспечения информационной безопасности гетерогенной платформы

Одновременное использование разнородных гипервизоров в гетерогенных платформах виртуализации формирует расширенную и многокомпонентную поверхность для потенциальных атак. Неоднородность гетерогенных платформ влечет за собой существенные риски ИБ, включая усложнение управления уязвимостями, несогласованность политик безопасности, трудности в мониторинге и аудите безопасности. Таким

образом, несмотря на все преимущества, гетерогенные платформы порождают комплексные риски ИБ.

Для управления рисками ИБ необходимо использовать математическое моделирование процессов обеспечения ИБ в гетерогенных средах, направленное на решения следующих задач.

1. Формализация и автоматизация политик безопасности. Создание строгих математических моделей позволяет перейти от эвристических правил к детерминированным алгоритмам принятия решений, что минимизирует человеческий фактор – частую причину уязвимостей и ошибок конфигурации.

2. Оптимизация обнаружения инцидентов ИБ. Моделирование позволяет выявлять характерные для кибератак аномалии в поведении ВМ и гипервизоров в условиях высокой сложности и неоднородности инфраструктуры, повышая точность и скорость детектирования угроз [4].

3. Сокращение времени реагирования на киберинциденты. Алгоритм обеспечения безопасности обеспечивает не только своевременное обнаружение, но и автоматизацию ответных мер, таких как изоляция скомпрометированной ВМ или ее миграция на защищенный узел, что критически важно для снижения ущерба [5].

4. Повышение отказоустойчивости и живучести системы. Применение различных платформ виртуализации и возможности динамической балансировки ВМ обеспечивают высокую отказоустойчивость гетерогенной платформы в целом.

Таким образом, актуальность исследований в данной области обусловлена их высокой прикладной значимостью для построения защищенной от кибератак ИТ-инфраструктуры, функционирующей на гетерогенной платформе виртуализации.

Формализация процессов обеспечения ИБ гетерогенных платформ заключается в разработке строгих детерминированных алгоритмов и критериев, регламентирующих обеспечение защищенности ресурсов платформы. Такой подход обеспечивает автоматизацию операций управления, что является необходимым условием для поддержания безопасности в масштабируемых гетерогенных системах. Критерии алгоритмического выбора интегрируют комплекс требований к обеспечению защищенности, включая данные от антивирусных систем, данные систем IPS/IDS, системы EDR, мониторинг аномалий в API-вызовах и прочее, в зависимости от конкретного ресурса.

Обозначим:

$P = \{p_1, p_2\}$ – набор платформ виртуализации;

$V = \{v_1, v_2, \dots, v_n\}$ – множество ВМ;

$R = \{AV, IDS, EDR, \dots\}$ – параметры состояния защищенности ВМ.

Каждая ВМ характеризуется вектором защищенности:

$$\vec{r}_i(t) = (r_i^{AV}(t), r_i^{IDS}(t), r_i^{EDR}(t), \dots), \quad (1)$$

где $\vec{r}_i(t)$ – вектор защищенности ВМ в момент времени t ; $r_i^{AV}(t)$, $r_i^{IDS}(t)$, $r_i^{EDR}(t)$ – значение параметров антивирусной системы, системы обнаружения вторжений и системы анализа телеметрии в момент времени t соответственно.

Зададим условия компрометации ВМ:

$$C(v_i) = \begin{cases} k, & \text{если } |\vec{r}_i(t)| \geq \delta_{trigger}^i, \\ 0, & \text{если } |\vec{r}_i(t)| < \delta_{trigger}^i, \end{cases} \quad (2)$$

где $C(v_i)$ – индикатор компрометации ВМ v_i ; $\delta_{trigger}^i$ – значение порога риска для v_i ; k – значение индикатора компрометации ВМ ($k \in N$).

Для каждой ВМ v_i платформы p_j значение порога риска $\delta_{trigger}^i$ может задаваться как согласно политик безопасности и являться равным для всех ВМ, так и различаться внутри одной платформы, учитывая специфику ВМ. Значение индикатора компрометации k устанавливается для каждой ВМ согласно критичности обрабатываемых ресурсов и влияния компрометации ВМ на другие ВМ и платформу в целом.

Определим значение уровня риска платформы виртуализации p_j в момент времени t :

$$S_j(t) = \sum_{i=1}^n C(v_i), \quad (3)$$

где $S_j(t)$ – значение уровня риска платформы p_j ; n – количество ВМ на платформе p_j .

Для каждой платформы виртуализации в соответствии с требованиями политик безопасности определяется пороговое значение риска платформы $\theta_{security}^j$ в зависимости от которого определяется текущее состояние защищенности платформы виртуализации p_j в целом.

В зависимости от соотношения $S_j(t)$ и $\theta_{security}^j$ определяются дальнейшие меры по поддержанию состояния защищенности платформы p_j согласно алгоритму обеспечения безопасности. Алгоритм предусматривает проведения независимой оценки состояния защищенности для каждой платформы p_j , входящей в состав гетерогенной платформы виртуализации. Алгоритм обеспечения ИБ представлен на рисунке и включает следующие этапы.

1. На первом этапе осуществляется расчет векторов защищенности $\vec{r}_i(t)$ для каждой ВМ v_i согласно формуле (1).

2. Согласно формуле (2) рассчитываются значения индикаторов компрометации ВМ $C(v_i)$.

3. При наличии одного и более $C(v_i) > 0$ соответствующие ВМ считаются скомпрометированными, в дальнейшем необходимо выполнить расчет уровня риска платформы и осуществить миграции ВМ. В противном случае осуществляется возврат к шагу 1.

4. Рассчитывается уровень риска $S_j(t)$ для платформы виртуализации p_j согласно формуле (3).

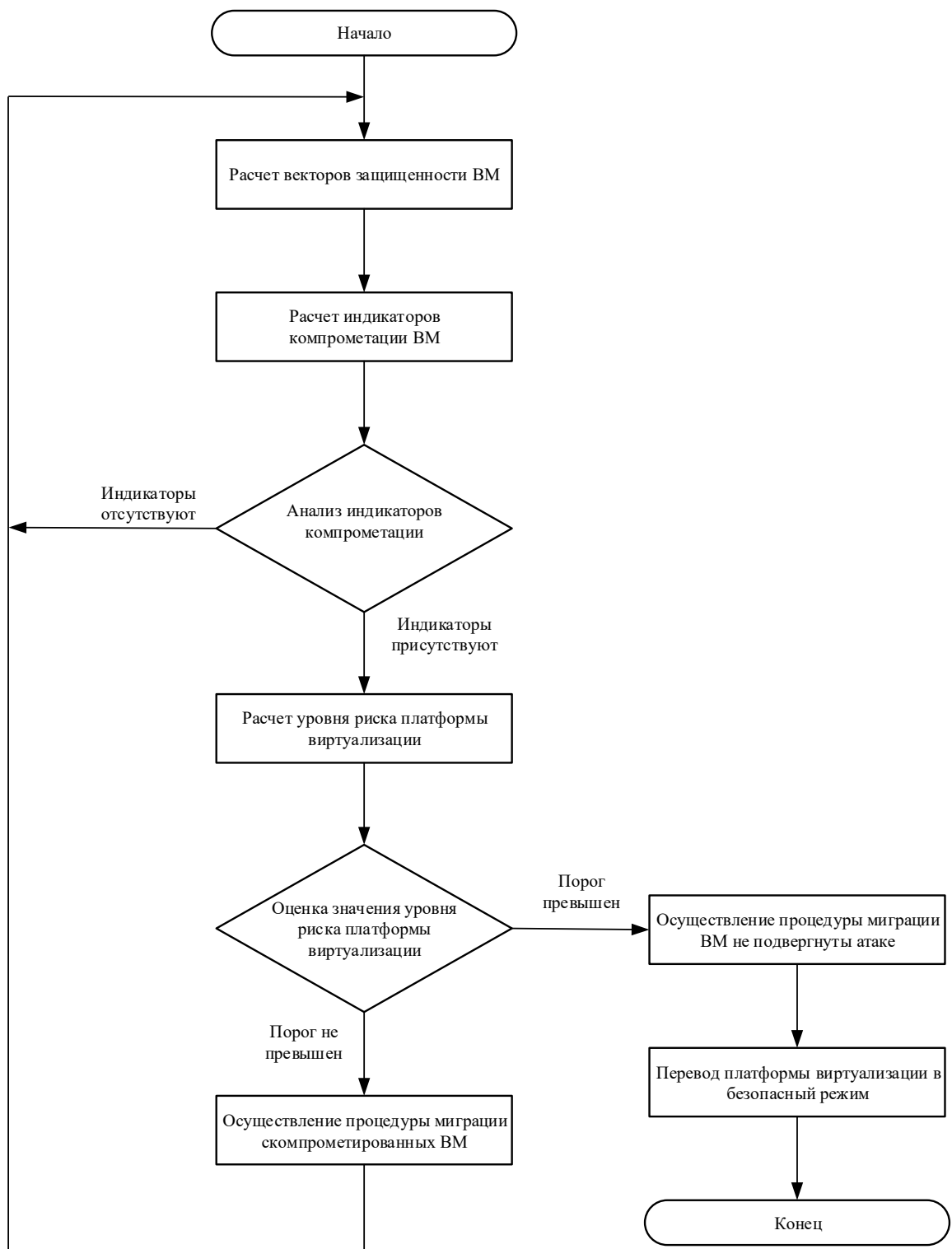
5. Рассчитанный уровень риска $S_j(t)$ сравнивается с установленным пороговым значением риска платформы $\theta_{security}^j$.

6. В случае если пороговое значение не превышено, т.е. $S_j(t) < \theta_{security}^j$, выполняются процедуры по изоляции с последующим переносом скомпрометированных ВМ [6]. Платформа p_j остается в работоспособном состоянии. Осуществляется возврат к шагу 1 алгоритма.

7. В случае если пороговое значение достигнуто или превышено, т.е. $S_j(t) \geq \theta_{security}^j$, платформа считается скомпрометированной. Осуществляется миграция ВМ, не подвергнутых кибератаке (индикатор компрометации $C(v_i) = 0$) на нескомпрометированную платформу виртуализации.

8. Скомпрометированная платформа переводится в защищенный режим: осуществляется изоляция сетевой инфраструктуры платформы; запрещается запуск новых ВМ на платформе; снимается дамп событий платформы виртуализации; для ресурсов платформы устанавливается режим доступа «только чтение».

9. Скомпрометированная платформа предоставляется подразделению кибербезопасности [7] для осуществления процедур реагирования на киберинциденты. После устранения последствий кибератаки платформа возвращается к нормальному функционированию.



Алгоритм обеспечения ИБ гетерогенной платформы виртуализации

4. Заключение

В данной статье продемонстрированы ключевые преимущества внедрения гетерогенных платформ виртуализации, их потенциал для повышения гибкости, отказоустойчивости и безопасности современной ИТ-инфраструктуры. Проведенный анализ показывает, что интеграция разнородных гипервизоров в единую гетерогенную платформу позволяет минимизировать риски зависимости от вендора, поддерживает реализацию гибридных и мультиоблачных стратегий и создает существенные препятствия для кибератак за счет разнородности среды.

Вместе с тем присущая таким платформам архитектурная сложность и расширенная поверхность атаки формируют значительные риски ИБ. Для нейтрализации данных рисков в работе аргументирована необходимость применения формализованного, модельно-ориентированного подхода к управлению ИБ. Предложенные математическая модель и алгоритм формируют основу для автоматизации процессов управления ИБ. Практическая значимость исследования заключается во вкладе в построение защищенных от кибератак ИТ-инфраструктур, функционирующих на основе гетерогенных платформ.

Библиографические ссылки

1. *Кочин В. П., Шанцов А. В.* Комплексная система защиты информации облачных ресурсов // Комплексная защита информации : материалы XXVI научно-практической конференции. Минск, 2021. С. 332–334.
2. *Кочин В. П., Шанцов А. В.* Гибридный режим функционирования облачных платформ // Информационно-коммуникационные технологии. Достижения, проблемы, инновации : материалы II научно-практической конференции, Полоцк, 30–31 марта 2022 г. / Полоцк : Полоцкий государственный университет. 2022. С. 189–193.
3. *Кочин В. П., Шанцов А. В.* Особенности построения комплексной системы защиты информации облачных ресурсов // Научная конференция студентов и аспирантов Белорусского государственного университета : материалы 78-й научной конференции студентов и аспирантов. Минск : БГУ, 2021. – С. 103–106.
4. Handbook for Computer Security Incident Response Teams (CSIRTs) / M. West-Brown [et al.]. Pittsburgh, PA : Carnegie Mellon Software Engineering Institute, 2003. URL: <http://resources.sei.cmu.edu/library/asset-view.cfm?assetid=6305> (date of access: 08.09.2025).
5. *Zimmerman C.* Ten strategies of a world-class cybersecurity operations center. Bedford : MITRE, 2014.
6. *Алексанков С. М.* Модель динамической миграции виртуальных машин с гибридным подходом // Научно-технический вестник информационных технологий, механики и оптики. 2017. № 4. С. 725–732.
7. *Кочин В. П., Шанцов А. В.* Методика создания и структура корпоративного подразделения информационной безопасности // Цифровая трансформация. 2022. № 3. С. 65–72.

О МОДУЛЯРНОМ РАЗДЕЛЕНИИ СЕКРЕТА В СПЕЦИАЛЬНОЙ ЛИНЕЙНОЙ ГРУППЕ

Г. В. Матвеев^{1), 2)}, А. А. Осиновская²⁾

¹⁾ Белорусский государственный университет,
Минск, Беларусь, Matveev@bsu.by

²⁾ Институт математики НАН Беларуси,
Минск, Беларусь, anna@im.bas-net.by

Модулярное разделение секрета в группе $SL_2(\mathbb{Z})$ было предложено в работе [1]. В настоящее время нами построена в явном виде вся фундаментальная область при действии левыми сдвигами главной конгруэнц-подгруппы на группе $SL_2(\mathbb{Z})$, что представляет дополнительные возможности для построения схем, так как эта область является пространством хранимых секретов схемы разделения секрета.

Ключевые слова: специальная линейная группа; конгруэнц-подгруппа; фундаментальная область.

ON MODULAR SECRET SHARING IN THE GROUP $SL_2(\mathbb{Z})$

G. V. Matveev^{a), b)}, A. A. Osinovskaya^{b)}

^{a)} Belarusian State University,
Minsk, Belarus, Matveev@bsu.by

^{b)} Institute of Mathematics of the National Academy of Sciences of Belarus,
Minsk, Belarus, anna@im.bas-net.by

Modular secret sharing in the group $SL_2(\mathbb{Z})$ was proposed in [1]. In this paper we have constructed in explicit form the entire fundamental domain under the action of left shifts of the principal congruence subgroup on the group $SL_2(\mathbb{Z})$, which presents additional possibilities for constructing schemes, since the domain is the space of stored secrets of the secret sharing scheme.

Keywords: special linear group; congruence subgroup; fundamental domain.

Модулярное разделение секрета в группе $SL_2(\mathbb{Z})$ было недавно предложено в работе [1]. В качестве пространства ключей и хранимого ключа используется фундаментальная область при действии левыми сдвигами главной конгруэнц-подгруппы на группе $SL_2(\mathbb{Z})$. Однако в [1] построена лишь часть этой области, но пригодная для реализации алгоритма разделения секрета.

Попытки построить всю фундаментальную область предпринимались и ранее [2], вне связи с разделением секрета, т.е. эта задача интересна и чисто с алгебраической точки зрения. Уместно будет напомнить, что в книге [3, с. 438–439] отмечается нетривиальный характер задачи подъема решений уравнения $xy - zt = 1$ по некоторой системе модулей до целочисленного решения. С нашей точки зрения задача подъема решений – это в точности алгоритм восстановления секрета. Вместе с тем, в работах [1, 4] показано, что знание фундаментальной области значительно облегчает восстановление секрета.

В настоящей работе нами построена в явном виде вся фундаментальная область при действии левыми сдвигами главной конгруэнц-подгруппы на группе $SL_2(\mathbb{Z})$, что представляет дополнительные возможности для разделения секрета в этой группе.

Пусть $m \neq 1$ – целое положительное число. Напомним определения конгруэнц-подгрупп в группе $SL_2(\mathbb{Z})$:

$$\Gamma_0(m) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \pmod{m} \right\},$$

$$\Gamma(m) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{m} \right\}.$$

Здесь и далее сравнимость матриц по модулю m означает их поэлементную сравнимость. Известно, что $\Gamma_0(m)$ и $\Gamma(m)$ действительно подгруппы и справедливы включения

$$\Gamma(m) \subset \Gamma_0(m) \subset SL_2(\mathbb{Z}).$$

Подгруппа $\Gamma(m)$ называется главной конгруэнц-подгруппой по модулю m .

Главная конгруэнц-подгруппа $\Gamma(m)$ действует левыми сдвигами на группах $\Gamma_0(m)$ и $SL_2(\mathbb{Z})$. Система представителей орбит называется фундаментальной областью. Две матрицы A и A' принадлежат одной орбите, если $A\Gamma(m) = A'\Gamma(m)$. Это равенство удобно трактовать иначе.

Лемма 1. [1] Условие $A\Gamma(m) = A'\Gamma(m)$ эквивалентно условию $A \equiv A' \pmod{m}$.

В свою очередь группа $\Gamma_0(m)$ действует левыми сдвигами на группе $SL_2(\mathbb{Z})$.

Лемма 2. Для матриц

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, A' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in SL_2(\mathbb{Z})$$

условие $A\Gamma_0(m) = A'\Gamma_0(m)$ эквивалентно условию $ac' \equiv a'c \pmod{m}$.

Лемма 3. Пусть число m составное и $m = aa_1$ – некоторое его разложение на нетривиальные множители. Для любого целого числа c , $0 < c \leq a_1$, $(c, a, a_1) = 1$, существует единственное целое число $c_1 \leq m / (a, a_1)$ такое, что $(c_1, a) = 1$ и $c_1 \equiv c \pmod{a_1}$.

Следствие 1. Если $m = p^k$, где p простое, то $c_1 = c$.

Рассмотрим произвольный составной модуль m . Для любой пары (a, c) , где a – нетривиальный делитель числа m , $m = aa_1$, $0 < c \leq a_1$, $(c, a, a_1) = 1$, определим матрицу

$$M(a, c) = \begin{pmatrix} a & b \\ c_1 & d \end{pmatrix} \in SL_2(\mathbb{Z}),$$

где c_1 такое же, как в лемме 3. Такие матрицы существуют (хотя и определены неоднозначно). Действительно, поскольку $(a, c_1) = 1$, то можно найти целочисленное разложение $az + c_1w = 1$ НОДа. Положив $d := z$ и $b := -w$, получаем искомую матрицу. Для каждой пары (a, c) фиксируем одну такую матрицу.

Лемма 4. Если $(a, c) \neq (a', c')$, то $M(a, c)\Gamma_0(m) \neq M(a', c')\Gamma_0(m)$.

Положим

$$M = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, N(c) = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$$

и определим множество

$$\mathcal{M} = \{M, N(c), 0 \leq c < m, M(a, c), m = aa_1, a, a_1 \neq 1, 0 < c \leq a_1, (c, a, a_1) = 1\}.$$

Теорема 1. Пусть m – некоторый целочисленный модуль. Тогда множество $\mathcal{M}$ составляет фундаментальную область группы $SL_2(\mathbb{Z})$ относительно подгруппы $\Gamma_0(m)$.

Укажем теперь представителей смежных классов группы $SL_2(\mathbb{Z})$ по подгруппе $\Gamma(m)$. Пусть $\varphi(m)$ – количество натуральных чисел $< m$ и взаимно простых с ним (функция Эйлера). Имеем

$$\varphi(m) = m \prod_{p|m} \left(1 - \frac{1}{p}\right).$$

Обозначим все такие числа через $\varepsilon_i \in \mathbb{Z}$, $1 \leq i \leq \varphi(m)$. Тогда для любого ε_i существует число $\varepsilon'_i \in \mathbb{Z}$ такое, что

$$\varepsilon_i \varepsilon'_i = 1 + mk_i$$

и ε_i, k_i однозначно определяются выбором ε_i . Введем матрицы

$$B_{i,j} = \begin{pmatrix} \varepsilon_i + jm & k_i + j\varepsilon'_i \\ m & \varepsilon'_i \end{pmatrix},$$

где $i = 1, \dots, \varphi(m)$, $j = 0, \dots, m-1$,

$$C_{i,j,l} = \begin{pmatrix} \varepsilon_i + jm & k_i + j\varepsilon'_i \\ m + (\varepsilon_i + jm)l & \varepsilon'_i + (k_i + j\varepsilon'_i)l \end{pmatrix},$$

где $i = 1, \dots, \varphi(m)$, $j, l = 0, \dots, m-1$, и

$$D_{i,j} = \begin{pmatrix} -m & -\varepsilon'_i \\ \varepsilon_i + jm & k_i + j\varepsilon'_i \end{pmatrix},$$

где $i = 1, \dots, \varphi(m)$, $j = 0, \dots, m-1$. Положим

$$\begin{aligned} \mathcal{B} &= \{B_{i,j} \mid i = 1, \dots, \varphi(m), j = 0, \dots, m-1\}, \\ \mathcal{C} &= \{C_{i,j,l} \mid i = 1, \dots, \varphi(m), j, l = 0, \dots, m-1\}, \end{aligned}$$

$$\mathcal{D} = \{D_{i,j} \mid i = 1, \dots, \varphi(m), j = 0, \dots, m-1\}.$$

Лемма 4. Пусть m – произвольный модуль.

- (i) $\mathcal{B} \subset \Gamma_0(m)$, $\mathcal{C}$ и $\mathcal{D} \subset SL_2(\mathbb{Z})$.
- (ii) Матрицы $C_{i,j,l}$ и $D_{i,j}$ попарно несравнимы по модулю m .
- (iii) Множество $\mathcal{B}$ составляет фундаментальную область группы $\Gamma_0(m)$ относительно подгруппы $\Gamma(m)$.

Теорема 2. Если $m = p$, то множество $\mathcal{C} \cup \mathcal{D}$ составляет фундаментальную область группы $SL_2(\mathbb{Z})$ относительно подгруппы $\Gamma(p)$.

Тем самым для простого модуля получены представители фундаментальной области в виде, удобном для реализации в схеме разделения секрета.

Теорема 3. Пусть m – некоторый целочисленный модуль. Тогда матрицы AB для $A \in \mathcal{M}$ и $B \in \mathcal{B}$ составляют фундаментальную область группы $SL_2(\mathbb{Z})$ по подгруппе $\Gamma(m)$.

Приведем в явном виде пример фундаментальной области для $m = 6$. Согласно предложению 2.1.1 из [5]

$$[\Gamma_0(6) : \Gamma(6)] = 12, [SL_2(\mathbb{Z}) : \Gamma_0(6)] = 12.$$

Имеем

$$\mathcal{B} = \left\{ \begin{pmatrix} 1+6j & j \\ 6 & 1 \end{pmatrix}, \begin{pmatrix} 5+6j & 4+5j \\ 6 & 5 \end{pmatrix}, j = 0, \dots, 5 \right\}.$$

$$\mathcal{M} = \left\{ \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}, c = 0, \dots, 5, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}, \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \right\}.$$

Тогда по теореме 3 фундаментальная область состоит из произведений матриц из этих множеств.

На основе найденной фундаментальной области можно построить схему разделения секрета с бóльшим пространством хранимых ключей, чем в работе [1]. Для этого удобно использовать матрицы $\mathcal{C}$, имеющие хорошую параметризацию. Они занимают большую часть фундаментальной области группы $SL_2(\mathbb{Z})$ относительно подгруппы $\Gamma(m)$, поскольку их количество равно $m^2\varphi(m)$, а при этом [5]

$$[SL_2(\mathbb{Z}) : \Gamma(m)] = m^2 \varphi(m) \prod_{p|m} \left(1 + \frac{1}{p}\right).$$

Работа выполнена при финансовой поддержке БРФФИ, договор № Ф25-012.

Библиографические ссылки

1. Янчевский В. И., Говорушко И. О., Матвеев Г. В. Разделение секрета в специальной линейной группе // Информатика. 2024. Т. 21, № 3. С. 23–31.
2. Harcos G. CEU lecture notes: a set of representatives for $\Gamma_0(q) \backslash SL_2(\mathbb{Z})$. URL: [https://users.renyi.hu/~gharcos/CEU_Gamma0\(q\).pdf](https://users.renyi.hu/~gharcos/CEU_Gamma0(q).pdf) (date of access: 25.08.2025).
3. Платонов В. П., Рапинчук А. С. Алгебраические группы и теория чисел. М. : Наука, 1991.
4. Матвеев Г. В., Осинковская А. А., Янчевский В. И. Фундаментальная область в специальной линейной группе $SL_2(\mathbb{F}_p[x])$ и схема разделения секрета на ее основе // Труды Института математики НАН Беларуси. 2024. Т. 32, № 2. С. 7–16.
5. Di Matteo G. The action of $SL_2(\mathbb{Z})$ on the upper-half complex plane. URL: <https://www.dimatteo.is/Mathematics/Courses/Modular-forms/02-SL2Z.pdf> (date of access: 25.08.2025).

ПОРТИРОВАНИЕ КРИПТОГРАФИЧЕСКИХ БИБЛИОТЕК С ПОМОЩЬЮ SWIG

М. Н. Мицкевич

*НИИ прикладных проблем математики и информатики,
Белорусский государственный университет,
Минск, Беларусь, mitskevichmn@bsu.by*

Приводятся результаты анализа инструментов высокоуровневого портирования. Описан пример портирования криптографической библиотеки Bce2 на язык C#.

Ключевые слова: портирование; обвязка; генерация кода; нативная функция.

CRYPTOGRAPHIC LIBRARIES PORTING WITH SWIG

M. N. Mitskevich

*Research Institute for Applied Problems of Mathematics and Informatics,
Belarusian State University,
Minsk, Belarus, mitskevichmn@bsu.by*

Results of the investigation of tools for high-level porting are considered. An example of cryptographic library porting to C# language is described.

Keywords: porting; wrapper; code generation; native function.

1. Введение

Большинство криптографических инструментов написано на языках Си и C++. Не является исключением и реализация стандартизированных в Республике Беларусь криптографических алгоритмов в библиотеке Bce2 [1], которая распространяется под открытой и свободной лицензией Apache License, Version 2.0.

Предыдущий опыт встраивания библиотеки Bce2 в решения, написанные на Java, показал трудоёмкость этого процесса и выявил наличие проблем совместимости. Сейчас актуальной является задача применения белорусской криптографии в решения, написанные на других языках, что явилось причиной проведения анализа инструментов портирования.

Было потрачено много сил и времени на написание, проверку и тестирование исходного кода реализации алгоритмов, поэтому переписывание

с нуля на каждом новом языке программирования изначально не рассматривалось.

Также хотелось покрыть максимально возможное количество языков и платформ, поэтому не рассматривались специализированные библиотеки, рассчитанные на портирование кода под определенный язык программирования, например, `ctypes` для Python.

Мы хотим получить с минимально возможными усилиями модуль на новом языке программирования (целевом), который будет вызывать функции библиотеки Bce2 привычным для целевого языка способом, используя обвязки для функций языка Си.

На наш взгляд, единственный инструмент, который удовлетворяет заявленным критериям отбора, доступен, развивается и поддерживает большое число целевых языков программирования – SWIG.

2. Обзор SWIG

SWIG (англ. *simplified wrapper and interface generator*) — свободный инструмент для связывания программ и библиотек, написанных на языках C и C++, с интерпретируемыми (Tcl, Perl, Python, Ruby, PHP) или компилируемыми (Java, C#, Scheme, OCaml) языками [2].

Основная цель — обеспечение возможности вызова функций, написанных на C и C++, из кода на других языках для расширения функциональности или повышения производительности.

Программист создаёт файл интерфейсов `.i` с описанием экспортируемых функций, а SWIG генерирует исходный код для склеивания C/C++ и нужного языка. В результате генерируется файл интерфейса для подключения существующей библиотеки или исходный код библиотеки функций-обвязок над функциями C/C++.

SWIG написан на языках C и C++, распространяется по лицензии, похожей на BSD, с февраля 1996 года. Лицензия SWIG позволяет использовать, распространять и модифицировать код SWIG для коммерческих и некоммерческих целей практически без ограничений.

3. Подключение Bce2

Рассмотрим пример вызова функций выработки детерминированной электронной цифровой подписи (ЭЦП) из библиотеки Bce2 в языке C#. Для этого мы должны включить модуль `bign` с функцией выработки ЭЦП, модуль `belt` с функцией хэширования и дополнительные модули для выделения памяти `mem`, для преобразования 16-ричных чисел `hex`, а также модуль `safe`, от которого зависят `bign` и `belt`.

Основной для SWIG файл `bee2.i` начинается с заголовочной секции. В ней прописывается директива `%module`, в которой указано название целевого модуля. Также подключаются модули стандартной библиотеки SWIG с помощью директивы `%include`, например, инструкция `%include "carrays.i"` подключает функции управления массивами.

```
%module bee2
#include "carrays.i"
#include "cdata.i"
#include "cpointer.i"
#ifdef SWIGPYTHON
#include "cstring.i"
#endif
#include "typemaps.i"
```

Далее следует секция включения `%{ ... %}`. Она позволяет включать в код обвязки необходимые для его компиляции заголовочные файлы, а также дополнительные функции, необходимые для корректной работы функций в целевом языке программирования. Например, для выполнения тестов нам потребовалась операция сдвига указателя на некоторое фиксированное смещение, а многие высокоуровневые языки прямо запрещают это делать. Поэтому мы добавили функцию `ptradd`, которая принимает указатель и смещение и возвращает новый смещенный указатель. В итоге секция включения в нашем примере выглядит так:

```
%{
#include "bee2/core/safe.h"
#include "bee2/core/mem.h"
#include "bee2/core/hex.h"

#include "bee2/crypto/belt.h"
#include "bee2/crypto/bign.h"

void* ptradd(void* ptr, int offset) {
    return ptr + offset;
}
%}
```

Далее следует секция объявления типов, массивов и структур, необходимых для того, чтобы SWIG понимал типы аргументов при разборе функций.

```
typedef unsigned int u32;
typedef u32 err_t;
```

```

#define ERR_OK      ((err_t)0)

%array_functions(u32, u32arr)
%apply SWIGTYPE* { size_t* count };
%apply SWIGTYPE* { size_t *ary };
%array_functions(size_t, sizeTarr)

%pointer_cast(octet*, void*, op2vp)
%pointer_cast(void*, octet*, vp2op)
%pointer_cast(unsigned char*, void*, bp2vp)
%pointer_cast(void*, unsigned char*, vp2bp)
%pointer_cast(void*, char*, vp2cp)

```

В самом конце идёт секция разбора. В ней содержатся инструкции `%include`, которые определяют заголовочные файлы, к функциям которых SWIG должен сгенерировать обвязки. Обычно все эти файлы должны быть указаны в секции включения, но для модуля `bee2/core/hex.h` генерируемые по умолчанию интерфейсы были неработоспособны, поэтому вместо подключения модуля пришлось явно объявить функции `hexTo` и `hexFrom`.

Кроме того, тут должны быть объявлены функции, реализация которых находится в секции включения (у нас это функция `ptradd`), чтобы к ним тоже были сгенерированы обвязки, и они стали доступными в целевом языке программирования.

```

#include "../include/bee2/core/safe.h"
#include "../include/bee2/core/mem.h"
#include "../include/bee2/crypto/belt.h"
#include "../include/bee2/crypto/bign.h"

void* ptradd(void* ptr, int offset);

void hexTo(void* dest, const char* src);

#ifdef SWIGPYTHON
%cstring_mutable(char* dest)
void hexFrom(char* dest, const void* src, size_t
count);
#else
%typemap(freearg) char * { }
%apply SWIGTYPE* { char *dest };

```

```
void hexFrom(char* dest, const void* src, size_t
count);
#endif
```

Следует отметить возможность условной компиляции, которая позволяет задавать разные правила генерации обвязок для разных языков программирования. Так, в примере выше видно, что с помощью библиотеки `cstring.i` получается более простое объявление указателя на изменяемую (mutable) строку. Так как эта библиотека SWIG реализована только для языка Python, то в других языках приходится использовать более сложные директивы библиотеки `typemaps.i`.

Также условная компиляция помогает при необходимости добавления функций, необходимых только в определенном языке. Так, например, в языке C# возникла проблема с получением константной строки символов из изменяемой, так как невозможно задать преобразование из типа `char*` в два разных типа C# (`string` и `SWIGTYPE_p_char`). Обойти это получилось, задавая все буферы памяти как тип `void*` и преобразовывая указатели при необходимости либо в строку с помощью конвертера `vp2cp`, либо в массив символов через следующую функцию:

```
#ifdef SWIGCSHARP
%{
char* getStr(void *dest) { return dest; }
%}
%apply SWIGTYPE* { char* };
char* getStr(void *dest);
#endif
```

4. Генерация библиотеки обвязок

В результате компиляции входного файла в SWIG создается готовая библиотека для языка C#, в которой нам важно выделить файл с обвязками `bee2_wrap.c` и главный модуль библиотеки `bee2cs.cs`.

Для всех функций из заданных заголовочных файлов библиотеки генерируется обвязка на стороне Си, реализующая интерфейс, пригодный для импорта в C#, и обвязка на стороне C#, которая скрывает детали реализации импорта функций за интерфейсом, соответствующим исходным функциям библиотеки. Например, для компилятора Си важен тип данных, на которые ссылается указатель.

Для C# также важны логические типы данных. В то время как реализация всех типов указателей одинакова, что учтено интерфейсом импорта нативных функций в C#. Поэтому следующая функция принимает

нетипизированный указатель, затем преобразует его в тип параметра функции `getStr` (также нетипизированный указатель), выполняет с помощью этой функции его преобразование в типизированный указатель на строку символов, затем обратно преобразуется в нетипизированный указатель для передачи в `C#`, где он будет преобразован в типизированный указатель `SWIGTYPE_p_char`, связанный со строками и позволяющий менять их содержимое.

```
SWIGEXPORT void * SWIGSTDCALL
CSharp_bcrypto_getStr(void * jarg1) {
    void * jresult ;
    void *arg1 = (void *) 0 ;
    char *result = 0 ;
    arg1 = (void *)jarg1;
    result = (char *)getStr(arg1);
    jresult = (void *)result;
    return jresult;
}
```

В файл с обвязками `bee2_wrap.c` включаются ещё и дополнительные функции, объявленные в файле `bee2.i`:

```
char* vp2cp(void* x) {
    return (char*) x;
}
char* getStr(void *dest) {
    return dest;
}
```

Параметр `SWIG -outfile bee2cs.cs` позволяет не разделять код `C#` на отдельные модули, а собрать весь код в одном файле, в котором содержатся все классы, соответствующие типам `Си`, а также класс библиотеки и класс нативных функций, о которых скажем отдельно.

В классе нативных функций `bee2PINVOKE` прописаны интерфейсы для сгенерированных нативных функций, выполняющих преобразование имен и типов во внутренние типы и имена библиотеки `C#`. Для каждой функции прописана точка входа и интерфейс функции, который она получает в `C#`.

```
[global::System.Runtime.InteropServices.DllImport(
    "bee2wrap", EntryPoint="CSharp_bcrypto_getStr")]
public static extern global::System.IntPtr getStr(
    global::System.Runtime.InteropServices.HandleRef
    jarg1);
```

По умолчанию имя библиотеки, из которой производится импорт функции, равно имени модуля SWIG. Имя модуля удобно задавать как имя исходной библиотеки, так как для имени класса для представления библиотеки в целевом языке также используется имя модуля. Для того, чтобы переопределить это имя на имя библиотеки с обвязками, используется параметр `SWIG -dllimport bee2wrap`, специфичный только для языка C#.

Основной класс `bee2` содержит набор статических функций, соответствующих функциям из заголовочных файлов, указанных в секции разбора.

В этих функциях вызываются обвязки, реализованные в Си-библиотеке, а также выполняется приведение типов, локализирующее использование нативных указателей внутри класса.

```
SWIGTYPE_p_char getStr(SWIGTYPE_p_void dest) {  
    global::System.IntPtr cPtr =  
    bee2PINVOKE.getStr(SWIGTYPE_p_void.getCPtr(dest));  
    SWIGTYPE_p_char ret =  
    (cPtr == global::System.IntPtr.Zero) ? null :  
    new SWIGTYPE_p_char(cPtr, false);  
    return ret;  
}
```

5. Особенности портирования

В процессе портирования библиотеки `Bee2` было отмечено несколько проблем портирования.

1. Преобразование типов данных.

Базовые типы Си хорошо переводятся в типы других языков. Проблемы возникают там, где в самом Си допускается разные варианты как размера типов (как в `size_t`, где размер беззнакового целого зависит от платформы), так и содержимого (как в `char`, который используется и для строк, и для целых чисел).

2. Перенос семантики типов данных.

Рассмотрим следующий пример:

```
typedef unsigned char u8;  
typedef u8 octet;
```

При таком введении новых типов через переопределение старых в целевом языке программирования теряется связь между типами, поэтому следует добавлять явные функции преобразования между типами, а также между указателями на эти типы. Кроме того, нужно указать преобразование из нетипизированного указателя в типизированный для каждого типа, который используется в заголовках функций. Например, минимальный набор функций для преобразования однобайтных массивов:

```
%pointer_cast(octet*, void*, op2vp)
%pointer_cast(void*, octet*, vp2op)
%pointer_cast(unsigned char*, void*, bp2vp)
%pointer_cast(void*, unsigned char*, vp2bp)
%pointer_cast(void*, char*, vp2cp)
```

3. Распространение нативных библиотек вместе с C#.

При компиляции библиотеки обвязок мы используем команду, которая размещает библиотеку обвязок в специальном каталоге вида `bee2net/runtimes/$os-$platform/native/$lib.so` библиотеки C#.

Чтобы библиотека обвязок копировалась вместе с библиотекой C# во все проекты, использующие её, нужно в файл проекта библиотеки C# `bee2net.csproj` добавить специальную инструкцию копирования содержимого каталога `runtimes` вместе с библиотекой [3].

Кроме того, нужно добавлять в каждую программу на C# вызов загрузчика, который будет выбирать подходящую библиотеку для каждой платформы.

6. Заключение

SWIG позволяет быстро интегрировать библиотеки Си и C++ во многие языки программирования. Это может быть полезно как для быстрого прототипирования, так и для создания работоспособного макета для полноценного портирования библиотеки.

Предполагается, что описанный подход позволит расширить использование функций защиты информации в программном обеспечении, а также увеличит популярность новых криптоалгоритмов.

Примеры портирования библиотеки Bee2 на C# и другие языки программирования доступны в репозитории `bee2ports` [4].

Библиографические ссылки

1. Agievich S. Bee2 // GitHub. URL: <https://github.com/agievich/bee2> (date of access: 05.08.2025).
2. SWIG. URL: <https://www.swig.org/> (date of access: 05.08.2025).
3. Where to put Native libraries inside a Project and how to reference them? // GitHub. URL: <https://github.com/dotnet/runtime/issues/11404> (date of access: 08.08.2025).
4. Bee2ports // GitHub. URL: <https://github.com/bcrypto/bee2ports/> (date of access: 11.08.2025).

ОБ АКТУАЛЬНЫХ ЗАДАЧАХ ОЦЕНКИ НАДЕЖНОСТИ КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ

В. Ю. Палуха, М. В. Мальцев, Ю. С. Харин, И. К. Пирштук

*НИИ прикладных проблем математики и информатики,
Белорусский государственный университет,
Минск, Беларусь, Palukha@bsu.by, Maltsew@bsu.by, Kharin@bsu.by, Pirshtuk@bsu.by*

Криптографические генераторы являются критически важными элементами систем защиты информации. Использование уязвимых генераторов снижает надежность таких систем и приводит к компрометации конфиденциальной информации. В связи с этим необходим тщательный анализ генераторов на соответствие установленным требованиям. В статье рассматриваются подходы к оценке надежности генераторов с использованием энтропийных характеристик и искусственных нейронных сетей.

Ключевые слова: искусственные нейронные сети; криптографический генератор; последовательность случайных чисел; статистическое тестирование; энтропия.

ON URGENT PROBLEMS OF ASSESSMENT CRYPTOGRAPHIC GENERATORS QUALITY

V. Yu. Palukha, M. V. Maltsew, Yu. S. Kharin, I. K. Pirshtuk

*Research Institute for Applied Problems of Mathematics and Informatics,
Belarusian State University,
Minsk, Belarus, Palukha@bsu.by, Maltsew@bsu.by, Kharin@bsu.by, Pirshtuk@bsu.by*

Cryptographic generators are critical elements of information security systems. Using of vulnerable generators reduces the reliability of such systems and leads to the compromise of confidential information. In this regard, a thorough analysis of generators is necessary for compliance with established requirements. The article is devoted to approaches to assessing the reliability of generators using entropy characteristics and neural networks.

Keywords: artificial neural networks; cryptographic generator; entropy; random sequence; statistical testing.

1. Введение

Надежность систем криптографической защиты информации критическим образом зависит от используемых в них генераторов случайных числовых последовательностей (криптографических генераторов). С помощью криптографических генераторов формируются ключи криптосистем,

синхропосылки, параметры криптографических алгоритмов и протоколов. Уязвимости генераторов приводят к снижению сложности подбора ключа злоумышленниками и к раскрытию конфиденциальной информации. Уязвимости генераторов проявляются в отклонениях их выходных последовательностей от модели «чистой случайности» – равномерно распределенной случайной последовательности (РРСП), элементы которой независимы в совокупности и одинаково распределены по равномерному закону. Для выявления таких отклонений традиционно используется статистическое тестирование, использующие методы статистической проверки гипотез. Известными примерами наборов (батарей) статистических тестов являются тесты NIST, тесты Diehard, тесты Д. Кнута. Однако данные батареи обладают рядом ограничений и недостатков: не фиксируется семейство альтернатив, не выявляются сложные зависимости высоких порядков, сравнительно простые генераторы могут пройти тесты [1]. В связи с этим актуальной задачей является разработка новых математических моделей и методов для анализа надежности криптографических генераторов. В статье рассматриваются подходы к решению данной задачи, использующие энтропийные характеристики, также приводится обзор публикаций, посвященный применению искусственных нейронных сетей в задачах защиты информации.

2. Применение нейронных сетей и машинного обучения для анализа надежности криптографических генераторов и смежных задач

Идеи использовать машинное обучение для решения задач защиты информации высказывались еще в начале 90-х годов Ривестом [2], но активное их применение началось в последнее десятилетие вследствие развития методов обучения и роста вычислительных мощностей. ИНС за счет использования больших объемов данных для обучения позволяют выявлять скрытые закономерности в исследуемых объектах, что позволяет применять их в задачах анализа криптографических алгоритмов. Приведем далее несколько примеров по применению ИНС в задачах защиты информации. Ряд работ посвящен криптоанализу классических шифров с применением методов машинного обучения. В работе [3] полносвязная ИНС, содержащая входной слой с 77 нейронами, два скрытых слоя с 128 и 23 нейронами и выходной слой с 23 нейронами, применялась для восстановления длины ключа шифра Виженера и продемонстрировала преимущество над традиционным используемым для этой цели методом – индексом совпадений. Авторы [4] разработали рекуррентную нейронную сеть (РНС) для дешифрования сообщений шифровальной машины «Энигма» (исследовался вариант с тремя роторами); использовалась такая разновидность

PHS, как LSTM (long short-term memory, сеть с долгой краткосрочной памятью). ИНС применяются также и для анализа криптографических алгоритмов, используемых в современных системах защиты информации. В работе [5] глубокие нейронные сети применяются для построения атак по сторонним каналам. В [6] для обнаружения вторжений в устройства интернета вещей использование сетей типа CVAE (Conditional Variational Auto-encoder – сверточный вариационный автоэнкодер) позволило добиться точности, близкой к стопроцентной. Работы [7, 8] посвящены применению ИНС в стеганографии.

Для анализа надежности криптографических генераторов применяются сверточные ИНС, рекуррентные ИНС, многослойные перцептроны [9–11]. В статье [9] сверточные ИНС используются для выявления отклонений от модели РРСП в выходных последовательностях блочных шифров Speck и Simon. Использовались две сети: одна сеть выделяла информативные признаки, вторая анализировала схожесть между последовательностями. В компьютерных экспериментах для обучения использовались наборы данных, содержащих 2500 фрагментов выходной последовательности генератора длины 32 бита каждый. Такой же набор данных формировался с использованием физического датчика случайных чисел. Задача обученной сети – различить фрагмент, полученный с помощью блочного шифра от фрагмента, полученного с помощью физического датчика. Точность различения, полученная в экспериментах на тестовом наборе данных, составила около 94%. В статье [11] рассматривалась задача оценки надежности квантового генератора случайных чисел с использованием ИНС на основе сверточных и рекуррентных слоев. ИНС обучались выявлять отклонения в выходных последовательностях генератора, вызванных погрешностями измерений, потерями при аналогово-цифровом преобразовании, электромагнитными шумами и др. Также существуют подходы к использованию ИНС для построения генераторов псевдослучайных последовательностей: в работе [12] предложен генератор на основе сети Элмана, в работе [13] предложен алгоритм поточного шифрования на основе однонаправленной функции нейронной сети Хопфилда, в [14, 15] на основе ИНС предлагаются блочный шифр и функция хэширования, которые также могут использоваться для генерации псевдослучайных данных.

3. Оценка надежности криптографических генераторов на основе энтропийных характеристик

Одним из методов оценки качества генераторов является энтропийный анализ их выходных последовательностей. Наблюдаемая двоичная последовательность делится на непересекающиеся фрагменты длины s (s -граммы), по ним вычисляется статистическая оценка энтропии, для

которой известно распределение вероятностей для РРСП, и по значению оценки принимается решение о (не)соответствии модели РРСП. Для уменьшения вероятности принятия ошибочного решения следует вычислить оценку при различных значениях s . Возникает вопрос о том, как связаны между собой оценки энтропии, вычисленные при различных значениях s , например, для s и $s + 1$. Далее приведены результаты, показывающие, что с ростом s коэффициент корреляции оценок энтропии Шеннона, вычисленных по s - и $(s + 1)$ -граммам, стремится к нулю, что свидетельствует об отсутствии линейной зависимости между оценками.

Пусть имеется двоичная последовательность $\{x_t\}$ длины T . Введем в рассмотрение гипотезу $H_0 = \{\{x_t\} \text{ является РРСП}\} = \{\{x_t\} - \text{независимые, одинаково распределенные случайные величины, } p_k = P\{x_t = k\} = p_k^0 = \frac{1}{2}, t = 1, \dots, T, k = 0, 1\}$. Пусть $T = Ms(s + 1)$, тогда $\{x_t\}$ можно разбить на $M(s + 1)$ непересекающихся s -грамм и Ms непересекающихся $(s + 1)$ -грамм.

Частотные оценки вероятностей s -грамм $\langle X_j \rangle, j = 1, \dots, M(s + 1)$, и $(s + 1)$ -грамм $\langle X'_l \rangle, l = 1, \dots, Ms$, определяются соотношениями:

$$\begin{aligned} \hat{p}_i(s) &= \frac{1}{M(s + 1)} \sum_{j=1}^{M(s+1)} \mathbf{I}\{\langle X_j \rangle = i\} = \frac{v_i}{M(s + 1)}, \quad i = 0, \dots, 2^s - 1; \\ \hat{p}_k(s + 1) &= \frac{1}{Ms} \sum_{l=1}^{Ms} \mathbf{I}\{\langle X'_l \rangle = k\} = \frac{v_k}{Ms}, \quad k = 0, \dots, 2^{s+1} - 1, \\ \mathbf{I}\{A\} &= \begin{cases} 1, & \text{если событие } A \text{ верно,} \\ 0, & \text{в противном случае.} \end{cases} \end{aligned} \quad (1)$$

Статистические оценки s - и $(s + 1)$ -мерной энтропии Шеннона на основе частотных оценок вероятностей (1) имеют вид

$$\begin{aligned} \hat{H}(s) &= - \sum_{i=0}^{2^s-1} \hat{p}_i(s) \ln \hat{p}_i(s) \in [0, s \ln 2], \\ \hat{H}(s + 1) &= - \sum_{k=0}^{2^{s+1}-1} \hat{p}_k(s + 1) \ln \hat{p}_k(s + 1) \in [0, (s + 1) \ln 2]. \end{aligned} \quad (2)$$

Для того, чтобы вычислить ковариацию оценок энтропии (2) $\text{cov}_0\{\hat{H}(s), \hat{H}(s + 1)\}$ (здесь и далее индекс «0» означает вычисление при истинной гипотезе H_0), вычислим вначале 2^{2s+1} ковариаций частотных

оценок вероятностей $d_{ik} = \text{cov}_0 \{ \hat{p}_i(s), \hat{p}_k(s+1) \}$ для всех возможных пар значений s - и $(s+1)$ -грамм (i, k) , $i = 0, \dots, 2^s - 1$, $k = 0, \dots, 2^{s+1} - 1$. Для подсчета ковариаций необходимо учитывать взаимное расположения s -грамм $\langle X_j \rangle$ и $(s+1)$ -грамм $\langle X'_l \rangle$ в последовательности. $(s+1)$ -грамма $\langle X'_l \rangle$ может пересекаться ровно с двумя s -граммами. На рис. 1 представлены возможные типы пересечений при $s = 3$.

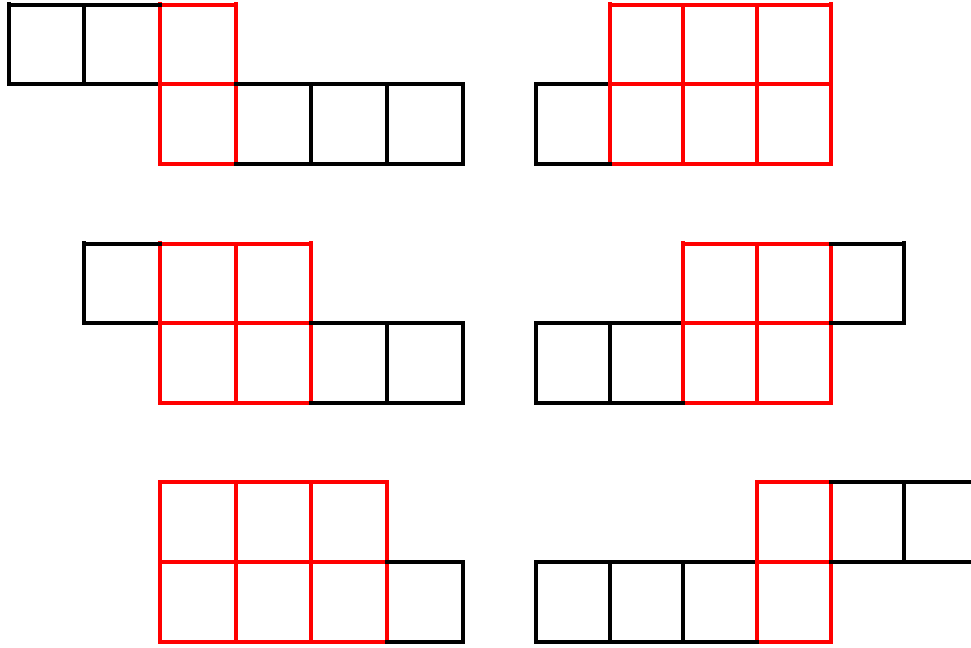


Рис. 1. Возможные типы пересечений 3- и 4-грамм

В следующей теореме, опирающейся на результаты [16], где установлено асимптотическое распределение ковариаций частотных оценок, получено совместное асимптотическое распределение оценок энтропии (2).

Теорема 1. Пусть $0 < \hat{p}_i \leq 2^{1-s}$, $i = 0, \dots, 2^s - 1$, $0 < \hat{p}_k \leq 2^{-s}$, $k = 0, \dots, 2^{s+1} - 1$. Тогда при истинной гипотезе H_0 для ковариации и коэффициента корреляции оценок энтропии $\hat{H}(s)$ и $\hat{H}(s+1)$ справедливы следующие асимптотические выражения при $M \rightarrow \infty$:

$$\text{cov}_0 \{ \hat{H}(s), \hat{H}(s+1) \} = \frac{2^{s+1} - s - 2}{M^2 s^2 (s+1)^2} + o\left(\frac{1}{M^2}\right), \quad (3)$$

$$\text{corr}_0 \{ \hat{H}(s), \hat{H}(s+1) \} = \frac{2(2^{s+1} - s - 2)}{s(s+1)\sqrt{(2^s - 1)(2^{s+1} - 1)}} + o(1). \quad (4)$$

Следствие 1. Из (4) следует, что с ростом s коэффициент корреляции стремится к нулю.

Следствие 1 обосновывает применение энтропийного профиля [16] для оценивания надежности криптографических генераторов: для принятия решения о справедливости гипотезы о равномерной распределенности наблюдаемой последовательности необходимо вычисление оценки энтропии при различных значениях длины фрагмента s .

Для демонстрации справедливости формул (3), (4) и следствия 1 проведены 2 серии компьютерных экспериментов. Рассматривалась псевдослучайная последовательность, полученная алгоритмом BelT [17] в режиме счётчика. В первой серии для $s = 2, \dots, 10$ из наблюдаемой последовательности брались $K = 1000$ фрагментов длины $T = Ms(s + 1)$ с фиксированным значением $M = 10000$. По полученным K фрагментам для каждого значения s вычислялись оценки энтропии (2), по которым затем вычислялись выборочные ковариации и коэффициенты корреляции. Также для указанных значений параметров вычислены теоретические значения ковариации и коэффициента корреляции оценок энтропии (2). На рис. 2 и 3 приведено сравнение выборочных и теоретических значений ковариации и коэффициента корреляции оценок энтропии соответственно. Во второй серии для $s = 2$ из наблюдаемой последовательности брались $K = 1000$ фрагментов длины $T = Ms(s + 1)$ со значениями M от 2000 до 16000 с шагом 2000. На рис. 4 и 5 приведено сравнение выборочных и теоретических значений ковариации и коэффициента корреляции оценок энтропии соответственно. Как видно из рисунков, полученные экспериментально значения ковариации и коэффициента корреляции оценок энтропии (2) близки к теоретическим значениям, а также первая серия экспериментов показала, что коэффициент корреляции стремится к 0 с ростом s .

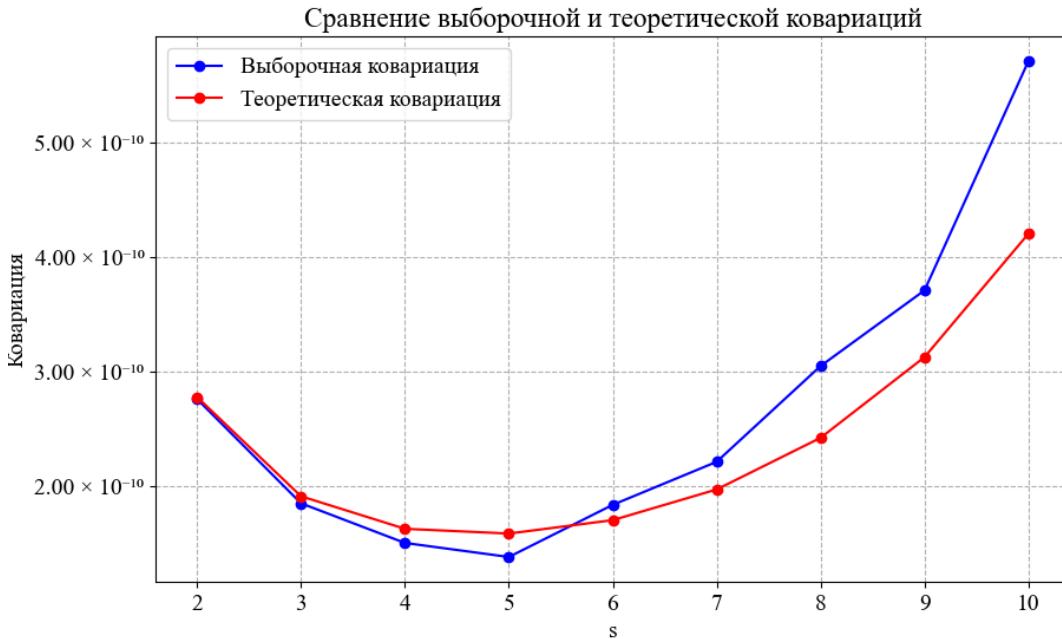


Рис. 2. Сравнение выборочной и теоретической ковариации оценок энтропии Шеннона, вычисленных по s - и $(s + 1)$ -граммам, в зависимости от s

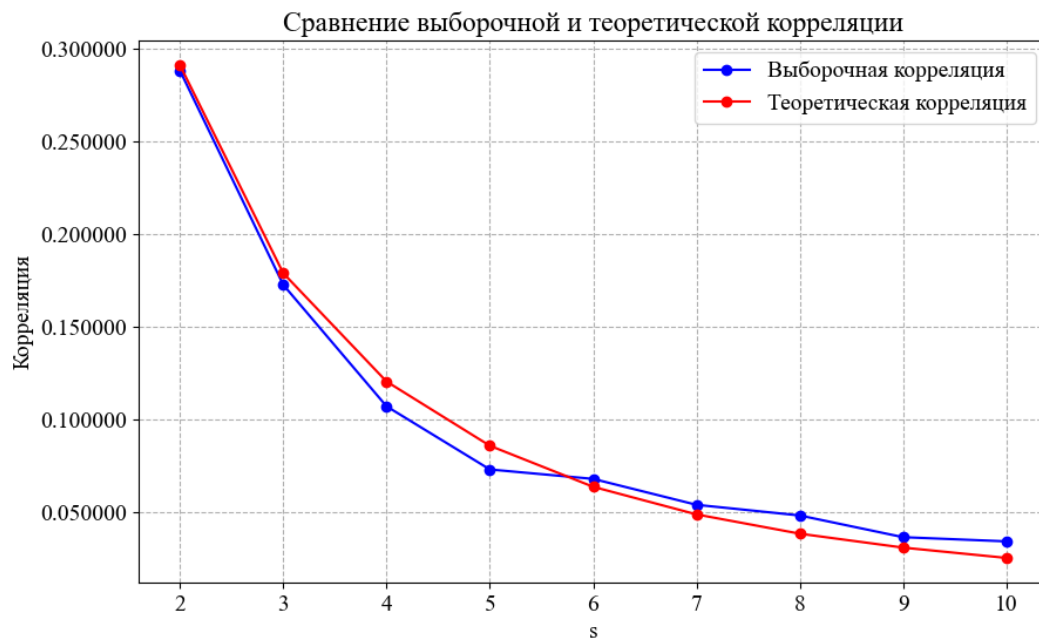


Рис. 3. Сравнение выборочного и теоретического коэффициента корреляции оценок энтропии Шеннона, вычисленных по s - и $(s + 1)$ -граммам, в зависимости от s

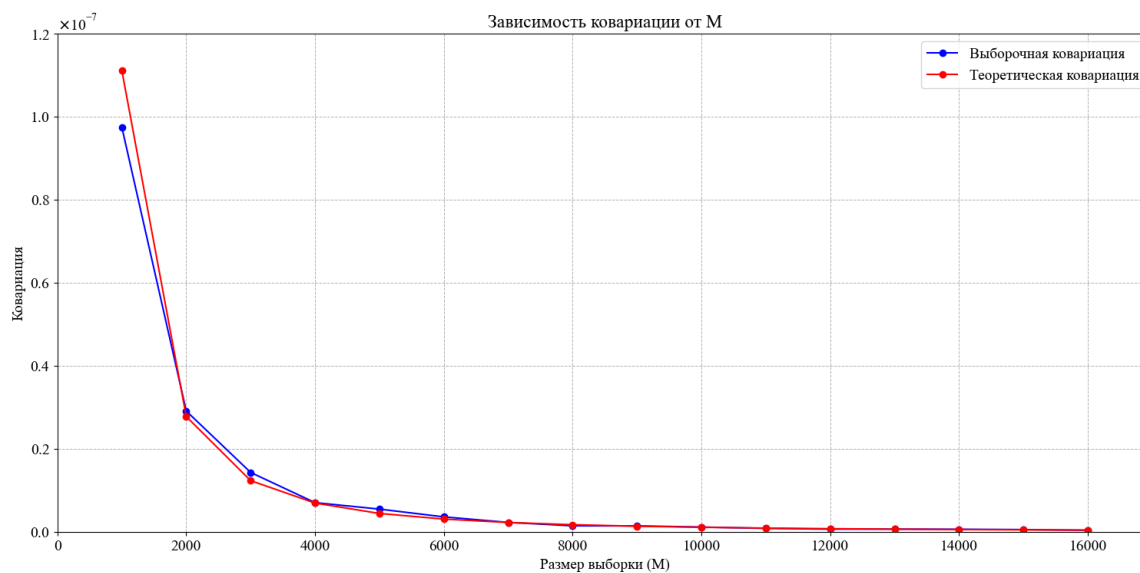


Рис. 4. Сравнение выборочной и теоретической ковариации оценок энтропии Шеннона, вычисленных по s - и $(s + 1)$ -граммам, в зависимости от M

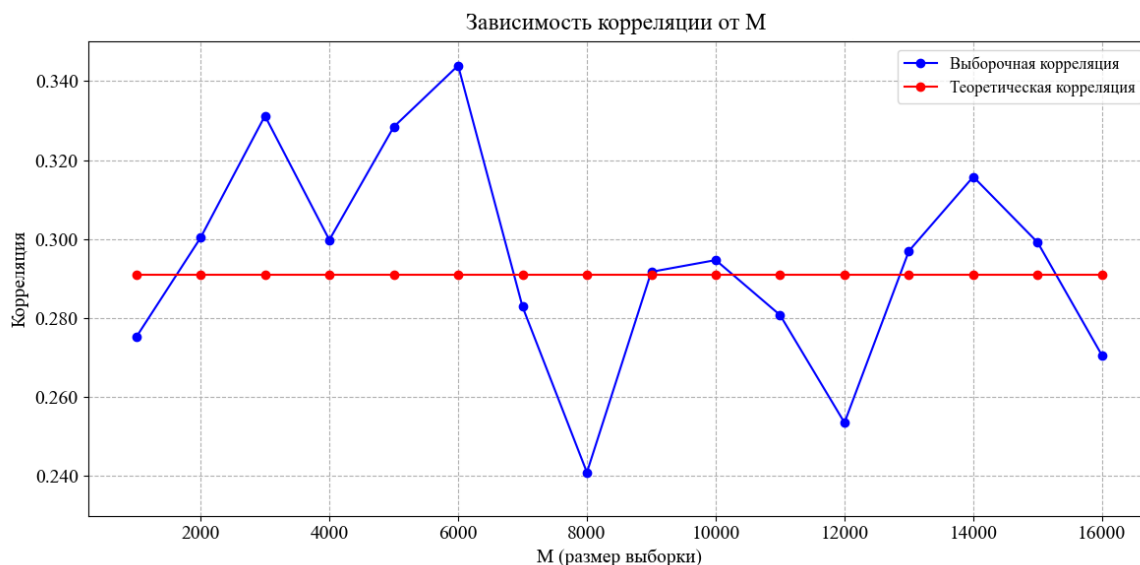


Рис. 5. Сравнение выборочного и теоретического коэффициента корреляции оценок энтропии Шеннона, вычисленных по s - и $(s + 1)$ -граммам, в зависимости от M

В заключение данного раздела для сравнения приведем результаты оценки надежности рассматриваемого генератора с использованием минимальной энтропии. Для проведения эксперимента оценивания минимальной энтропии формировались 5 бинарных последовательностей псевдослучайных чисел размером 1 МБ каждая. В настоящей статье для оценивания минимальной энтропии используется 5 наиболее эффективных с точки зрения быстродействия тестов: частотный, коллизий, марковской зависимости, сжатия и прогнозирования задержки [18]. Тесты минимальной энтропии, как правило, применяются в случае неизвестной природы случайной последовательности. Во время эксперимента также измерялось время выполнения каждого теста. Результаты эксперимента для оценок значений статистик минимальной энтропии приведены в табл. 1, а результаты быстродействия алгоритмов приведены в табл. 2. Как видно из табл. 1, для псевдослучайных данных все минимальные значения энтропии обеспечивает тест коллизий.

Таблица 1

Значения статистик тестов минимальной энтропии

Название теста \ номер последовательности	1	2	3	4	5
Частотный тест	0,998599	0,998110	0,998015	0,998566	0,998172
Тест коллизий	0,946371	0,985282	0,946583	0,943320	0,943574
Тест марковской зависимости	0,999098	0,998683	0,997779	0,998512	0,998815
Тест сжатия	0,976419	0,985980	0,957144	0,959197	0,969160
Тест прогнозирования задержки	0,998948	0,987015	0,999333	0,998552	0,998343

4. Заключение

В статье рассмотрены подходы к оценке надежности криптографических генераторов на основе энтропийных характеристик и искусственных нейронных сетей. Установлены статистические свойства оценок энтропии двоичной последовательности, вычисленных по фрагментам длины s и $s + 1$, в случае справедливости нулевой гипотезы о том, что наблюдаемая последовательность является РРСП. Найдены асимптотические выражения для ковариации и коэффициента корреляции оценок энтропии, показано, что с ростом s коэффициент корреляции стремится к 0, что свидетельствует о слабой зависимости между оценками энтропии. Проведены две серии компьютерных экспериментов, подтверждающие полученные результаты, приведено сравнение с оценками минимальной энтропии.

Библиографические ссылки

1. Зубков А. М., Серов А. А. Проверка пакета статистических критериев NIST на специальных псевдослучайных последовательностях // Математические вопросы криптографии. 2019. Т. 10, вып. 2. С. 89–96.
2. Rivest R. Cryptography and machine learning // Advances in Cryptology. ASIACRYPT 91. 1991. P. 427–439.
3. Millichap C., Yau Y. An artificial neural network approach to finding the key length of the Vigenere cipher. URL: <https://arxiv.org/abs/2312.09956> (date of access: 01.09.2025).
4. Greydanus S. Learning the Enigma with Recurrent Neural Networks. URL: <https://arxiv.org/abs/1708.07576> (date of access: 01.09.2025).
5. Lerman L., Bontempi G., Markowitch O. Power analysis attack: an approach based on machine learning // International Journal of Applied Cryptography. 2014. Vol. 3. P. 97–115.
6. Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in IoT / M. Lopez-Martin [et al.] // Sensors. 2017. Vol. 17. Article no. 1967.
7. Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source mismatch / L. Pibre [et al.] // Electronic Imaging. 2016. Vol. 8. P. 1–11.
8. Jarusek R., Volna E., Kotyrba M. Neural network approach to image steganography techniques // Advances in Intelligent Systems and Computing. 2015. P. 317–327.
9. Ansaroudi Z., Zaccagnino R., D'Arco P. On Pseudorandomness and Deep Learning: A Case Study // Applied Sciences. 2023. Vol. 13. Article no. 3372.
10. Boanca S. Exploring patterns and assessing the security of pseudorandom number generators with machine learning // Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART 2024). 2024. Vol. 3. P. 186–193.
11. Machine learning cryptanalysis of a quantum random number generator / N. Truong [et al.] // IEEE Transactions on Information Forensics and Security. 2019. Vol. 14, no. 2. P. 403–414.
12. Desai V., Deshmukh V., Rao D. Pseudo Random Number Generator Using Elman Neural Network // 2011 IEEE Recent Advances in Intelligent Computational Systems. 2011. P. 251–254.

13. *Long H.* Stream Cipher Method Based on Neural Network // Proceedings of the 2012 National Conference on Information Technology and Computer Science, CITCS. 2012. P. 414–417.
14. *Lian S.* A block cipher based on chaotic neural networks // Neurocomputing. 2009. Vol. 72. P. 1296–1301.
15. *Lian S., Sun J., Wang Z.* One-way Hash Function Based on Neural Network. URL: <https://arxiv.org/abs/0707.4032> (date of access: 01.09.2025).
16. On statistical testing of random and pseudorandom sequences based on entropy functionals / U. Yu. Palukha [et al.] // Computer Data Analysis and Modeling: Stochastics and Data Science : Proc. of the XIII Intern. Conf., Minsk, Sept. 6–10, 2022 / Belarusian State University ; eds.: Yu. Kharin [et al.]. Minsk : BSU, 2022. P. 148–162. URL: <https://elib.bsu.by/handle/123456789/291848> (date of access: 01.09.2025).
17. Информационные технологии и безопасность. Криптографические алгоритмы генерации псевдослучайных чисел. СТБ 34.101.47–2017. Минск : Госстандарт, 2017. III.
18. Recommendation for the Entropy Sources Used for Random Bit Generation / M. S. Turan [et al.]. NIST. 2018. Special Publication 800-90B. DOI: [10.6028/NIST.SP.800-90B](https://doi.org/10.6028/NIST.SP.800-90B).

АНАЛИЗ АНОМАЛЬНОГО ТРАФИКА В ЛОКАЛЬНОЙ СЕТИ ПРИ ПОМОЩИ ПРОТОКОЛА SNMP

С. Г. Саевич, И. О. Петроченко, А. П. Мателенок

*Полоцкий государственный университет имени Евфросинии Полоцкой,
Новополоцк, Беларусь, a.matelenok@psu.by*

Рассмотрены фундаментальные вопросы диагностики и защиты сетевых ресурсов от аномальных воздействий, приводящих к полной или частичной потере работоспособности, уничтожению, искажению, утечке информации, или несанкционированному доступу к сетевым ресурсам. Показано, как SNMP позволяет обнаруживать аномальную сетевую активность и предоставляет данные, на основе которых можно принять меры. SNMP сам по себе «не борется», но является ключевым инструментом для обнаружения и сбора информации, необходимой для борьбы.

Ключевые слова: аномальный трафик; протокол SNMP; информационные риски; анализ защищенности.

ANALYSIS OF ANOMALOUS TRAFFIC IN A LOCAL NETWORK USING SNMP PROTOCOL

S. G. Saevich, I. O. Petrochenko, A. P. Matelenok

*Polotsk State University named after Euphrosyne of Polotsk,
Novopolotsk, Belarus, a.matelenok@psu.by*

The fundamental issues of diagnostics and protection of network resources from abnormal impacts leading to full or partial loss of functionality, destruction, distortion, information leakage, or unauthorized access to network resources are considered. It is shown how SNMP allows detecting abnormal network activity and provides data on the basis of which measures can be taken. SNMP itself “does not fight”, but is a key tool for detection and collection of information necessary for the fight.

Keywords: abnormal traffic; SNMP protocol; information risks; security analysis.

1. Введение

Конгломерат различных компьютерных сетей, систем и их объединения посредством Интернета составляют информационную структура современной компании. Сложность логической и физической организации локальной сети приводит к объективным трудностям при решении вопросов об ее управлении и защиты [1]. Соответственно центральный вопрос –

оперативное обнаружение ее состояния, приводящего к потере полной или частичной ее работоспособности, уничтожению, искажению или утечки информации в результате несанкционированной деятельности хакеров, вирусов и т.д. Раннее обнаружение таких состояний позволит своевременно устранить их причину, а также предотвратить возможные нежелательные последствия. Для диагностики таких состояний применяется значительный спектр специализированных систем [2]: межсетевые экраны, антивирусы, система обнаружения атак (Intrusion Detection System) система контроля целостности, криптографические средства защиты. Однако, указанные системы обнаружения имеют ряд недостатков [3]:

- периодическое применение при наличии признаков вмешательства, так как постоянное их использование приводит к избыточному потреблению ресурсов системы;
- методы анализа специализированных систем направлены на обнаружение известных типов воздействия, но за частую не в состоянии обнаружить их модификации.

Таким образом, в настоящее время является актуальным вопрос поиска более эффективных методов выявления недопустимых аномалий в работе сети, являющихся следствием технических сбоев или несанкционированного воздействия. Основным требованием к этим методам является возможность обнаружения произвольных типов аномалий, в том числе новых, а также воздействий, распределенных во времени. Одним из таких методов, на наш взгляд, является выявление аномальной активности с помощью протокола SNMP.

Целью работы является показать, как SNMP позволяет обнаруживать аномальную сетевую активность и предоставляет данные, на основе которых можно принять меры. SNMP сам по себе «не борется», но является ключевым инструментом для обнаружения и сбора информации, необходимой для борьбы.

2. Проведение атаки и демонстрация

В рамках данного исследования была проведена демонстрация возможностей SNMP для обнаружения аномальной сетевой активности. Была смоделирована типичная ЛВС, включающая сетевые устройства, поддерживающие SNMP, и систему мониторинга, агрегирующую данные, полученные по этому протоколу (рис. 1).

Основные выполненные настройки.

Конфигурация сервера: на сервере под управлением ОС Kali Linux (IP-адрес 192.168.6.164) произведена настройка сетевого интерфейса e0, обеспечивающего его подключение к центральному коммутатору (SW) и

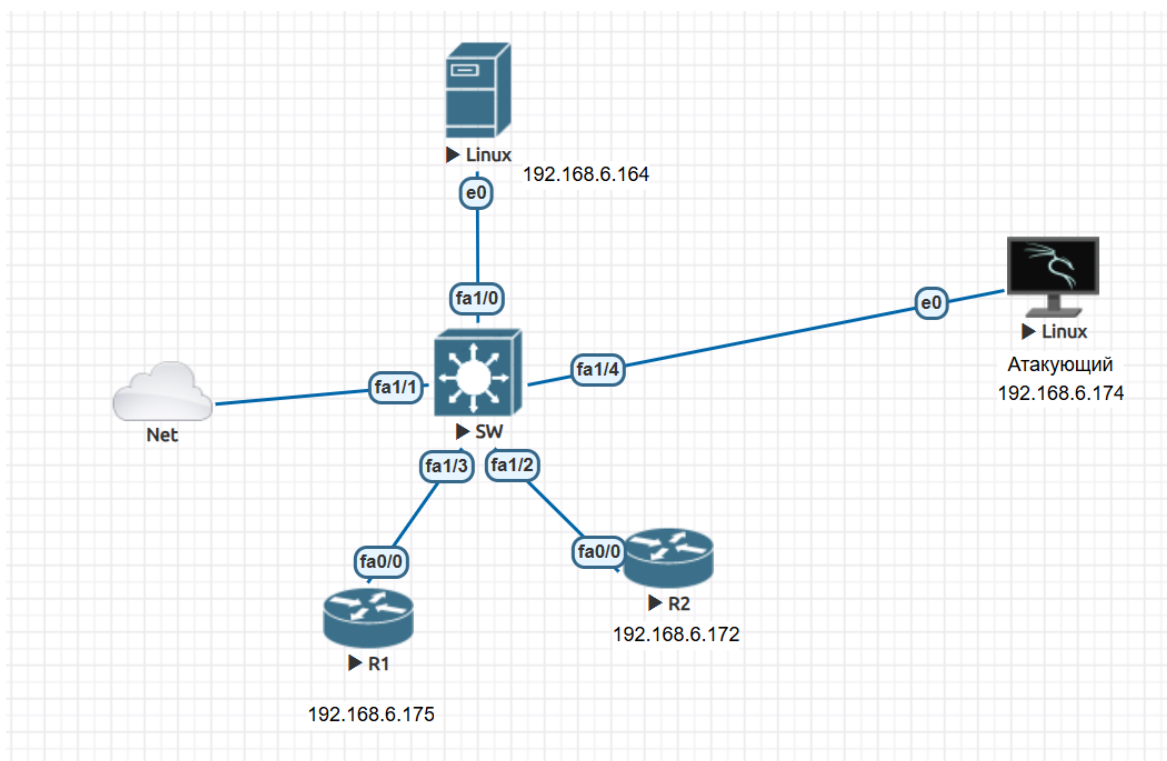


Рис. 1. Локальная сеть

доступность в локальной сети. Ключевым элементом конфигурации сервера стало развертывание системы сетевого мониторинга Observium. Данная система предназначена для автоматического обнаружения сетевых устройств, сбора данных об их состоянии и производительности, а также визуализации полученной информации непосредственно с данного сервера.

Конфигурация атакующего компьютера: компьютер под управлением ОС Kali Linux, обозначенный как «Атакующий» (IP-адрес 192.168.6.174). Для него выполнена базовая настройка сетевого интерфейса e0 для обеспечения взаимодействия в рамках локальной сети. Данный узел используется для осуществления атаки на локальную сеть путём UDP Flood.

Настройка протокола SNMP на маршрутизаторах: на маршрутизаторах R1 (IP-адрес 192.168.6.175) и R2 (IP-адрес 192.168.6.172) была выполнена базовая конфигурация протокола SNMP. Активация SNMP-агентов на данных устройствах позволяет системе Observium осуществлять опрос маршрутизаторов для сбора ключевых метрик, таких как загрузка интерфейсов, состояние устройства, таблицы маршрутизации и другие параметры, необходимые для анализа работы сети.

Таким образом, созданная конфигурация представляет собой функциональный сегмент сети, где ключевые устройства (сервер и

маршрутизаторы) подготовлены для внешнего мониторинга с использованием стандартных протоколов и специализированного программного обеспечения.

Покажем, как Observium фиксирует аномальный рост трафика на интерфейсе одного из роутеров. Будем атаковать R2 (IP: 192.168.6.172) на его интерфейс fa0/0 (рис. 2–3).

На представленных скриншотах из системы Observium для маршрутизатора R2 (с IP-адресом 192.168.6.172) видно, что система мониторинга успешно подключилась к устройству и начала процесс сбора данных.

Текущие графики трафика показывают небольшую сетевую активность на интерфейсе FastEthernet0/0. Эта активность соответствует фоновому обмену данными (например, служебный трафик протоколов маршрутизации, ARP-запросы). Долгосрочные графики пока не информативны, так как сбор данных только начался.

Реализация атаки UDP Flood с компьютера атакующего (IP-адрес 192.168.6.174) на интерфейс FastEthernet0/0 роутера R2 представлена на рис. 4.

С помощью команды `sudo hping3 --flood --rand-source -p 80 --udp 192.168.6.172` была инициирована сетевая атака. Программа `hping3` используется для отправки специально сформированных сетевых пакетов, а ключ `sudo` предоставляет ей необходимые права администратора. Опция `--flood` указывает программе отправлять пакеты максимально быстро, не дожидаясь ответов. Ключ `--rand-source` заставляет программу

«мирное время»

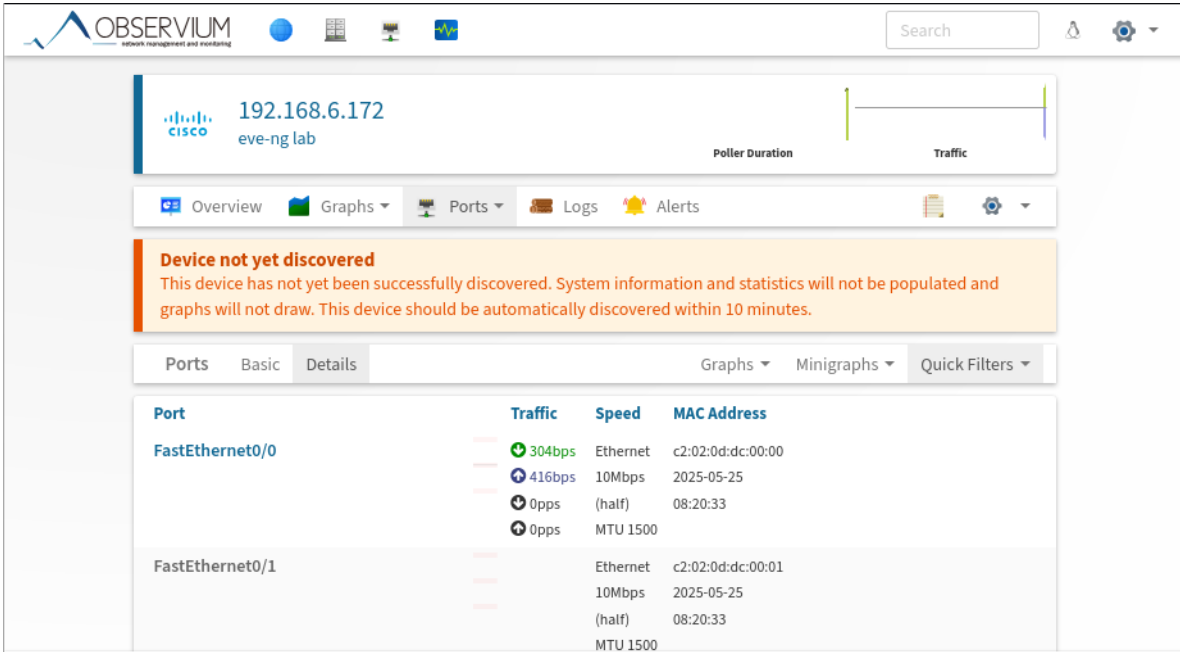


Рис. 2. Общий вид устройства (R2)

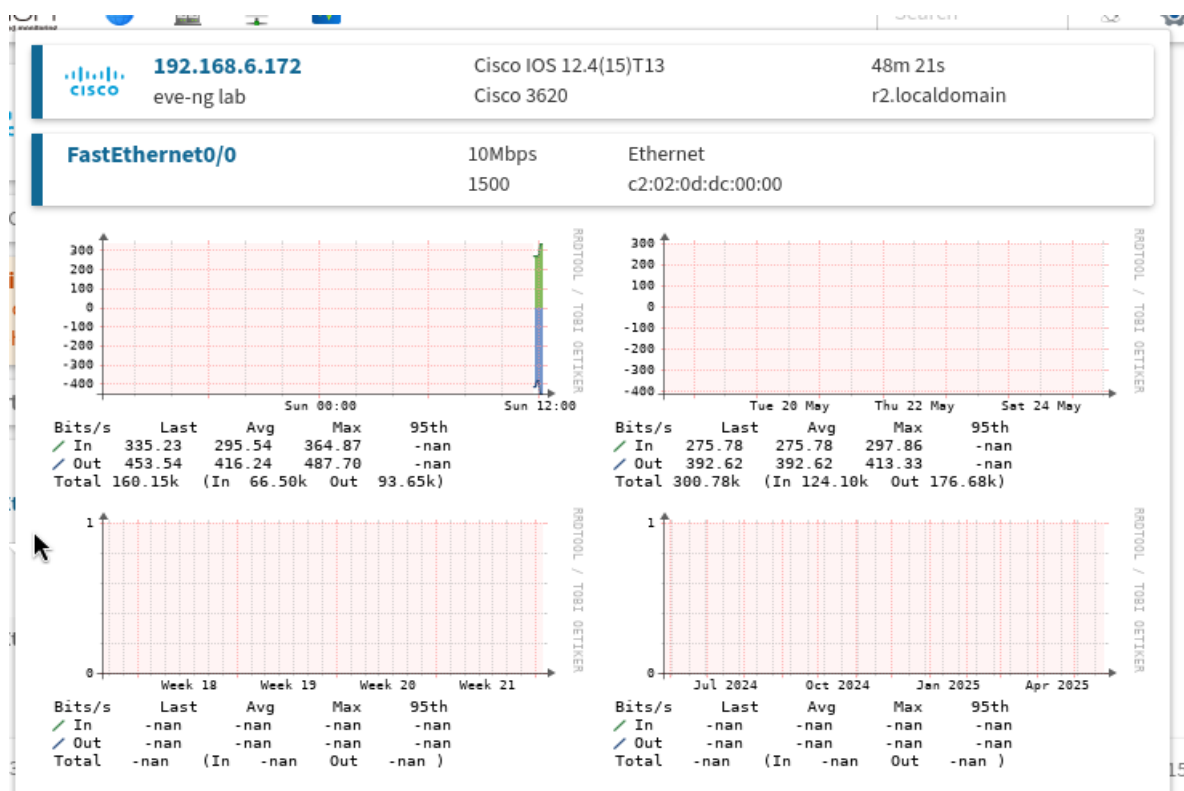


Рис. 3. Детальные графики для порта FastEthernet0/0

```
(kali@kali)-[~]
$ sudo hping3 --flood --rand-source -p 80 --udp 192.168.6.172
HPING 192.168.6.172 (eth0 192.168.6.172): udp mode set, 28 headers + 0 data bytes
hping in flood mode, no replies will be shown
^C
— 192.168.6.172 hping statistic —
24443437 packets transmitted, 0 packets received, 100% packet loss
round-trip min/avg/max = 0.0/0.0/0.0 ms
```

Рис. 4. Окно терминала на атакующем компьютере

использовать случайные (поддельные) IP-адреса отправителя для каждого пакета, что затрудняет идентификацию источника атаки. Параметр `-p 80` задает в качестве целевого порта на атакуемом устройстве порт 80, который обычно используется для веб-трафика (HTTP), а `--udp` определяет, что будут отправляться пакеты протокола UDP. Адрес 192.168.6.172 является IP-адресом цели атаки, в данном случае это маршрутизатор R2.

Таким образом, была проведена атака типа UDP Flood. Суть этой атаки заключается в отправке огромного количества UDP-пакетов на целевое устройство. Поскольку UDP – это протокол без установления соединения, он не требует подтверждения доставки, что позволяет атакующему генерировать очень большой объем трафика с минимальными затратами собственных ресурсов.

Цель атаки – исчерпать пропускную способность сетевого канала к атакуемому устройству или перегрузить его ресурсы (процессор, память) обработкой входящего потока пакетов, что может привести к отказу в обслуживании.

Из вывода команды видно, что было передано более 24 миллионов пакетов, при этом ни одного ответа получено не было, что характерно для данного типа атаки, особенно при использовании опции.

После начала атаки мы сразу замечаем изменения выводимых данных Observium (рис. 5).

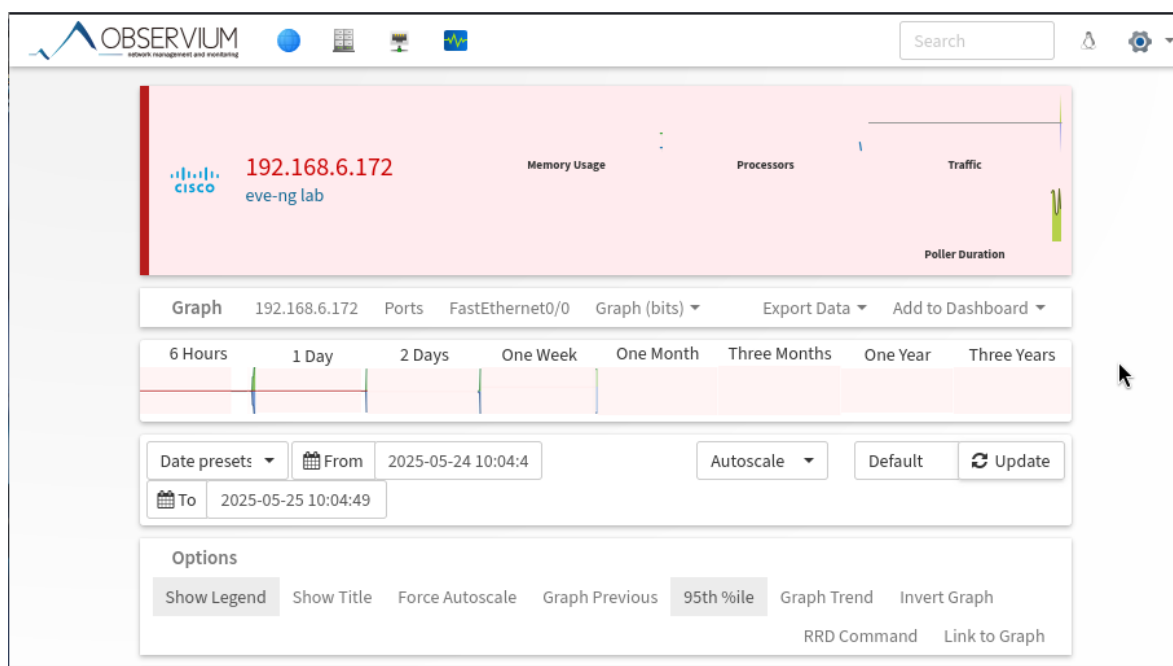


Рис. 5. Общий вид графика трафика и элементы управления

Верхняя часть: мы видим информационную панель для устройства 192.168.6.172. Справа есть мини-графики: Memory Usage (Использование памяти), Processors (Загрузка процессоров) и Traffic (Трафик). На мини-графике Traffic отчетливо виден огромный вертикальный всплеск, указывающий на резкое и значительное увеличение сетевого трафика. Это прямое следствие UDP Flood атаки, когда атакующий компьютер завалил маршрутизатор пакетами.

Главное здесь – подтверждение системой мониторинга факта аномально высокого трафика на атакованное устройство (рис. 6).

Это увеличенный график трафика для интерфейса FastEthernet0/0 роутера R2 за определенный период.

Ключевой момент: Правее отметки “Sun 12:00” (Воскресенье, 12:00) виден резкий, почти вертикальный скачок трафика (зеленая и синяя

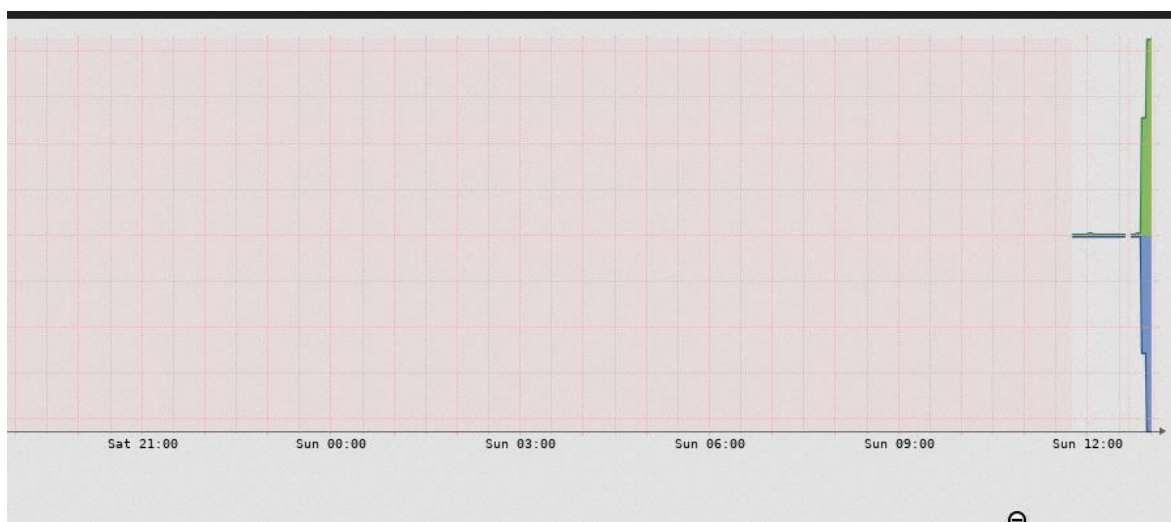


Рис. 6. Детальный график трафика

области). Это и есть визуализация UDP Flood атаки. Зеленая область, входящий трафик, синяя – исходящий. Высота этого пика показывает интенсивность потока данных во время атаки. До и после этого пика трафик был минимальным.

График на рис. 7 наглядно демонстрирует момент начала и окончания атаки, а также ее интенсивность.

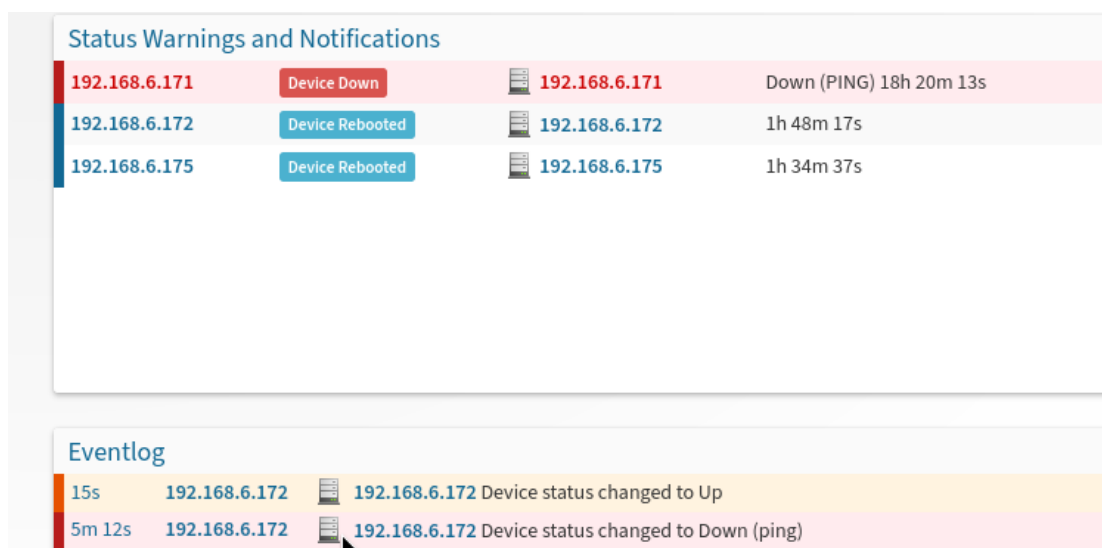


Рис. 7. Status Warnings and Notifications + Eventlog (журнал событий)

Верхний блок “Status Warnings and Notifications” (Предупреждения и Уведомления о состоянии).

192.168.6.172 Device Rebooted 1h 48m 17s: очень важное сообщение! Атакованный маршрутизатор 192.168.6.172 был перезагружен 1 час 48

минут назад. Это следствие успешной атаки. Устройство не справилось с нагрузкой и «пало».

Нижний блок “Eventlog” (продолжение Журнала событий).

15s 192.168.6.172 Device status changed to Up: 15 секунд назад атакованный R2 снова стал доступен (вернулся в состояние “Up”) после перезагрузки.

5m 12s 192.168.6.172 Device status changed to Down (ping): это то же событие, что мы видели на предыдущем скриншоте лога – когда R2 «упал» из-за атаки.

Эти логи показывают полный цикл: атака -> устройство недоступно -> устройство перезагружается -> устройство снова доступно. Это классический сценарий последствий атаки UDP Flood.

Ошибки, возникшие после атаки и перезагрузки устройства R2, представлены на рис. 8.

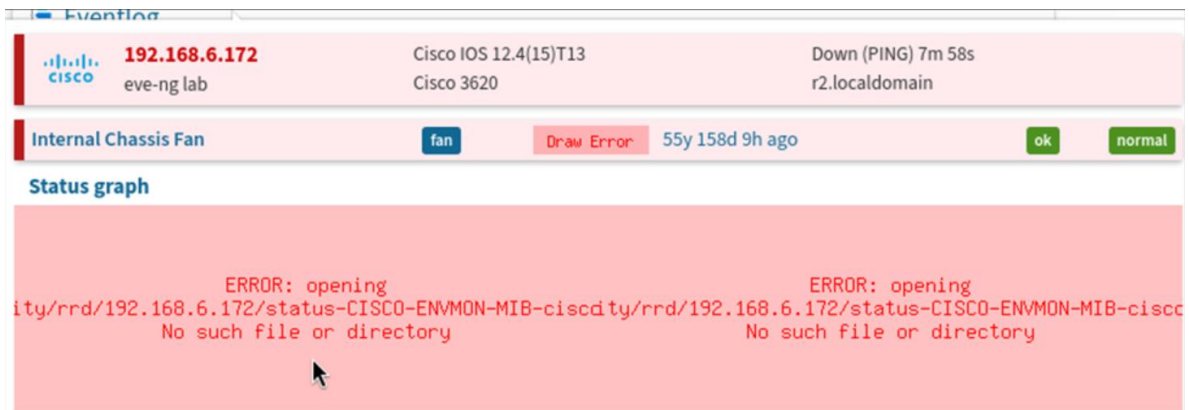


Рис. 8. Страница устройства R2 с ошибками

ERROR: opening /var/lib/observium/rrd/192.168.6.172/status-CISCO-ENVMON-MIB-ciscosity.rrd: No such file or directory. Эти ошибки говорят о том, что Observium не может найти или открыть файлы RRD (Round Robin Database), в которых хранятся исторические данные для некоторых сенсоров (в данном случае, относящихся к CISCO-ENVMON-MIB – MIB для мониторинга окружающей среды Cisco, например, температуры, состояния вентиляторов).

Когда устройство аварийно перезагружается или SNMP-сервис на нем перестает отвечать корректно из-за высокой нагрузки, сбор некоторых данных может прерваться. В результате файлы RRD для этих метрик могут не создаваться, повредиться или Observium потеряет к ним доступ на время.

Скриншот на рис. 8 показывает остаточные проблемы после атаки и перезагрузки: Observium не может получить или отобразить некоторые специфические данные мониторинга для R2, вероятно, из-за сбоя в их сборе во время или сразу после инцидента.

3. Общие выводы

Проведенная UDP Flood атака была достаточно мощной, чтобы вызвать значительный всплеск трафика на маршрутизаторе R2 (192.168.6.172). Это привело к тому, что маршрутизатор перестал отвечать (стал “Down”), а затем аварийно перезагрузился. После перезагрузки он снова стал доступен. Система Observium успешно зафиксировала как атаку (всплеск трафика).

Протокол SNMP сыграл ключевую роль в том, что система мониторинга Observium смогла обнаружить, отобразить и зарегистрировать как саму UDP Flood атаку, так и ее последствия для маршрутизатора R2 (192.168.6.172). Вот как это проявилось на каждом этапе.

1. Обнаружение всплеска трафика (см. рис. 5–6).

Именно благодаря SNMP-запросам к маршрутизатору R2 Observium смог получить данные со счетчиков трафика на его интерфейсе FastEthernet0/0. SNMP позволяет системе мониторинга регулярно опрашивать устройство и узнавать, сколько байт/пакетов прошло через каждый порт. Без SNMP Observium не смог бы построить график, наглядно демонстрирующий аномальный скачок трафика во время атаки. Он бы просто не знал, что происходит на сетевом уровне устройства.

2. Определение состояния устройства (см. рис. 7 – “Down”, “Up”, “Rebooted”).

Хотя сообщение “Down (ping)” указывает на отказ ICMP (ping) запросов, SNMP предоставляет гораздо больше информации о состоянии устройства.

До атаки: SNMP позволял Observium регулярно проверять работоспособность SNMP-агента на R2, собирать данные о времени его непрерывной работы (uptime), загрузке процессора, использовании памяти и т.д.

Во время атаки: когда R2 был перегружен, его SNMP-агент, скорее всего, также перестал отвечать или отвечал с огромными задержками. Это является дополнительным сильным индикатором серьезных проблем для Observium.

Обнаружение перезагрузки: изменение значения uptime устройства (которое Observium получает по SNMP) является прямым указанием на то, что устройство было перезагружено. Сообщение “Device Rebooted” генерируется именно на основе этой информации.

Восстановление: когда R2 снова стал доступен (“Device status changed to Up”), Observium смог возобновить успешные SNMP-запросы, подтвердив восстановление работоспособности и начав заново собирать метрики.

3. Сбор детальной информации и ошибки (см. рис. 8 – ошибки ENVMON-MIB).

Сообщения об ошибках ERROR: opening ... status-CISCO-ENVMON-MIB-ciscosity.rrd напрямую связаны с SNMP. CISCO-ENVMON-MIB – это специальный набор объектов (MIB – Management Information Base), который устройства Cisco предоставляют по SNMP для мониторинга параметров окружающей среды (температура, состояние вентиляторов, напряжение и т.д.).

Observium пытался по SNMP запросить эти данные у R2.

Сразу после атаки/перезагрузки SNMP-агент на R2 мог быть недоступен, или конкретно эти данные были недоступны. В результате Observium не смог получить значения, и файлы RRD (где хранятся исторические данные) для этих метрик не были обновлены или созданы корректно, что и вызвало ошибку при попытке их отобразить. Это также свидетельствует о сбое в работе устройства, зафиксированном благодаря попыткам SNMP-опроса.

Таким образом, достоинства SNMP заключается в следующем:

- **предоставление данных для визуализации:** SNMP позволил «увидеть» атаку в виде всплеска трафика на графиках;
- **Диагностика состояния:** SNMP дал Observium возможность отслеживать работоспособность R2, фиксировать его недоступность, перезагрузку и последующее восстановление;
- **детализированный мониторинг:** даже ошибки при сборе специфических данных (как с CISCO-ENVMON-MIB) через SNMP являются важной диагностической информацией, указывающей на проблемы с определенными подсистемами или службами на устройстве.

Без SNMP система Observium была бы значительно «слепее» и не смогла бы предоставить такой подробной картины произошедшего инцидента и его влияния на сетевое оборудование. SNMP – это «глаза и уши» системы мониторинга в сети.

Наше исследование доказывает ценность SNMP как инструмента для постфактум анализа и обнаружения аномалий. Однако истинный потенциал протокола раскрывается, когда он становится частью проактивной системы реагирования. Вместо того чтобы маршрутизатор пассивно «принимал удар» до полного отказа, можно разработать механизмы, где данные SNMP служат триггером для автоматических контрмер. Например, можно было бы сконфигурировать маршрутизатор таким образом, чтобы при достижении критических порогов загрузки интерфейса (значения которых Observium получает по SNMP) не просто регистрировалось событие, а инициировались защитные действия. Это могла бы быть временная блокировка порта, с которого идет аномальный трафик, или применение заранее подготовленных списков контроля доступа (ACL), отсекающих подозрительные пакеты. SNMP в этом сценарии не только сигнализирует

администратору через Observium (например, с помощью SNMP-трапов, отправляемых маршрутизатором немедленно при событии, а не ожидая опроса), но и может служить каналом для получения команды на изменение конфигурации от управляющей системы.

4. Заключение

Таким образом, в данном исследовании была продемонстрирована базовая конфигурация SNMP и её применение для мониторинга самых очевидных метрик, таких как загрузка интерфейса и общее состояние устройства. Современные реализации SNMP (особенно SNMPv3 с его улучшенной безопасностью) и обширные MIB-базы (Management Information Bases), открывают доступ к огромному количеству детализированной информации. SNMP способен предоставлять данные о загрузке каждого процессорного ядра, количестве ошибок на интерфейсах, состоянии таблиц маршрутизации, очередях пакетов, параметрах QoS, специфических для вендора показателях производительности и даже данных о попытках несанкционированного доступа. Это позволяет строить гораздо более сложные и тонкие профили нормального поведения сети и с большей точностью и на более ранних стадиях выявлять аномалии.

Библиографические ссылки

1. *Рябенко Б. Я., Фионов А. Н.* Криптографические методы защиты информации: учеб. пособие. М. : Горячая линия – Телеком, 2013.
2. *Олифер В., Олифер Н.* Компьютерные сети. Принципы, технологии, протоколы: Юбилейное издание. Учебник для ВУЗов. 6-е изд. СПб. : Питер, 2020.
3. *Костенко Е. Ю., Дуйсенгалиев Р. Р., Барабанова Е. А.* Система мониторинга для контроля трафика технологических сетей передачи данных // Вестник АГТУ. Серия: Управление, вычислительная техника и информатика. 2015. № 4. С. 101–109. URL: <https://cyberleninka.ru/article/n/sistema-monitoringa-dlya-kontrolya-trafika-tehnologicheskikh-setey-peredachi-dannyh> (дата обращения: 02.10.2025).

НОРМАЛИЗАЦИЯ СОБЫТИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ОПЕРАЦИОННЫХ СИСТЕМ С ИСПОЛЬЗОВАНИЕМ MITRE ATT&CK

А. М. Соболев¹⁾, В. П. Кочин²⁾

¹⁾ Государственное предприятие «Центр информационных
ресурсов и коммуникаций»,

Минск, Беларусь, sobol@niks.by

²⁾ Белорусский государственный университет,

Минск, Беларусь, kochyn@bsu.by

В статье рассматривается нормализация событий информационной безопасности операционных систем с использованием фреймворка MITRE ATT&CK. Приведены примеры нормализованных событий для Windows и Linux, показано, как сопоставление с MITRE ATT&CK упрощает анализ угроз и построение цепочек атак. Результаты исследования показывают, что нормализация событий является критически важным этапом построения интеллектуальных систем мониторинга безопасности.

Ключевые слова: информационная безопасность; центр кибербезопасности; нормализация событий; техника MITRE ATT&CK; журнал событий.

NORMALIZATION OF INFORMATION SECURITY EVENTS IN OPERATING SYSTEMS USING MITRE ATT&CK

A. M. Sobol^{a)}, V. P. Kochyn^{b)}

^{a)} State Enterprise “Center for Information Resources and Communications”,

Minsk, Belarus, sobol@niks.by

^{b)} Belarusian State University,

Minsk, Belarus, kochyn@bsu.by

This article discusses the normalization of information security events in operating systems using the MITRE ATT&CK framework. Examples of normalized events for Windows and Linux are provided, and it is shown how mapping to MITRE ATT&CK simplifies threat analysis and attack chain construction. The results of the study show that event normalization is a critical step in building intelligent security monitoring systems.

Keywords: information security; cybersecurity center; event normalization; MITRE ATT&CK technique; event log.

1. Введение

Современные центры кибербезопасности (далее – SOC) ежедневно сталкиваются с обработкой огромных объёмов данных, формируемых из различных источников: журналы событий операционных систем семейства Windows и Linux, телеметрии EDR- и антивирусных решений, сетевых датчиков, а также облачных сервисов. По прогнозу аналитиков Gartner, к 2025 году количество событий безопасности, требующих анализа, увеличится более чем втрое по сравнению с 2020 годом [1].

Главная сложность в таких данных заключается не только в масштабе, но и в **разнородности** источников событий. Каждая операционная система или средство защиты информации использует собственный формат журналов событий и структуру полей. Для решения указанной проблемы применяется процесс **нормализации событий информационной безопасности**, заключающийся в приведении разных форматов событий информационной безопасности к унифицированной модели данных.

Для повышения ценности собранных событий в расследовании киберинцидентов необходимо обеспечить их привязку к **контексту угроз**. Наиболее распространённым инструментом для этого является база знаний **MITRE ATT&CK**, которая представляет собой таксономию тактик, техник и процедур (далее – TTPs), применяемых злоумышленниками [2]. Привязка нормализованных событий к техникам ATT&CK позволяет:

- отнести событие к определённому этапу атаки;
- выявить скрытые взаимосвязи между действиями в различных операционных системах;
- автоматизировать реагирование и построение сценариев противодействия атакам.

Таким образом, нормализация событий информационной безопасности операционных систем с использованием MITRE ATT&CK является не только техническим механизмом упрощения анализа, но и стратегическим инструментом, который повышает эффективность SOC, сокращает время расследования и позволяет выявить сложные многоэтапные атаки.

2. Метод нормализации событий операционных систем

Нормализация событий информационной безопасности – это процесс приведения разнородных логов к единой структуре с унифицированной терминологией и форматами данных [3]. Целью нормализации событий является:

- устранение разнородности логов разных ОС;
- создание единой схемы для аналитики и корреляции;

- сопоставление событий с TTPs по MITRE ATT&CK [2];
- повышение эффективности реагирования SOC за счёт сокращения ложных срабатываний.

Нормализация событий ИБ включает несколько ключевых этапов, обеспечивающих структурирование, унификацию и обогащение данных [4, 5].

Этап 1. Синтаксический анализ и структурирование на основе матрицы MITRE.

Синтаксический анализ позволяет извлекать ключевые элементы событий (например, Event ID, имя пользователя, PID процесса, имя файла и путь к нему) на основе существующих тактик MITRE и преобразовывать их в структурированные объекты (JSON, XML, CSV). Данный этап упрощает последующую обработку и корреляцию событий.

Этап 2. Унификация терминологии.

Унификация терминологии заключается в приведении разрозненных обозначений событий к единому стандарту. Например, `failedlogin` в Windows и `authenticationfailure` в Linux объединяются под общим термином «неудачная попытка входа».

Этап 3. Обогащение события.

Обогащение включает добавление дополнительной информации для повышения аналитической ценности события. Это могут быть идентификаторы техник MITRE ATT&CK, данные о хосте, IP-адреса, геолокация, параметры запуска процессов.

Нормализация журнала событий на основе MITRE обеспечивает единый аналитический формат, пригодный для различных источников, и имеет следующие преимущества:

- стандартизация анализа угроз: унифицированная система тактик (Tactics) и техник (Techniques) позволяет анализировать события независимо от ОС;
- семантическая привязка: каждое событие получает контекст атакующих действий, упрощая построение цепочек атак;
- совместимость с аналитическими системами: SIEM и DL-модели проще обрабатывают события, сопоставленные с TTPs;
- повышение точности корреляции: объединение событий по тактикам снижает количество ложных срабатываний и упрощает построение правил реагирования.

Для унификации событий важно понимать, какие тактики атак встречаются в журнале событий информационной безопасности операционных систем. Основные тактики, часто выявляемые в ОС Windows и Linux, представлены в табл. 1.

Таблица 1

**Тактики, выявляемые в журнале событий операционных систем
Windows и Linux**

Тактика	Краткое описание
Credential Access	Попытки получения учетных данных пользователей (логины, пароли, токены).
Privilege Escalation	Повышение привилегий пользователей или процессов.
Defense Evasion	Обход средств защиты: отключение антивируса, очистка журналов событий, обфускация команд.
Discovery	Сбор информации о системе и сети: просмотр процессов, сканирование сетевых ресурсов.
Execution	Выполнение команд, скриптов или бинарных файлов, включая удаленное выполнение.

В табл. 2 рассмотрены тактики, представленные в табл. 1, описание и необходимые поля для их выявления в операционных системах Windows и Linux.

Таблица 2

Ключевые признаки в зависимости от тактики и техники

Тактика / Техника	Описание	Примеры Windows	Примеры Linux	Ключевые признаки для нормализации
Первоначальный доступ / Существующие учетные записи (T1078)	Использование легитимных учетных записей для доступа к системе. Атакующий может применять украденные или заранее созданные учетные записи, обходя защиту.	RDP вход с существующей учетной записью	SSH вход с существующей учетной записью	User, EventID (логин), src_ip, протокол, Time
Выполнение / Интерпретаторы командной строки и сценариев (T1059)	Запуск команд или скриптов для выполнения произвольного кода и автоматизации задач.	PowerShell, cmd.exe, запуск скриптов	Bash, sh, Python, cron-скрипты	User, Name_File, Type_Oper (execute), аргументы командной строки, PID, Time

Тактика / Техника	Описание	Примеры Windows	Примеры Linux	Ключевые признаки для нормализации
Закрепление / Автозапуск при загрузке или входе (T1547.001)	Автоматический запуск процессов или скриптов при старте системы или входе пользователя для устойчивого присутствия.	Добавление ключей автозагрузки, изменение реестра	Добавление cron-задания, systemd unit, rc.local	User, Name_File, dir, Type_Oper (create/modify, Time, Content (скрипт))
Получение учетных данных / Дамп процесса lsass (T1003.001)	Кража учетных данных пользователей или токенов доступа через память процессов или системные файлы.	Дамп процесса LSASS	Чтение /etc/shadow, ptrace процессов	User, Name_File (shadow, passwd, процесс), Type_Oper (read), PID, Content, Time
Перемещение внутри периметра / Службы удаленного доступа / RDP (T1021.001)	Перемещение злоумышленника внутри сети для расширения контроля. Используются легитимные протоколы удаленного доступа.	RDP соединения, SMB доступ	SSH, SFTP, SCP, Telnet	User, src_ip, dst_ip, протокол, Type_Oper (remote_access), PID, Time
Деструктивное воздействие / Удаление данных (T1485)	Уничтожение файлов или логов для саботажа или сокрытия следов вторжения.	Удаление файлов, очистка логов	rm, shred, > /var/log/auth.log	User, Name_File, dir, Type_Oper (delete/truncat, PID, Content, Time)

Перечисленные события могут быть как нормальным поведением, так и потенциальной угрозой. На основании данных, представленных в табл. 2, можно сформулировать входной вектор для дальнейшего анализа:

$$x_i = \{EventID, ID\_proc, User, Type\_Oper, Name\_File, dir, Time, Content, Tech\}, \quad (1)$$

где *EventID* – код события Windows или Linux; *ID_proc* – уникальный идентификатор процесса; *User* – учётная запись, инициировавшая

событие; *Type_Oper* – запуск процесса, изменение файла, создание пользователя; *Name_File* – исполняемый файл или изменяемый ресурс; *dir* – директория, откуда запускается файл, *Time* – дата и время события; *Content* – родительский процесс, права, нагрузка CPU; *Tech* – техника согласно MITRE ATT&CK.

На основании данного вектора в табл. 3 представлены нормализуемые события, связанные с первоначальным доступом, для операционной системы Windows и Linux.

Таблица 3

Пример нормализации события

Параметры нормализованного вектора	Пример для ОС Windows	Пример для ОС Linux
EventID	4624	1001
ID_proc	1240	2233
User	alice	bob
Type_Oper	execute	execute
Name_File	RDP	/usr/bin/ssh
Dir	C:\Windows\System32	/home/bob
Src_IP	192.168.11.10	192.168.13.15
Dst_IP	192.168.1.20	192.168.1.13
Protocol	RDP	SSH
Time	2025-08-09 08:54:36	2025-09-01 05:03:39
Content	Успешныйвход	Успешныйвход
Tech	T1078	T1078

3. Заключение

Нормализация событий ИБ с использованием MITRE ATT&CK обеспечивает структурирование и унификацию данных из различных источников. Этот процесс позволяет создавать единый поток событий, готовый для анализа и корреляции в SIEM, аналитических моделях и моделях, основанных на глубоком обучении [6, 7].

Полученные нормализованные данные обладают необходимой структурой для подачи в DL-модели, что в будущем позволит:

- обнаруживать аномалии и неизвестные атаки, выявляя сложные паттерны поведения;
- прогнозировать потенциальные сценарии атак на основе исторических данных;
- автоматически классифицировать события по тактикам и техникам MITRE ATT&CK, ускоряя процесс анализа и снижая нагрузку на специалистов ИБ.

Таким образом, нормализация событий является ключевым этапом построения интеллектуальных систем мониторинга безопасности и открывает возможности для интеллектуального анализа угроз в масштабируемых инфраструктурах.

Библиографические ссылки

1. Gartner. Security Information and Event Management Market Forecast. 2021. URL: <https://www.gartner.com/en> (date of access: 29.08.2025).
2. MITRE. MITRE ATT&CK®: Knowledge Base of Adversary Tactics and Techniques. 2025. URL: <https://attack.mitre.org/> (date of access: 05.08.2025).
3. Scarfone K., Mell P. Guide to Computer Security Log Management. NIST. 2006. Special Publication 800-92.
4. Козлов Д. В. Методика нормализации событий, получаемых от различных источников и отображающих любую ситуацию в системе или сети автоматизированной системы управления // I-methods. 2022. №4. URL: <https://cyberleninka.ru/article/n/metodika-normalizatsii-sobytiy-poluchaemyh-ot-razlichnyh-istochnikov-i-otrazhayuschih-lyubuyu-situatsiyu-v-sisteme-ili-seti-avtomatizirovannoy-sistemy-upravleniya> (дата обращения: 22.08.2025).
5. Долгачев М. В., Москвичев А. Д. Нормализация журналов событий с использованием древа фиксированной глубины // Вопросы кибербезопасности. 2020. №2 (36). URL: <https://cyberleninka.ru/article/n/normalizatsiya-zhurnalov-sobytiy-s-ispolzovaniem-dereva-fiksirovannoy-glubiny> (дата обращения: 28.08.2025).
6. Sobol A., Kochyn V. Modeling the State of Information Security of a Smart Campus // Open Semantic Technologies for Intelligent Systems : Research Papers Collection / Belarusian State University of Informatics and Radioelectronics ; eds.: V. V. Golenkov [et al.]. Minsk : BSUIR, 2024. Issue 8. P. 353–358. URL: <https://libeldoc.bsuir.by/handle/123456789/55563> (date of access: 28.08.2025).
7. Sobol A., Kochyn V., Huk A. Information Security System Models in Smart Campuses // Proceedings, 2024 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), Sochi, Russia, May 20–24, 2024 / IEEE, 2024. P. 750–754. DOI: [10.1109/ICIEAM60818.2024.10553706](https://doi.org/10.1109/ICIEAM60818.2024.10553706).

ПОДАВЛЕНИЕ ПЭМИН В ВИДЕОИНТЕРФЕЙСАХ ПОСРЕДСТВОМ ВЫСОКОЧАСТОТНОЙ ШИМ ПОМЕХИ

И. А. Третьяков, Я. И. Рушечников, В. В. Данилов

*ФГБОУ ВО «Донецкий государственный университет»,
Донецк, ДНР, Российская Федерация, i.tretiakov@mail.ru*

В настоящей работе рассмотрена возможность применения широтно-импульсной модуляции для создания помех в требуемом диапазоне частот и подавления гармоник побочных электромагнитных излучений и наводок в видеоинтерфейсах офисной вычислительной техники.

Ключевые слова: ПЭМИН; видеоинтерфейс; широтно-импульсная модуляция; STM32; радиоподавление.

SUPPRESSION OF TEMPEST IN VIDEO INTERFACES BY MEANS OF HIGH-FREQUENCY PWM INTERFERENCE

I. A. Tretiakov, Ia. I. Rushechnikov, V. V. Danilov

*Donetsk State University,
Donetsk, Donetsk People's Republic, Russian Federation, i.tretiakov@mail.ru*

In this paper, the possibility of using pulse width modulation to create interference in the required frequency range and to add harmonics of electromagnetic side effects and interference in video interfaces of office computing equipment is considered.

Keywords: TEMPEST; video interface; pulse width modulation; STM32; radio suppression.

1. Введение

Побочные электромагнитные излучения и наводки (ПЭМИН) представляют собой набор спектральных компонентов, анализ которых позволяет оценить риск или возможность перехвата информации из цифровых систем обработки и передачи информации. Исследования этого явления в условиях постоянно усложняющейся электромагнитной обстановки становятся все более актуальными [1–3], а разработка эффективных способов подавления таких ПЭМИН являются актуальной научно-технической задачей в области информационной безопасности и защиты информации [4–7].

2. Влияние высокочастотного ШИМ-сигнала на ПЭМИН

Высокочастотный широтно-импульсно-модулированный (ШИМ)-сигнал может использоваться не только для управления различными инертными потребителями, но и для создания помех в заданном диапазоне частот, например, на тех частотах, где наблюдаются ПЭМИН сигналы. Гибкость в настройке частоты основной гармоник и подавлении четных и нечетных гармоник позволяет генерировать помехообразный сигнал, способный подавить слабые ПЭМИН сигналы. На рис. 1 представлена структурная схема комплекса для исследования средств подавления ПЭМИН.

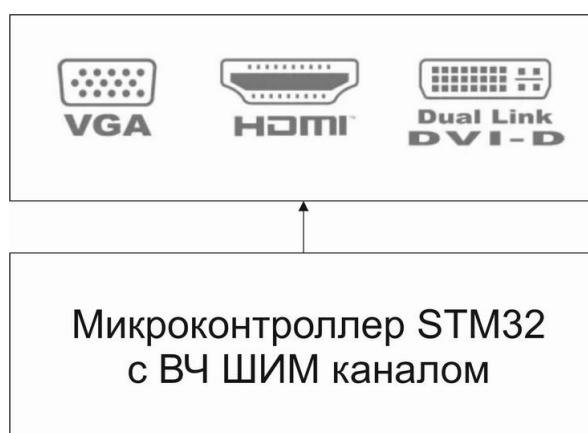


Рис. 1. Структурная схема подавления ПЭМИН сигнала средствами микроконтроллера с ШИМ

На данной схеме микроконтроллер, генерирующий ШИМ, подключается к фольгированному экранированию интерфейсов типа VGA, DVI, HDMI, что позволяет ему подавлять значительную часть ПЭМИН, представляющих собой потенциальный канал утечки информации. Такой комплекс может быть встроен как в системный блок, так и в монитор, адаптируясь к частотам ПЭМИН и подавляя их как основной гармоникой, так и кратными, что помогает зашумить возможные каналы утечки информации.

Оценка качества мероприятий по подавлению ПЭМИН проводилась по методике, описанной в [8]. Эта методика, используя разностный анализ экспериментальных данных, позволяет локализовать участки с ПЭМИН в частотной области для дальнейшего анализа. Спектрограммы (рис. 2–3) показывают, что частоты ПЭМИН остались прежними, но на них появилась мешающая составляющая, значительно снижающая вероятность декодирования ПЭМИН.

Метод измерений включает захват радиоэфира с многократным усреднением. Сравнивая спектрограммы измерений без источника помех (рис. 2) и с источником помех в виде высокочастотного ШИМ-сигнала (рис. 3),

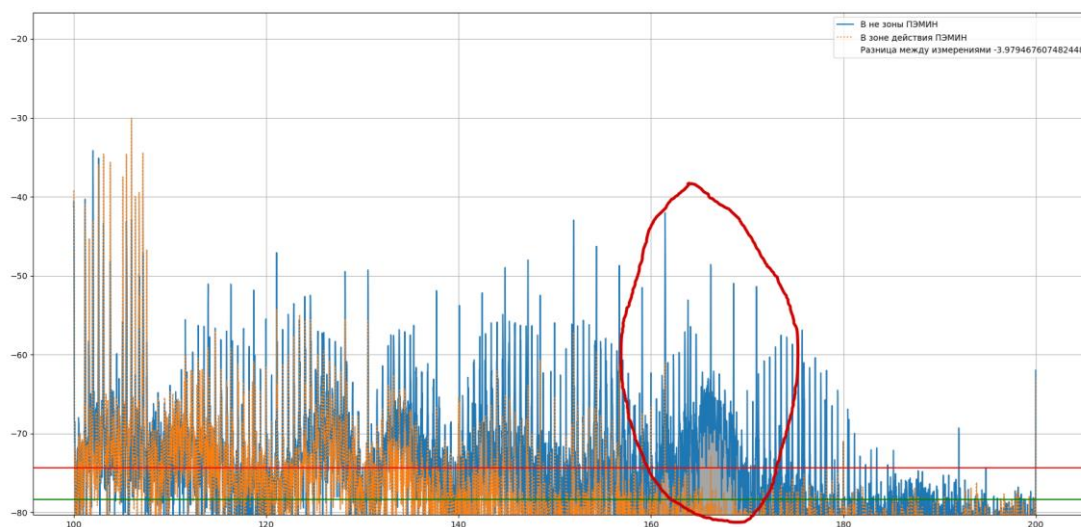


Рис. 2. Исходный сигнал ПЭМИН

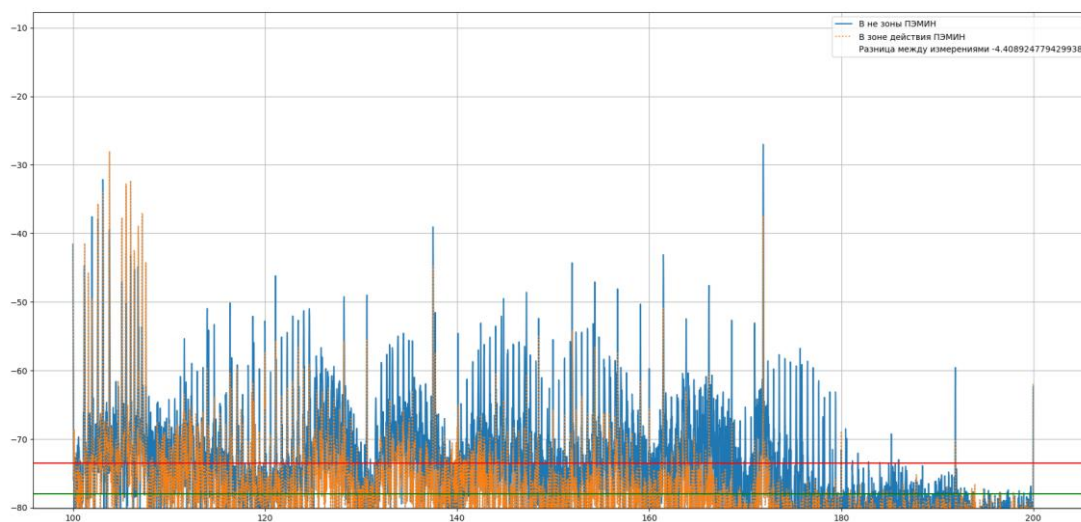


Рис. 3. Сигнал ПЭМИН, подавленный ВЧ ШИМ помехой

можно заметить следующую закономерность: в отсутствие мешающего сигнала на спектрограмме четко видны ПЭМИН на частотах 160–170 МГц (обведено красным на рис. 2), тогда как второе измерение в этом диапазоне не демонстрирует такой явной разницы. Отсутствие спектральной разницы между экспериментами с мешающим сигналом и без него значительно снижает различимость ПЭМИН на фоне окружающих помех, что, в свою очередь, уменьшает риск утечек информации через побочные каналы.

Попытки декодировать сигнал ПЭМИН и демодулировать его с помощью SDR-платформ [9] и перестраиваемых фильтров с автоматической регулировкой усиления показали, что сигнал сильно зашумлен, и его восстановление в реальном времени маловероятно.

3. Заключение

Экспериментальные результаты проведенного спектрального анализа свидетельствуют о возможности гибкого управления подавлением мало-мощного сигнала ПЭМИН, изменяя всего один параметр ШИМ-сигнала (коэффициент заполнения). Доступность и низкая стоимость этой технологии позволяют интегрировать подобный механизм в вычислительные системы без ущерба для их общей производительности и функциональности.

Исследование проводилось в рамках НИР «Исследование природы каналов побочных электромагнитных излучений и наводок элементов и устройств офисной вычислительной техники» (регистрационный номер 124012400347-2).

Библиографические ссылки

1. Антипов Д. А. Анализ утечек информации на основе побочных электромагнитных излучений // Доклады Томского государственного университета систем управления и радиоэлектроники. 2018. Т. 21, № 2. С. 27–32. DOI: [10.21293/1818-0442-2018-21-2-27-32](https://doi.org/10.21293/1818-0442-2018-21-2-27-32).
2. Шилевой А. А., Персичкин А. А. О возможности аналитического обнаружения сигнала ПЭМИН в видеоинтерфейсах стандарта HDMI // Вестник Балтийского федерального университета им. И. Канта. Серия: Физ.-мат. и техн. науки. 2022. № 1. С. 34–39.
3. Методика оценки побочных электромагнитных излучений посредством алгоритмической обработки массива экспериментальных спектрограмм / Я. И. Рушечников [и др.] // Вестник Донецкого национального университета. Серия Г: Технические науки. 2024. № 3. С. 126–134. DOI: [10.5281/zenodo.14018656](https://doi.org/10.5281/zenodo.14018656).
4. Землянухин П. А., Кондратьев А. В., Свидельский С. С. Исследование характеристик формирователя шумового сигнала, как источника шума в многоканальном генераторе шума // Известия ЮФУ. Технические науки. 2020. № 5(215). С. 111–123. DOI: [10.18522/2311-3103-2020-5-111-123](https://doi.org/10.18522/2311-3103-2020-5-111-123).
5. Кулешов И. А., Спивак А. И., Львов А. Е. Анализ способов некриптографической защиты информации от побочных электромагнитных излучений и наводок на объектах связи // Техника средств связи. 2021. № 3(155). С. 67–73.
6. Ерохин В. В., Зайнулин Е. В. Автоматизированный программно-аппаратный комплекс в среде LabView для исследования эффективности подавления ПЭМИН // Вестник научных конференций. 2021. № 5-2(69). С. 23–24.
7. Паршуткин А. В., Неаскина М. Р. Повышение защищенности информации от утечки через побочные электромагнитные излучения // Вопросы кибербезопасности. 2022. № 3(49). С. 82–89. DOI: [10.21681/2311-3456-2022-3-82-89](https://doi.org/10.21681/2311-3456-2022-3-82-89).
8. Рушечников Я. И., Данилов В. В. Информационная технология радиомониторинга на основе программно-определяемой радиосистемы // Вестник Донецкого национального университета. Серия Г: Технические науки. 2020. № 1. С. 31–36.
9. Tretiakov I. A., Rushechnikov Ia. I., Danilov V. V. Algorithm for automated estimation of the volume of TEMPEST signals // Information, Control, and Communication Technologies (ICCT-2024) : Proceedings of the 8th International Scientific Conference (Vladikavkaz, 01–05 October 2024) / Vladikavkaz : V. A. Trapeznikov Institute of Control Sciences of RAS, 2024. P. 114–116.

SQL-ИНЪЕКЦИИ КАК УГРОЗА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ: СПЕЦИФИКА POSTGRESQL И OWASP TOP-10

Д. Д. Черезов, П. М. Каратаева, Д. А. Хватов, М. Д. Савкин

*Балтийский федеральный университет им. И. Канта,
Калининград, Россия ddcherezov@stud.kantiana.ru*

В работе рассматриваются SQL-инъекции как устойчивый класс угроз безопасности веб-приложений. Показано их влияние на основные компоненты модели CIA-триады и значимость для современных корпоративных и облачных систем. В качестве объекта исследования обоснован выбор PostgreSQL 17 с учётом её архитектурных особенностей, формирующих специфический профиль угроз. Рассмотрена классификация атак (classic, blind, time-based, UNION-based), проанализированы ограничения сигнатурных средств защиты и обозначена перспективность гибридных подходов, включающих методы глубокого обучения. Работа имеет как теоретическую ценность, так и прикладное значение для практики обеспечения информационной безопасности.

Ключевые слова: SQL-инъекции; PostgreSQL 17; CIA-триада.

SQL INJECTIONS AS A THREAT TO INFORMATION SECURITY: SPECIFICS OF POSTGRESQL AND THE OWASP TOP 10

D. D. Cherezov, P. M. Karataeva, D. A. Khvatov, M. D. Savkin

*Immanuel Kant Baltic Federal University,
Kaliningrad, Russia, ddcherezov@stud.kantiana.ru*

The paper examines SQL injections as a persistent class of web application security threats. Their impact on the core components of the CIA triad model and their significance for modern enterprise and cloud systems are demonstrated. PostgreSQL 17 is substantiated as the research subject, taking into account its architectural features that shape a specific threat profile. A classification of attacks (classic, blind, time-based, UNION-based) is presented, the limitations of signature-based protection mechanisms are analyzed, and the prospects of hybrid approaches incorporating deep learning methods are outlined. The study has both theoretical value and practical significance for the field of information security.

Keywords: SQL injections; PostgreSQL 17; CIA triad.

1. Введение

SQL-инъекции (SQL Injection, SQLi) на протяжении более двух десятилетий остаются одной из наиболее опасных и распространенных угроз безопасности веб-приложений. Их актуальность подтверждается международными рейтингами и отраслевыми отчетами. В OWASP Top-10 (2021) они

отнесены к категории A03:2021 – Injection [1], в списке CWE Top-25 (2022) уязвимость «SQL Injection» занимает третье место [2], а согласно отчёту Verizon DBIR (2023) около 12% атак на веб-приложения были связаны именно с этим видом атак [3]. В модели MITRE ATT&CK SQL-инъекции классифицируются как техника начального доступа и сбора данных, что подчёркивает их ключевую роль в сложных атаках на корпоративные и облачные системы [4].

Опасность SQL-инъекций определяется их универсальностью и множеством сценариев эксплуатации. Злоумышленники, внедряя произвольный SQL-код в запросы приложения, могут обходить механизмы аутентификации, извлекать конфиденциальные сведения, изменять или удалять данные и инициировать отказ в обслуживании. В научной литературе отмечается, что SQL-инъекции обладают «высокой скрытностью, вариативностью и способностью обходить классические средства защиты, что делает их обнаружение одной из важнейших задач информационной безопасности» [5, с. 2]. Ситуацию осложняют особенности отдельных систем управления базами данных. Так, в PostgreSQL показательными примерами являются использование функции `pg_sleep()` при time-based атаках и механизма COPY PROGRAM, позволяющего запускать внешние программы, что фактически открывает путь к удалённому выполнению кода (RCE) [6].

Выбор PostgreSQL в качестве объекта исследования обусловлен её широкой популярностью и применением в критически важных сервисах. Согласно Stack Overflow Developer Survey (2024), PostgreSQL признана самой «любимой» СУБД среди разработчиков [7]. Её используют Spotify, Apple, Reddit и ряд социальных сетей, что подтверждает стратегическую значимость данной системы для современных высоконагруженных приложений. Архитектурные особенности PostgreSQL — строгая типизация данных, использование метакоманд и многоверсионное управление конкурентным доступом (MVCC) — формируют специфический профиль угроз, требующий отдельного анализа в контексте SQL-инъекций.

Отечественные исследования также подтверждают актуальность проблемы: российские специалисты отмечают, что рост сложности веб-приложений и активное распространение облачных решений делают SQL-инъекции одной из наиболее устойчивых и опасных угроз, требующих применения многоуровневой защиты и интеграции механизмов мониторинга на уровне СУБД [8; 9; 12]. Исследование SQL-инъекций применительно к PostgreSQL 17 является актуальной задачей, позволяющей объединить международные подходы к классификации уязвимостей (OWASP, CWE, MITRE) с анализом архитектурных особенностей конкретной СУБД и выработкой рекомендаций по защите корпоративных и облачных приложений.

2. Основная часть

Классификация SQL-инъекций является важным элементом анализа угроз, так как различные типы атак используют разные особенности языка SQL и механизмы обработки запросов. В литературе и отраслевых отчётах выделяют четыре основных разновидности: classic, blind, time-based и UNION-based [6; 9; 10]. Классическая инъекция основывается на использовании таутологий, например конструкции `OR 1=1`, что позволяет изменять логику запроса и обходить аутентификацию при слабой фильтрации пользовательского ввода.

Слепые инъекции применяются в тех случаях, когда система не возвращает ошибок и не выводит результаты запросов; злоумышленник делает выводы о содержимом базы данных на основе различий в поведении приложения при выполнении условий «истина/ложь». Time-based-инъекции используют измерение задержки ответа, вызванной выполнением ресурсоёмких функций. В PostgreSQL для этого характерно применение функции `pg_sleep()`, которая активно используется в подобных атаках. UNION-based-инъекции строятся на объединении легитимного и вредоносного запроса при помощи оператора `UNION`, что позволяет извлекать системную информацию о базе данных, включая версию СУБД или имя текущего пользователя. Таблица отражает ключевые характеристики перечисленных классов атак и специфику их проявления в PostgreSQL.

Классификация SQL-инъекций

Тип атаки	Механизм	Пример	Специфика в PostgreSQL
Classic	Изменение логики запроса таутологией	<code>OR 1=1</code>	Работает при слабой фильтрации
Blind (Boolean)	Различие ответов при истинности/ложности	<code>AND substring(password, 1, 1)='a'</code>	Позволяет поэтапно извлекать данные
Time-based	Задержка выполнения	<code>pg_sleep(5)</code>	Поддержка <code>pg_sleep()</code> упрощает атаку
UNION-based	Объединение легитимных и вредоносных запросов	<code>UNION SELECT version()</code>	Доступ к системной информации

Для подтверждения изложенной классификации была проведена демонстрация классической SQL-инъекции. На тестовой таблице `users` (рис. 1) запрос `SELECT * FROM users WHERE login = " OR '1'=1"`; возвращает все записи, обходя аутентификацию (рис. 2).

```

postgres=# CREATE TABLE users (
postgres(#      id SERIAL PRIMARY KEY,
      login TEXT NOT NULL,
      password_hash TEXT NOT NULL,
      role TEXT DEFAULT 'user'
);

INSERT INTO users (login, password_hash, role)
VALUES
('admin', '5f4dcc3b5aa765d61d8327deb882cf99', 'admin'), -- passwo
rd
('user1', '12dea96fec20593566ab75692c994959', 'user'),
('user2', '5ebe2294ecd0e0f08eab7690d2a6ee69', 'user');
      id SERIAL PRIMARY KEY,
postgres(#      login TEXT NOT NULL,
      password_hash TEXT NOT NULL,
      role TEXT DEFAULT 'user'
);

```

Рис. 1. Создание таблицы и добавление тестовых пользователей в PostgreSQL

```

postgres=# SELECT * FROM users
postgres-# WHERE login = '' OR '1'='1';
WHERE login = '' OR '1'='1';

```

id	login	password_hash	role
1	admin	5f4dcc3b5aa765d61d8327deb882cf99	admin
2	user1	12dea96fec20593566ab75692c994959	user
3	user2	5ebe2294ecd0e0f08eab7690d2a6ee69	user

```

(3 rows)

postgres=# █

```

Рис. 2. Классическая SQL-инъекция с использованием таутологии (OR '1'='1')

Классификация SQL-инъекций демонстрирует разнообразие атакующих техник и объясняет, почему их обнаружение остается нетривиальной задачей. Для PostgreSQL значимыми оказываются встроенные механизмы — строгая типизация, функции задержки и расширенные возможности работы с системными командами, — что формирует специфический профиль угроз и требует особого внимания исследователей и практиков [6; 9; 10].

Особенности PostgreSQL в контексте SQL-инъекций определяются как архитектурными решениями данной СУБД, так и расширенными функциональными возможностями, которые при некорректной настройке или эксплуатации могут использоваться злоумышленниками. В отличие от MySQL, допускающей неявные преобразования типов, PostgreSQL опирается на строгую систему типизации данных, что в ряде случаев снижает вероятность успешных примитивных атак. Однако это преимущество не является абсолютной защитой, так как злоумышленники могут использовать явное приведение типов, например функции CAST() или оператор ::text, для обхода

фильтрации пользовательского ввода [5; 8]. Дополнительным фактором риска выступают метакоманды, характерные для клиентского инструмента `psql`. Несмотря на то, что они не могут быть вызваны напрямую через SQL-инъекцию, получение злоумышленником доступа к консоли открывает возможность быстрого изучения структуры базы данных, выполнения команд по экспорту информации и ускорения развития атаки [6].

Значительную угрозу для PostgreSQL представляют встроенные функции, такие как `pg_sleep()`, позволяющие приостанавливать выполнение запроса. Именно они используются в сценариях time-based blind SQL-инъекций, когда результат определяется по задержке ответа системы (рис. 3). Подобные механизмы существенно упрощают проведение атак и многократно отмечаются в специализированной литературе как характерная уязвимость PostgreSQL [4; 9].

```
postgres=# -- Проверка с использованием pg_sleep:
postgres=# SELECT CASE WHEN (substring(password_hash,1,1)='5')
                THEN pg_sleep(5) ELSE pg_sleep(0) END
FROM users WHERE id=1;
SELECT CASE WHEN (substring(password_hash,1,1)='5')
postgres-#                THEN pg_sleep(5) ELSE pg_sleep(0) END
FROM users WHERE id=1;
                THEN pg_sleep(5) ELSE pg_sleep(0) END
postgres-# FROM users WHERE id=1;
FROM users WHERE id=1;

pg_sleep
-----
(1 row)
```

Рис. 3. Пример использования функции `pg_sleep()` при реализации time-based blind SQL-инъекции в PostgreSQL

Наиболее опасным элементом выступает команда `COPY PROGRAM`, предоставляющая возможность копировать данные во внешнюю программу, исполняемую на сервере. При наличии соответствующих прав злоумышленник способен вызвать выполнение произвольных команд, что фактически означает переход от угрозы компрометации базы данных к удалённому выполнению кода (Remote Code Execution). Например, команда `COPY users TO PROGRAM 'cat /etc/passwd'`; может быть использована для извлечения системных файлов. Исследователи подчеркивают, что именно наличие подобных возможностей в PostgreSQL значительно расширяет атакующую поверхность и требует повышенного внимания со стороны администраторов [6; 10].

Архитектурные особенности PostgreSQL — строгая типизация данных, наличие метакоманд, встроенные функции задержки и механизм COPY PROGRAM — формируют уникальный профиль угроз. Их совокупность отличает PostgreSQL от других СУБД и определяет необходимость отдельного анализа в рамках исследования SQL-инъекций применительно к версии 17.

Рассматривая SQL-инъекции в модели угроз информационной безопасности, необходимо отметить, что они оказывают воздействие сразу на все три базовых компонента модели CIA-триады. Конфиденциальность нарушается вследствие несанкционированного доступа к данным, включая учетные записи, финансовые сведения и персональные данные пользователей. Целостность базы данных подрывается при модификации или удалении записей, а также при изменении ролей, что ведет к несанкционированному расширению прав. Доступность системы страдает при выполнении ресурсоемких запросов или использовании функций задержки, таких как `pg_sleep()`, что может вызвать отказ в обслуживании и недоступность системы для легитимных пользователей [5; 6].

Международные классификации подтверждают критичность данной уязвимости. В OWASP Top-10 (2021) SQL-инъекции отнесены к категории A03:2021 – Injection, одной из наиболее опасных для веб-приложений [1]. В рейтинге CWE Top-25 (2022) они занимают третью позицию, что указывает на их устойчивую распространённость и универсальность [2]. Кроме того, в модели MITRE ATT&CK SQL-инъекции классифицируются как техника начального доступа и сбора данных, что подчёркивает их значение в многоступенчатых атаках, включающих горизонтальное перемещение по сети и установку webshell-модулей [4].

Современные исследования подтверждают, что традиционные методы защиты, такие как сигнатурные межсетевые экраны приложений (WAF) и статический анализ кода, оказываются недостаточно эффективными против новых и обфусцированных форм атак. Как отмечают Sun и соавторы, правила, основанные исключительно на сигнатурах, демонстрируют низкую устойчивость к изменяющимся методам эксплуатации, тогда как использование методов глубокого обучения позволяет достигать точности детектирования выше 97–98% [5, с. 3]. Эту тенденцию подтверждают сравнительные исследования алгоритмов машинного обучения для обнаружения SQL-инъекций, выявляющие преимущества гибридных и ансамблевых подходов [11], а в отечественных работах подчёркивается значимость методов интеллектуального анализа данных и интеграции их в практики разработчиков и ИБ-подразделений [12]. Это указывает на необходимость внедрения комплексных стратегий защиты, сочетающих безопасные практики разработки, параметризацию запросов, строгую валидацию пользовательского ввода и интеллектуальные системы мониторинга в реальном времени.

SQL-инъекции следует рассматривать не только как локальную уязвимость уровня базы данных, но и как элемент системной модели угроз. Их наличие способно привести к масштабным последствиям, включая утечку конфиденциальной информации, компрометацию бизнес-логики приложения и полный контроль над сервером.

3. Заключение

Проведённый анализ подтверждает, что SQL-инъекции представляют собой устойчивую и чрезвычайно опасную угрозу информационной безопасности. Их актуальность подтверждается международными рейтингами (OWASP Top-10, CWE Top-25) и отраслевыми отчётами (Verizon DBIR), фиксирующими значительную долю успешных атак данного типа [1–3]. Особую роль SQL-инъекции играют в контексте современных систем управления базами данных, где архитектурные особенности могут усиливать воздействие атакующих сценариев.

Для PostgreSQL характерно сочетание особенностей, которое одновременно снижает риск примитивных атак и создаёт новые векторы эксплуатации. Строгая типизация ограничивает простейшие инъекции, однако использование явного приведения типов и функций, таких как `pg_sleep()` или `COPY PROGRAM`, открывает путь к более сложным сценариям, включая удалённое выполнение кода [5; 6]. Именно поэтому PostgreSQL требует отдельного анализа в рамках исследования SQL-инъекций. Выбор PostgreSQL 17 в качестве объекта исследования обусловлен её широким применением в критически важных сервисах и корпоративных приложениях. Данная версия отражает актуальное состояние развития системы и позволяет сопоставить теоретические положения международных классификаций угроз с реальными техническими особенностями популярной СУБД.

Практическая ценность проведённого исследования заключается в систематизации классификации атак, выявлении специфики PostgreSQL, повышающей риски эксплуатации, а также в обосновании необходимости комплексного подхода к защите. Перспективными направлениями являются внедрение современных средств обнаружения атак, включая методы машинного обучения и глубокие нейронные сети, ориентированные на работу в реальном времени и сокращение ложноположительных срабатываний — в том числе для PostgreSQL [10], а также сравнительно-аналитические и прикладные разработки, демонстрирующие эффективность ML-подходов при обнаружении SQL-инъекций [11; 12].

Исследование SQL-инъекций применительно к PostgreSQL 17 имеет как теоретическую значимость, так и прикладное значение для практики обеспечения информационной безопасности. В условиях, когда PostgreSQL активно

используется крупнейшими компаниями и в критических инфраструктурах, выявление и анализ подобных угроз становятся необходимыми элементами построения устойчивых и надёжных систем защиты данных.

Библиографические ссылки

1. OWASP Foundation. OWASP Top Ten Web Application Security Risks – 2021. URL: <https://owasp.org/Top10> (date of access: 06.09.2025).
2. MITRE. 2022 CWE Top 25 Most Dangerous Software Weaknesses. URL: https://cwe.mitre.org/top25/archive/2022/2022_cwe_top25.html (date of access: 06.09.2025).
3. 2023 Data Breach Investigation Report : Public Sector Snapshot. Verizon, 2023.
4. Sun H., Du Y., Li Q. Deep Learning-Based Detection Technology for SQL Injection Research and Implementation // Applied Sciences. 2023. Vol. 13. Article no. 9466. DOI: [10.3390/app13169466](https://doi.org/10.3390/app13169466).
5. Exploring the Intersection of Network Security and Database Communication: a PostgreSQL Socket Connection Case Study / S. Ye [et al.] // Transactions on Computer Science and Intelligent Systems Research. 2023. Vol. 3.
6. Salunke S. V., Ouda A. A Performance Benchmark for the PostgreSQL and MySQL Databases // Future Internet. 2024. Vol. 16. Article no. 382. DOI: [10.3390/fi16100382](https://doi.org/10.3390/fi16100382).
7. Li T., Chen J., Zhang W. SQL Injection Detection Using Hybrid Deep Learning Models // Journal of Information Security and Applications. 2024. Vol. 80. Article no. 103723.
8. Wang Y., Liu H., Zhao F. AI-Enhanced Methods for SQL Injection Prevention in Cloud Databases // Computers & Security. 2024. Vol. 136. Article no. 103481.
9. Иванов С. А., Кузнецов А. В. Современные тенденции защиты баз данных веб-приложений от SQL-инъекций // Информационные технологии. 2023. № 5. С. 15–27.
10. Zhang Q., Zhou Y., Li K. Optimizing Convolutional Neural Networks for Real-Time SQL Injection Detection in PostgreSQL Databases // ArXiv preprint. 2023. arXiv:2312.04941.
11. Chen L., Wu Z., Fang H. Comparative Study of SQL Injection Detection Algorithms Using Machine Learning // Procedia Computer Science. 2024. Vol. 222. P. 456–465.
12. Петров И. В., Сидоров Н. П. Методы интеллектуального анализа для выявления атак SQL-инъекций // Вестник КГУ. Серия: Информационная безопасность. 2024. № 2. С. 45–54.

РАЗДЕЛ II

БИОИНФОРМАТИКА И ПРИЛОЖЕНИЯ

DE NOVO ГЕНЕРАЦИЯ ПОТЕНЦИАЛЬНЫХ ИНГИБИТОРОВ БЕЛКА KASА MYCOBACTERIUM TUBERCULOSIS

А. В. Гончар¹⁾, К. В. Фурс¹⁾, А. В. Тузиков¹⁾, А. М. Андрианов²⁾

¹⁾ Объединенный институт проблем информатики,
Национальная академия наук Беларуси,

Минск, Республика Беларусь, tuzikov@newman.bas-net.by

²⁾ Институт биоорганической химии, Национальная академия наук Беларуси,
Минск, Республика Беларусь, alexande.andriano@yandex.ru

Разработана генеративная состязательная нейронная сеть с частичным привлечением учителя, обученная на графовых эмбедингах и использованная для генерации потенциальных ингибиторов фермента KasA микобактерии туберкулеза. Проведена оценка потенциала ингибиторной активности новых сгенерированных нейронной сетью соединений против белка KasA методами молекулярного моделирования. На основе анализа полученных данных были идентифицированы шесть соединений, перспективных для создания новых эффективных препаратов для терапии лекарственно-устойчивых форм туберкулеза.

Ключевые слова: Микобактерия туберкулеза; белок KasA; генеративная состязательная нейронная сеть; графовые эмбединги; виртуальный скрининг; молекулярный докинг; молекулярная динамика.

DE NOVO GENERATION OF POTENTIAL INHIBITORS OF THE KasA PROTEIN OF MYCOBACTERIUM TUBERCULOSIS

H. V. Hanchar^{a)}, K. V. Furs^{a)}, A. V. Tuzikov^{a)}, A. M. Andrianov^{b)}

^{a)} United Institute of Informatics Problems, National Academy of Sciences of Belarus,
Minsk, Republic of Belarus, tuzikov@newman.bas-net.by

^{b)} Institute of Bioorganic Chemistry, National Academy of Sciences of Belarus,
Minsk, Republic of Belarus, alexande.andriano@yandex.ru

A semi-supervised generative adversarial network was developed, trained on graph embeddings, and applied to the generation of potential inhibitors of the KasA enzyme of Mycobacterium tuberculosis. The inhibitory potential of the newly generated compounds against the KasA protein was evaluated using molecular modeling methods. Based on the analysis of the obtained data, six compounds were identified as promising candidates for the development of new effective drugs for the treatment of drug-resistant forms of tuberculosis.

Keywords: Mycobacterium tuberculosis; KasA protein; generative adversarial network; graph embeddings; virtual screening; molecular docking; molecular dynamics.

1. Введение

Туберкулез (ТБ) по-прежнему входит в число ведущих причин смертности во всем мире и представляет особую угрозу для пациентов с ВИЧ и сахарным диабетом [1]. Рост лекарственной устойчивости существенно осложняет терапию и делает актуальными исследования по созданию новых противотуберкулезных препаратов. На ранних стадиях разработки лекарств все более востребованы технологии виртуального скрининга и методы машинного обучения, позволяющие ускорить поиск активных соединений, сократить время и затраты, необходимые для их создания.

Среди потенциальных мишеней особое значение имеет фермент KasA, играющий ключевую роль в синтезе жирных кислот клеточной стенки микобактерии туберкулеза (МБТ). Известно, что потеря активности белка KasA приводит к лизису клеток бактерии, свидетельствуя о том, что этот фермент имеет ключевое значение для жизненного цикла МБТ и, следовательно, является важной терапевтической мишенью для разработки новых эффективных ингибиторов лекарственно-устойчивого ТБ [2].

Цель настоящего исследования заключалась в создании генеративной модели нейронной сети для *de novo* дизайна малых молекул, потенциально активных против фермента KasA МБТ. Для достижения этой цели были проведены исследования, которые включали:

- разработку архитектуры нейронной сети для генерации низкомолекулярных химических соединений, обладающих высоким сродством к белку KasA;
- формирование обучающей библиотеки малых молекул, содержащих элементы структуры, способные к селективным взаимодействиям с активным центром фермента;
- обучение модели и ее тестирование на наборе соединений из созданной молекулярной библиотеки;
- генерацию новых молекул, потенциально активных против белка KasA;
- молекулярный докинг сгенерированных молекул с целевым белком;
- молекулярную динамику перспективных соединений;
- оценку физико-химических свойств лучших соединений и отбор потенциальных ингибиторов МБТ.

2. Материалы и методы

Для решения поставленных задач нами была разработана модель генеративной состязательной нейронной сети с частичным привлечением учителя (SGAN, Semi-Supervised Generative Adversarial Network) (рис. 1), которая использует графовые эмбединги, полученные из латентного

пространства вариационного автоэнкодера JTVAE (Junction Tree Variational Autoencoder) [3]. Обучение SGAN проводили на наборе молекул из обучающей выборки с использованием в неявном виде значений энергии связывания с белком KasA. Молекулы в обучающем наборе данных разделяли на две группы, включавшие соответственно соединения с низкими (ниже $-8,2$ ккал/моль) и высокими значениями энергии связывания с целевым белком, рассчитанными с помощью методов молекулярного докинга. Это позволило SGAN генерировать новые молекулы, похожие на соединения с высоким сродством к мишени.

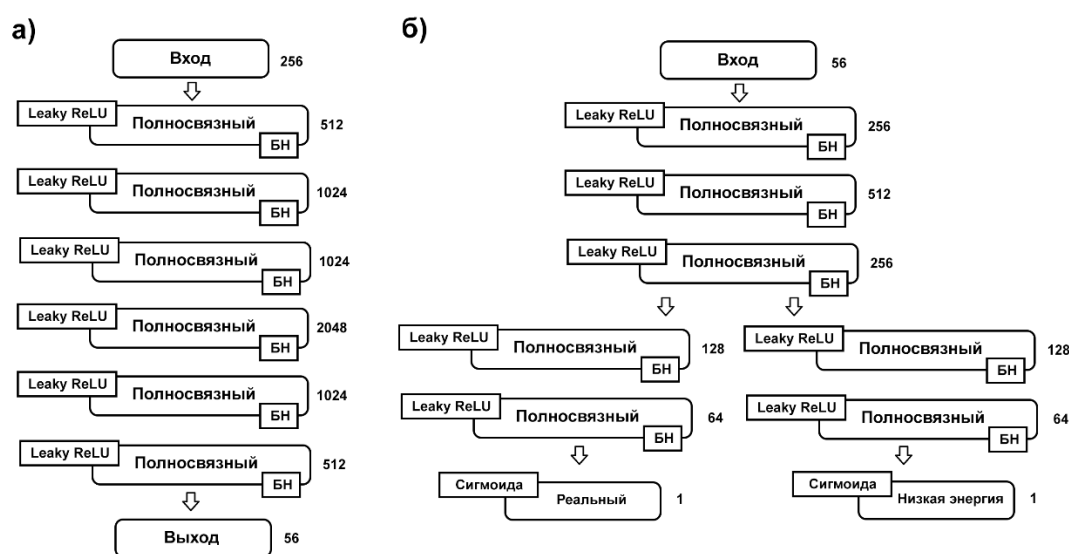


Рис. 1. Архитектура SGAN:
а – генератор; б – дискриминатор

Формирование обучающего набора данных

Для формирования обучающего набора данных был проведен виртуальный скрининг библиотек Zinc15, ChemSpace и ChemDiv на веб-сервере Pharmit с использованием трех фармакофорных моделей, построенных на основе комплексов белка C171Q KasA с ингибиторами TLM и TLM5 (PDB ID: 4C6X и 4C72 соответственно). С использованием Python 3 и программного пакета RDKit (<https://www.rdkit.org>) из отобранных молекул были удалены дубликаты, и для каждого соединения получены канонические представления SMILES. В результате размер обучающей молекулярной библиотеки, сформированной с помощью фармакофорного анализа баз данных веб-сервера Pharmit, составил 58 815.

Для генерации графовых эмбедингов, необходимых для обучения SGAN, из формата SMILES, полученного на этапе фармакофорного поиска, использовали предобученную модель нейронной сети JTVAE,

которую разработчики обучали на выборке из 250 000 соединений, отобранных случайным образом в библиотеке Zinc15.

Для расчета значений свободной энергии связывания молекул из обучающего набора данных с целевым белком с помощью программы QuickVina2 был проведен молекулярный докинг этих лигандов с ферментом C171Q KasA в свободном состоянии. Структура белка извлекалась из комплекса C171Q KasA в кристалле с ингибитором TLM5 (ID PDB: 4C72). Структуры белка и лигандов были подготовлены к расчетам с помощью программного пакета MGLTools (<https://ccsb.scripps.edu/mgltools/>). Ячейка для докинга включала малонил-связывающий сайт KasA и имела следующие размеры: $\Delta X = 20,67 \text{ \AA}$, $\Delta Y = 24,8 \text{ \AA}$, $\Delta Z = 16,46 \text{ \AA}$ с центром $X = -7,24 \text{ \AA}$, $Y = -19,9 \text{ \AA}$, $Z = 6,75 \text{ \AA}$. Значение параметра охвата конформационного пространства, определяющего широту поиска, было установлено равным 100.

Обучение SGAN

Для 58815 малых молекул из сформированного обучающего набора данных с помощью сети JTVAE были получены графовые эмбединги, а затем эта виртуальная библиотека была разделена на две выборки – тренировочную и тестовую, состоящие из 47052 и 11763 векторов соответственно. Путем многочисленных экспериментов параметр, задающий количество эпох обучения, был выбран равным 150. Отношение частоты обучения генератора к дискриминатору было выбрано равными 0,3 к 0,7.

Функции потерь дискриминатора и генератора представлены в формулах (1) и (2):

$$D_{loss} = BCE(D_{out1}(G(noise)), 0) + BCE(D_{out2}(real\_data), energy\_class), \quad (1)$$

$$G_{loss} = BCE(D_{out1}(G(noise)), 1), \quad (2)$$

где BCE – функция потерь бинарная кросс-энтропия; D – дискриминатор; D_{out1} – выход дискриминатора, отвечающий за предсказание реальности молекулы; D_{out2} – выход дискриминатора, отвечающий за предсказание класса энергии молекулы; G – генератор; $noise$ – вектор гауссовского шума размерности 256; $real\_data$ – вектор, соответствующий молекуле из тренировочной выборки; $energy\_class$ – 1 или 0 в зависимости от того, принадлежит ли молекула классу с низкой энергией связывания или нет.

Графики функций потерь генератора и дискриминатора SGAN представлены на рис. 2. Сплошная синяя линия соответствует потерям генератора, а красная пунктирная линия – потерям дискриминатора. Анализ

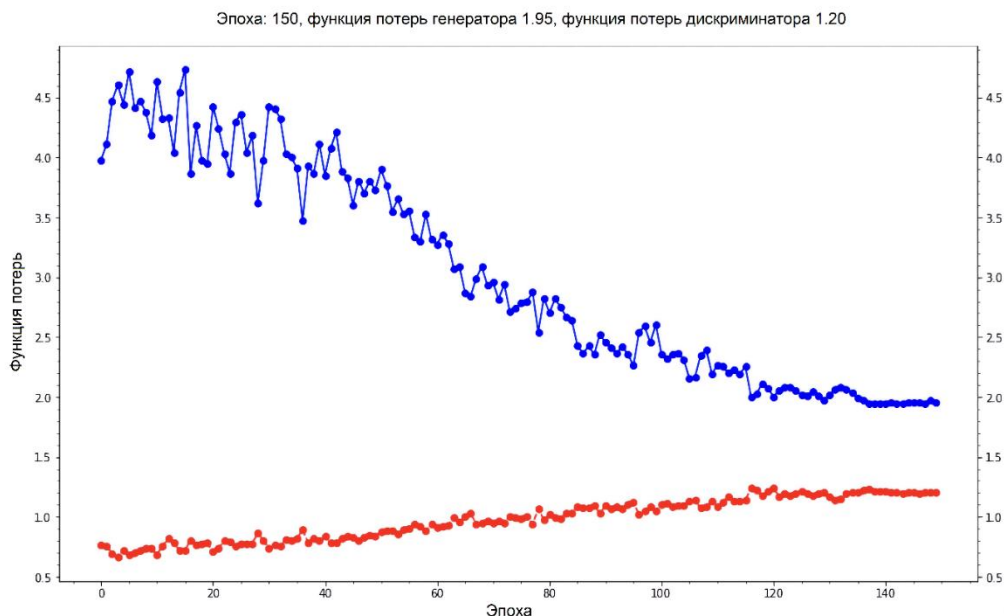


Рис. 2. Функции потерь генератора (синяя линия) и дискриминатора (красная линия) SGAN

графиков функций потерь дает основания полагать, что модель была обучена корректно, так как функции потерь сходятся, но не пересекаются.

На основе случайных векторов длины 256, имеющих гауссовское распределение, с помощью генератора SGAN было сгенерировано 200 000 векторов длиной 56 и получены соответствующие предсказания дискриминатора. Для оценки генеративных возможностей SGAN были выбраны два региона векторов. Первый (перспективный) регион включал вектора с вероятностью реальности выше 50% и вероятностью низкого энергетического класса выше 50%, а второй регион представлял область с вероятностью реальности молекул выше 50% и вероятностью низкого энергетического класса ниже 50%. Всего было получено 2565 и 2745 графовых эмбедингов для первого и второго регионов соответственно, которые после декодирования в JTVAE, удаления дубликатов и проверки корректности дали 1755 и 1882 уникальные молекулы.

Для выбора наиболее перспективных лигандов среди всех сгенерированных SGAN соединений (из обеих областей), значения свободной энергии связывания были переоценены с использованием оценочных функций RFScore-4 и NNScore 2.0. На основе предсказанных с помощью оценочных функций AutoDock Vina, RFScore-4, NNScore 2.0 значений энергии связывания для каждого соединения был рассчитан экспоненциальный консенсусный ранг (ECR) [4] и отобраны 20 лучших соединений для молекулярно-динамических расчетов.

Молекулярно-динамическое (МД) моделирование комплексов лиганд/белок проводили в воде с использованием Amber18 [5] и силовых

полей Amber ff14SB (белок) и GAFF (лиганды). После минимизации, постепенного нагрева до 300 К, уравнивания давления и достижения равновесия система моделировалась в ансамбле NPT при 300 К и 1 атм в течение 200 нс. В качестве контрольного соединения использовали ингибитор KasA TLM5 ($K_d = 25,6$ мкМ).

Свободную энергию связывания рассчитывали методом MM/GBSA для 40 комплексов, извлечённых из последних 80 нс траекторий (шаг 2 нс). Энтропийный вклад определяли с помощью Nmode из AmberTools19, а анализ траекторий проводили модулем CPPTRAJ.

3. Результаты и обсуждение

Результаты молекулярного докинга для новых сгенерированных соединений

Для оценки потенциала разработанной модели нейронной сети были получены прогнозные показатели дискриминатора для 11 763 векторов из тестового набора данных, из которых в перспективную и вторую области вошли соответственно 3438 и 4448 молекулярных эмбеддингов. Для сгенерированных молекул и соединений из тестового набора был проведен молекулярный докинг с белком KasA с использованием того же вычислительного протокола, что и для соединений из обучающего набора данных. Результаты молекулярного докинга представлены в табл. 1.

Таблица 1

Результаты молекулярного докинга для сгенерированных соединений и соединений из тестовой выборки

1 – перспективная область 2 – вторая область		Количество соединений	Процент соединений с низкой энергией связывания ($< -8,2$ ккал/моль)	Средняя энергия связывания, (ккал/моль)
Сгенерированные соединения	1	1755	67	-8,5
	2	1882	40	-8,0
Соединения из тестовой выборки	1	3438	76	-8,7
	2	4448	21	-7,6

Данные табл. 1 указывают на то, что дискриминатор SGAN может с высокой точностью предсказывать класс энергии на реальных (не сгенерированных) данных, что подтверждает корректность обучения модели. Полученные результаты свидетельствуют о способности сети SGAN генерировать различные химические соединения с высоким сродством к белку KasA, особенно в перспективном регионе. Эти данные также указывают на то, что разделение молекул на классы на основе энергии связывания при обучении модели является эффективной стратегией для создания соединений с высоким сродством к целевому белку.

Анализ полученных траекторий молекулярной динамики позволил идентифицировать 12 соединений, у которых значения свободной энергии связывания (ΔG) оказались ниже, чем у контрольного соединения TLM5, рассчитанного по тому же вычислительному протоколу. Ниже представлены значения свободной энергии связывания и их стандартные отклонения по результатам молекулярной динамики для шести наиболее перспективных соединений (рис. 3), физико-химические параметры которых представлены в табл. 2. Учитывая стандартную ошибку метода MM/GBSA, составляющую около 2,9 ккал/моль, полученные данные указывают на то, что сродство этих соединений к KasA значительно выше по сравнению с ингибитором KasA TLM5. Также из табл. 2 видно, что отобранные соединения полностью соответствуют требованиям, предъявляемым к потенциальному лекарству согласно «правилу пяти» Липинского [6].

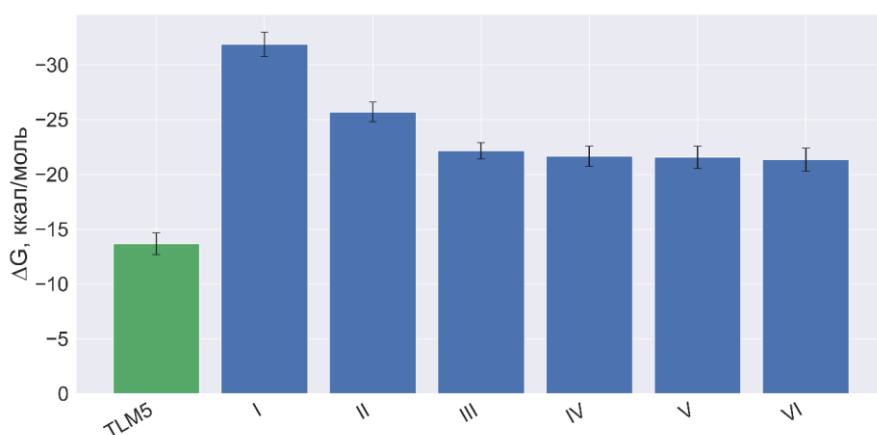


Рис. 3. Средние значения ΔG , рассчитанные для динамических моделей комплексов лиганд/KasA

Таблица 2

Физико-химические параметры идентифицированных соединений, связанные с «правилом пяти» Липинского [2; 4], и их синтетическая доступность

Лиганд	Химическая формула	Молекулярная масса, Да	LogP _{o/w}	Синтетическая доступность	Число доноров водородной связи	Число акцепторов водородной связи
I	C ₂₃ H ₂₉ N ₃ O ₄ S	443,56	3,14	5,15	1	5
II	C ₂₃ H ₃₀ N ₆ O ₃	438,52	2,58	4,22	0	6
IV	C ₁₈ H ₂₂ N ₆ O ₂ S	386,47	1,78	4,06	1	5
V	C ₁₉ H ₁₄ FN ₅ O ₂ S	395,41	2,81	3,47	1	6
III	C ₂₃ H ₂₆ N ₃ O ₃ ⁺	392,47	2,69	3,8	2	5
VI	C ₂₃ H ₃₅ N ₅ O ₂	413,56	3,22	4,95	1	5

Примечание. Приведенные данные получены с помощью веб-сервера открытого доступа SwissADME (<http://www.swissadme.ch>); LogP_{o/w} – липофильность соединения.

4. Заключение

В ходе проведенного исследования была разработана генеративная состязательная нейронная сеть (SGAN) с частичным привлечением учителя, адаптированная для поиска потенциальных ингибиторов фермента KasA, являющегося ключевым элементом биосинтеза клеточной стенки микобактерии туберкулеза.

Для реализации проекта была сформирована молекулярная библиотека соединений, содержащая низкомолекулярные аналоги известных ингибиторов KasA. В процессе исследования проведена оценка генеративных возможностей модели, выполнен молекулярный докинг сгенерированных соединений, а также оценка их сродства к мишени с использованием оценочных функций AutoDock Vina, RFScore-4 и NNScore 2.0. Проведенные МД расчеты подтвердили стабильность взаимодействия перспективных соединений с ферментом KasA, что делает их перспективными кандидатами на дальнейшие экспериментальные исследования.

Результаты работы позволили выделить шесть соединений, обладающих наилучшими характеристиками по сродству к KasA и удовлетворяющих требованиям фармакокинетики. Эти соединения могут стать основой для дальнейших экспериментальных исследований, направленных на разработку новых эффективных препаратов для терапии лекарственно-устойчивых форм туберкулеза.

Работа выполнена при поддержке Консорциума и Программы по разработке портала по лекарственно устойчивому туберкулезу (<https://tbportals.niaid.nih.gov>) и поддержана Белорусским республиканским фондом фундаментальных исследований (проекты Ф23-007, Ф24КИТГ-016).

Библиографические ссылки

1. WHO. Global tuberculosis report 2024. World Health Organization, 2024. URL: <https://www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports/global-tuberculosis-report-2024> (date of access 01.09.2025).
2. Identification of KasA as the cellular target of an anti-tubercular scaffold / K. A. Abrahams [et al.] // Nat. Commun. 2016. Vol. 7. Article no. 12581. DOI: [10.1038/ncomms12581](https://doi.org/10.1038/ncomms12581).
3. Jin W., Barzilay R., Jaakkola T. Junction tree variational autoencoder for molecular graph generation // International conference on machine learning. 2018. P. 2323–2332.
4. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking / K. Palacio-Rodriguez [et al.] // Sci. Rep. 2019. Vol. 9. Article no. 5142.
5. AMBER 2018 / D. A. Case [et al.]. San Francisco : University of California, 2018.
6. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings / C. A. Lipinski [et al.] // Advanced Drug Delivery Reviews. 2001. Vol. 46, No. 1–3. P. 3–26. DOI: [10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0).

ИДЕНТИФИКАЦИЯ НЕКАНОНИЧЕСКИХ ПОЛИПЕПТИДОВ ЛЕЙКОЗНЫХ КЛЕТОК ЧЕЛОВЕКА, ОСНОВАННАЯ НА К-МЕРНОМ АНАЛИЗЕ ДАННЫХ RNA-SEQ

Е. В. Гузова, В. В. Гринев

*Белорусский государственный университет,
Минск, Беларусь, grinev_vv@bsu.by*

В статье описан комбинированный анализ транскриптомных и протеомных данных, основанный на к-мерном подходе. Результатом применения такого подхода стал каталог неканонических полипептидов, обнаруженных в лейкозных клетках человека. Обнаруженные полипептиды имеют менее 100% гомологии с известными белками человека и могут рассматриваться как новые. В общей сложности было детектировано 33 новых полипептида и описано четыре источника их образования.

Ключевые слова: лейкозные клетки человека; транскриптом; протеом; новые полипептиды; к-меры.

IDENTIFICATION OF NON-CANONICAL POLYPEPTIDES FROM HUMAN LEUKEMIA CELLS BASED ON K-MER ANALYSIS OF RNA-SEQ DATA

K. V. Huzava, V. V. Grinev

*Belarusian State University,
Minsk, Belarus, grinev_vv@bsu.by*

The article describes a combined analysis of transcriptomic and proteomic data based on the k-mers approach. A catalog of non-canonical polypeptides expressed in human leukemia cells was developed using this approach. The detected polypeptides have less than 100% homology with known human proteins and can be considered as new ones. In total, 33 new polypeptides were detected and four sources of their formation were described.

Keywords: human leukemia cells; transcriptome; proteome; new polypeptides; k-mers.

1. Введение

Развитие методов молекулярной биологии, генной инженерии и биоинформатики обеспечило становление новых направлений иммунотерапии онкологических заболеваний. К настоящему времени опубликованы результаты ряда успешных клинических испытаний адаптивных TCR(CAR)

Т-клеток, а также вакцин на основе дендритных клеток [1]. В отношении лейкозов их применение сегодня сдерживается главным образом отсутствием молекулярных мишеней, способных вызывать устойчивый иммунный ответ. В качестве таковых принято рассматривать опухоль-ассоциированные и опухоль-специфические поверхностные антигены. Первые сверхэкспрессированы в раковых клетках, однако обнаруживаются и на поверхности нормальных клеток, ввиду чего их использование затруднено проявлением цитотоксичности по отношению к последним [2]. В связи с этим предпочтение отдается опухоль-специфическим антигенам, или неоантигенам, порождаемым генными мутациями, возникающими в ходе малигнизации клеток, альтернативным сплайсингом молекул РНК, а также изменениями на уровне посттрансляционной модификации белков [3].

Некоторые подходы к идентификации неоантигенов основываются на реконструкции аминокислотных последовательностей полноразмерных аберрантных белков. Однако в данном случае необходимо понимать, что алгоритмы сборки кодирующих их молекул РНК, равно как и алгоритмы, используемые для предсказания у последних кодирующего потенциала, несовершенны и потому не позволяют оценить все многообразие белков, экспрессирующихся в лейкозных клетках. В то же время известно, что при поиске молекулярных мишеней, пригодных для иммунотерапии, интерес представляют пептиды длиной 8–16 а.о., способные стабильно взаимодействовать с молекулами главного комплекса гистосовместимости класса I. В связи с этим мы задались целью разработать подход, альтернативный полноразмерным сборкам, но обеспечивающий эффективную идентификацию лейкозоспецифических полипептидов.

2. Материалы и методы

Обобщенная схема примененного биоинформатического конвейера изображена на рис. 1. Для реконструкции лейкозоспецифических фрагментов зрелых молекул РНК использовалась методика, изложенная в работе [4]. В соответствии с ней в качестве эталонных были взяты публичные данные секвенирования РНК, экстрагированных из медуллярных эпителиальных клеток тимуса (mTECs, от англ. medullary thymic epithelial cells) и миелоидных клеток-предшественников (MPCs, от англ. myelocytic precursor cells) (6 и 11 образцов соответственно). В роли исследуемых клеток выступили клетки модельной линии острого миелоидного лейкоза Kasumi-1, положительной по транслокации t(8;21)(q22;q22). Транскриптомные данные, ранее описанные в работе [5], представляли собой результаты шести независимых повторов парноконцевого секвенирования

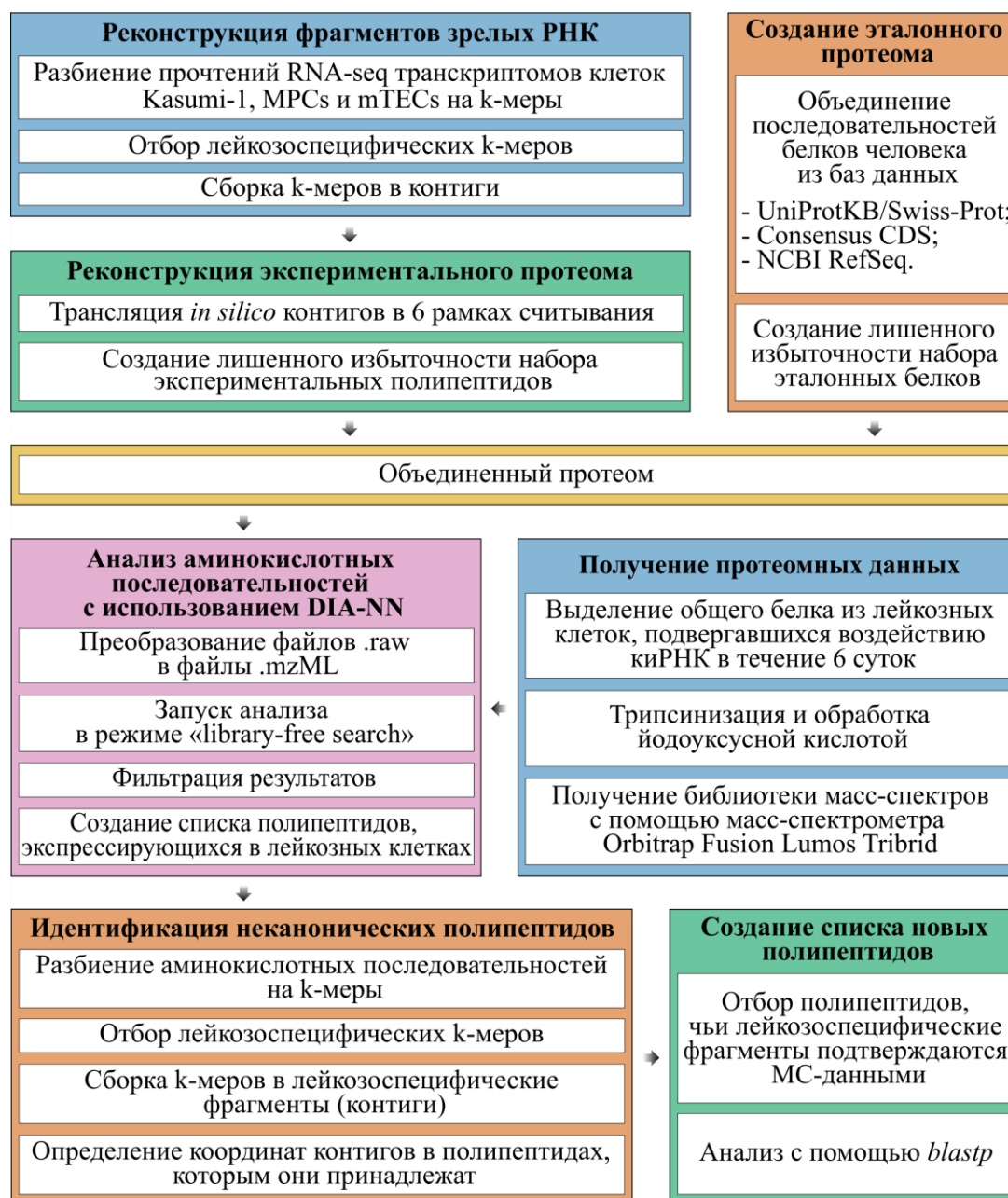


Рис. 1. Основные этапы идентификации новых полипептидов в лейкозных клетках человека

молекул РНК, выделенных из этих клеток (по три библиотеки прочтений для клеток с нокдауном онкогена *RUNX1::RUNX1T1* и без него).

Данные RNA-Seq описанных типов клеток проходили стандартную процедуру предпроцессинга. Затем прочтения каждой из включенных в анализ библиотек разбивались на k-меры протяженностью 33 нуклеотида. Из списка лейкозных k-меров для дальнейшей работы отбирались лишь встречающиеся более 2 раз в каждом образце одной экспериментальной группы. Подобная фильтрация позволила снизить количество шума,

представленного последовательностями, несущими ошибочно прочитанные основания.

На следующем этапе из k-меров лейкозных клеток исключались k-меры mTECs и MPCs. Полученные таким образом лейкозоспецифические k-меры собирались в контиги. Последние ввиду отсутствия информации о цепи ДНК, выступившей в роли матрицы для их синтеза, подвергались трансляции *in silico* в шести рамках считывания (+0, +1 и +2 в обеих ориентациях).

Поиск среди предсказанных *in silico* полипептидов тех, чья экспрессия подтверждается экспериментальными данными, осуществлялся посредством нейронной сети *DIA-NN* [6]. На вход ей подавался файл *.mzML, содержащий результаты масс-спектрометрического (МС) анализа тотального белка, выделенного из клеток Kasumi-1 и прошедшего обработку трипсином, а также йодоуксусной кислотой. Кроме того, в качестве входного выступил FASTA файл, содержащий последовательности предсказанных *in silico* полипептидов и эталонных белков человека. Общее количество последних, извлеченных из баз данных UniProtKB/Swiss-Prot, Consensus CDS и NCBI RefSeq, составило 105681.

Анализ запускался в режиме “library-free search”. Для обработки возвращаемых *DIA-NN* результатов была разработана R-функция *parseDIANNoutput()*, в основе которой лежит фильтрация данных по нескольким параметрам:

- 1) условная вероятность случайного события при идентификации пептида $\leq 0,01$;
- 2) q-значение для пептидов, детектируемых в каждом запуске, $\leq 0,01$;
- 3) количественное содержание пептида выше нуля не менее чем в 3 из 5 образцов одной экспериментальной группы;
- 4) q-значение для белков, детектируемых в каждом запуске, $\leq 0,01$;
- 5) число уникальных пептидов, подтверждающих экспрессию белка, должно быть ≥ 1 .

Аминокислотные последовательности полипептидов с подтвержденной экспрессией разбивались на k-меры длиной 8 а.о. Аналогичная работа проводилась с последовательностями эталонных белков, после чего из списка лейкозных k-меров вычитались k-меры эталонных белков. Прошедшие подобную фильтрацию лейкозоспецифические k-меры далее собирались в контиги. Каждая из описанных манипуляций осуществлялась с помощью R-функции собственной разработки [7].

Наконец, для создания списка неканонических полипептидов, экспрессируемых лейкозными клетками, собранные белковые фрагменты совместно с МС-пептидами, подтверждающими их экспрессию, подвергались анализу с помощью биоинформатического инструмента *blastp* [8].

Полипептид классифицировался как новый в том случае, если принадлежащий ему лейкозоспецифический фрагмент, равно как и выравнивающиеся относительно него МС-пептиды, имели идентичность менее 100% по отношению к любым известным белкам человека.

3. Результаты и их обсуждение

Исходный список уникальных лейкозных k-меров включил в себя свыше 64 млн последовательностей. В результате их негативной фильтрации с использованием k-меров mTECs и MPCs это число было сокращено до 916764 (1,41%). Ожидаемо более существенный вклад на данном этапе был внесен k-мерами MPCs (рис. 2, а).

В ходе сборки лейкозоспецифических k-меров в более протяженные фрагменты и последующей их трансляции *in silico* было получено 618910 полипептидов протяженностью 8 а.о. и более. Характер распределения их длин относительно длин эталонных белков человека (рис. 2, б) свидетельствует о том, что подавляющее большинство предсказанных *in silico* полипептидов представляли собой не полноразмерные белки, а лишь их фрагменты. Очевидно, это ограничивало количество МС-пептидов, способных подтвердить их экспрессию, и должно было учитываться при анализе результатов работы *DIA-NN*. По этой причине в настоящей работе для подтверждения экспрессии отдельного полипептида достаточным считалось наличие хотя бы одного соответствующего МС-пептида.

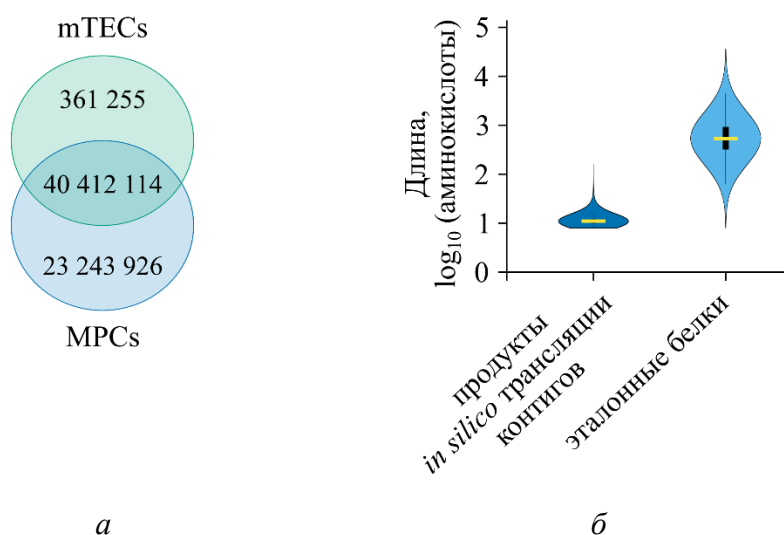


Рис. 2. k-меры mTECs и MPCs, для которых были обнаружены совпадения среди лейкозных k-меров (а); а также распределение длин экспериментальных полипептидов (б)

В конечном итоге каталог экспрессирующихся в лейкозных клетках полипептидов сократился до 318 записей. Однако более детальное изучение всех предсказанных *in silico* и подвергнутых анализу полипептидов показало, что для 371759 (60%) из них обнаружение пептидов в МС-данных оказалось невозможным ввиду особенностей организации их первичной структуры. Это, в свою очередь, позволяет предположить, что среди реконструированных фрагментов РНК число последовательностей, обладающих кодирующим потенциалом, на самом деле может быть выше.

Из 318 полипептидов с подтвержденной экспрессией 33 по результатам анализа с использованием *blastp* были классифицированы как неканонические. При этом с помощью инструмента *BLAT* удалось установить генную принадлежность только для 16 фрагментов РНК, кодирующих новые полипептиды (рис. 3). Причем 2 из них картировались на участки генома, в которых перекрывалось несколько генов. Поскольку установить, какому из перекрывающихся генов принадлежит фрагмент, не представлялось возможным, далее они будут рассматриваться как самостоятельные последовательности РНК (в связи с чем их количество увеличилось до 18).



Рис. 3. Результаты анализа генной принадлежности фрагментов зрелых РНК, кодирующих новые экспериментальные полипептиды

Интересно, что 5 из 18 фрагментов представляли собой РНК-продукты генов, у которых на сегодняшний день, согласно аннотациям Ensembl, не было выявлено кодирующего потенциала. Остальные же являлись результатом транскрипции белок-кодирующих генов и

принадлежали различным областям соответствующих полноразмерных молекул РНК.

Так, 5 фрагментов выравнивались относительно транслируемых областей. При этом несоответствия между первичной структурой канонического белка и полипептида, кодируемого тем или иным фрагментом РНК, обуславливались наличием однонуклеотидных замен (таблица). Среди них наиболее интересная – трансверсия, выявленная в протоонкогене *PBXIP1*, высокая экспрессия которого ассоциирована с прогрессией ряда онкологических заболеваний [9]. На уровне белка она влечет за собой замену гидрофобного лейцина на гидрофильный глутамин, что потенциально может отразиться на пространственной укладке и функционировании белка.

Однонуклеотидные замены, выявленные в кодирующих областях генов

Ген	Хромосома	Позиция в хромосоме	Вариант нуклеотидной замены	Вариант аминокислотной замены	Расстояние Грэнтэма
<i>PBXIP1</i>	1	154946463	A>T	L>Q	113
<i>FKBP5</i>	6	35637137	C>T	G>S	56
<i>STIM1</i>	11	4091617	C>T	T>I	89
<i>IRAK4</i>	12	43772969	G>A	R>Q	43
<i>WIZ</i>	19	15425333	C>T	R>H	29

Еще 5 фрагментов РНК принадлежали интронам, ранее неаннотированным в базе данных Ensembl как сохраняемые. В 3 из них были обнаружены стоп-кодона.

Наконец, оставшиеся 3 фрагмента РНК принадлежали нетранслируемым областям известных транскриптов соответствующих генов, что позволяет выдвинуть гипотезу о возможности их трансляции в составе малых открытых рамок считывания.

4. Заключение

Таким образом, разработанный авторами статьи [4] k-мерный анализ транскриптомных данных был успешно дополнен собственным алгоритмом перехода от нуклеотидных последовательностей к аминокислотным и дальнейшим анализом полученного массива протеомных данных. Применение такого комбинированного биоинформатического подхода позволило создать каталог неканонических полипептидов, экспрессирующихся в лейкозных клетках человека. На основе полученных результатов может быть выделено несколько их источников: 1) вариабельность нуклеотидных последовательностей генов, 2) альтернативный сплайсинг РНК, 3) малые

открытые рамки считывания, 4) трансляционная активность РНК-продуктов генов, ранее аннотированных как некодирующие.

Следующим шагом может стать дополнение описанного конвейера алгоритмом поиска среди детектированных полипептидов неоантигенов, пригодных для противолейкозной терапии.

Библиографические ссылки

1. Advances in cancer immunotherapy: historical perspectives, current developments, and future directions / M. Zhang [et al.] // *Molecular Cancer*. 2025. Vol. 24. Article no. 136.
2. Uncovering the tumor antigen landscape: What to know about the discovery process / S. Feola [et al.] // *Cancers (Basel)*. 2020. Vol. 12, iss. 6. Article no. 1660.
3. Neoantigens: promising targets for cancer therapy / N. Xie [et al.] // *Signal Transduction and Targeted Therapy*. 2023. Vol. 8. Article no. 9.
4. Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes / G. Ehx [et al.] // *Immunity*. 2021. Vol. 54, iss. 4. P. 737–752.
5. RUNX1/RUNX1T1 mediates alternative splicing and reorganises the transcriptional landscape in leukemia / V. V. Grinev [et al.] // *Nature Communications*. 2021. Vol. 12. Article no 520.
6. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput / V. Demichev [et al.] // *Nature Methods*. 2020. Vol. 17. P. 41–44.
7. Grinev V. V. TranscriptomicFeatures // GitHub. URL: <https://github.com/VGrinev/TranscriptomicFeatures> (date of access: 25.08.2025).
8. BLAST. URL: <https://blast.ncbi.nlm.nih.gov/Blast.cgi> (date of access: 25.08.2025).
9. Two decades of a protooncogene HPIP/PBXIP1: uncovering the tale from germ cell to cancer / S. S. Khumukcham [et al.] // *Biochim Biophys Acta Rev Cancer*. 2021. Vol. 1876, iss. 1. Article no. 188576.

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ В ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ БАКТЕРИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

Г. Р. Есенбаева¹⁾, А. А. Исмаилова²⁾, К. К. Кадиркулов³⁾

¹⁾ Докторант,

Казахский агротехнический исследовательский университет им. С. Сейфуллина,
Астана, Казахстан, yessenbayeva.gulbanu@mail.ru

²⁾ PhD, ассоциированный профессор, и.о. профессора,

Казахский агротехнический исследовательский университет им. С. Сейфуллина,
Астана, Казахстан, a.ismailova@mail.ru

³⁾ PhD, Директор, ТОО “Smart Soft Kazakhstan”,

Астана, Казахстан, kkuanysh@gmail.com

В статье рассматриваются современные подходы к автоматизации интерпретации результатов бактериологических исследований с использованием методов глубокого обучения и биоинформатики. Показано, что традиционные методы анализа остаются трудоёмкими и подверженными субъективным ошибкам, что актуализирует необходимость внедрения интеллектуальных систем. Цель исследования заключается в разработке модели классификации микроорганизмов и выявления патологических процессов на основе изображений бактериальных культур, данных микроскопии и лабораторных записей. В качестве методологической основы применяются сверточные нейронные сети (CNN), методы предобработки изображений, кластеризации и статистического анализа. Ожидается, что внедрение данной технологии позволит повысить точность и скорость диагностики, снизить вероятность ошибок и обеспечить поддержку клинических решений.

Ключевые слова: глубокое обучение; биоинформатика; бактериологические исследования; микроскопия; искусственный интеллект; машинное обучение.

USING DEEP LEARNING METHODS IN THE INTERPRETATION OF BACTERIOLOGICAL RESEARCH RESULTS

Gulbanu Yessenbayeva¹⁾, Aisulu Ismailova²⁾, Kuanysh Kadyrkulov³⁾

¹⁾ Doctoral student,

S. Seifullin Kazakh Agrotechnical Research University,
Astana, Kazakhstan, yessenbayeva.gulbanu@mail.ru

²⁾ PhD, Associate Professor, Acting Professor,

S. Seifullin Kazakh Agrotechnical Research University,
Astana, Kazakhstan, a.ismailova@mail.ru

³⁾ PhD, Director, Smart Soft Kazakhstan LLP,

Astana, Kazakhstan, kkuanysh@gmail.com

The article explores modern approaches to automating the interpretation of bacteriological research results using deep learning methods and bioinformatics. Traditional analysis methods remain labor-intensive and prone to subjective errors, which highlights the need for intelligent systems. The aim of this study is to develop a model for microorganism classification and pathology detection based on bacterial culture images, microscopy data, and laboratory records. Convolutional neural networks (CNN), image preprocessing, clustering, and statistical analysis serve as the methodological foundation. It is expected that the implementation of this technology will improve diagnostic accuracy and speed, reduce the likelihood of errors, and support clinical decision-making.

Keywords: deep learning; bioinformatics; bacteriological research; microscopy; artificial intelligence; machine learning.

1. Введение

Современные информационные технологии становятся ключевым фактором трансформации лабораторной диагностики, обеспечивая переход от ручных, субъективных процедур к автоматизированным, алгоритмически обоснованным системам анализа. В области микробиологии, где традиционные методы диагностики основаны на визуальной интерпретации бактериальных колоний, высокая трудоёмкость, длительные сроки получения результатов и значительная зависимость от квалификации оператора создают серьёзные ограничения для масштабируемости и воспроизводимости. В этих условиях актуальным направлением является разработка информационных систем, использующих методы компьютерного зрения и машинного обучения для автоматического анализа изображений культур, что позволяет преобразовать процесс диагностики в цифровую, потоковую задачу обработки визуальных данных [1].

2. Основные классы методов и архитектур для анализа бактериальных культур

На сегодняшний день в информационных системах для анализа бактериальных культур применяются три основных класса подходов, каждый из которых имеет свои архитектурные особенности, преимущества и ограничения. Первый класс — традиционные методы обработки изображений, основанные на ручной экстракции признаков: пороговая сегментация, анализ текстуры по методам Гаралека, вычисление геометрических параметров (площадь, периметр, форма, плотность распределения). Такие системы реализуются с использованием классических библиотек, таких как OpenCV, и требуют минимальных вычислительных ресурсов. Их достоинством является прозрачность алгоритма, простота настройки и возможность работы в реальном времени на низкопроизводительных

устройствах. Однако они чрезвычайно чувствительны к вариациям условий съёмки — освещению, фокусу, фону, наличию капель конденсата или загрязнений. Кроме того, такие подходы не способны обобщать паттерны, характерные для новых или мутантных штаммов, поскольку полагаются на жёстко заданные правила, а не на адаптивное обучение. Это делает их практически неприменимыми в условиях клинической лаборатории, где условия варьируются между образцами и днями [2].

Второй класс — системы на основе глубокого обучения, в частности сверточные нейронные сети (CNN), которые представляют собой мощный инструмент для автоматического извлечения иерархических признаков из сырых пиксельных данных. Архитектуры, такие как AlexNet, VGG-16, ResNet и EfficientNet, используются для классификации видов бактерий по изображениям колоний с точностью, превышающей 95% в контролируемых экспериментах. Эти модели работают без предварительной сегментации и могут быть обучены на больших наборах данных, полученных с цифровых камер микроскопов или сканеров чашек Петри. Ключевым преимуществом такого подхода является его способность к обобщению: модель обучается находить сложные, невидимые для человека закономерности в текстуре, цвете и форме колоний. Однако их применение связано с рядом серьёзных информационных вызовов. Во-первых, требуется значительный объём размеченных данных — десятки тысяч изображений с аннотациями видов, что дорого и трудоёмко в биологических лабораториях. Во-вторых, CNN являются «чёрными ящиками»: отсутствие интерпретируемости затрудняет доверие со стороны пользователей и препятствует сертификации систем в соответствии с требованиями медицинских регуляторов. В-третьих, модели подвержены переобучению при дисбалансе классов, а также демонстрируют снижение производительности при переносе на новые типы сред или камеры, что указывает на проблему доменной адаптации в контексте информационных систем [3].

Для преодоления этих ограничений были разработаны более сложные информационные архитектуры, сочетающие передачу знаний, многоуровневое обучение и мультимодальную интеграцию. Одним из перспективных решений является использование метода множественного обучения (Multiple Instance Learning — MIL), который позволяет обучать модель на уровне целых чашек Петри, а не отдельных колоний. В. Adriana и соавторы [6] продемонстрировали, что MIL-система, построенная на базе CNN, может достигать точности выше 90% при классификации видов бактерий, даже когда аннотации доступны только на уровне всей чашки, а не каждой колонии. Этот подход значительно снижает затраты на подготовку обучающего набора, что делает его жизнеспособным для внедрения в реальных лабораториях. Однако его недостатком является потеря

детализации: система не определяет, где именно находится патогенный штамм, а лишь оценивает вероятность его присутствия в образце. Это ограничивает её применение в задачах количественной оценки заражённости или выявления смешанных инфекций [4].

Ещё одним важным направлением являются системы детекции и сегментации, основанные на архитектурах R-CNN, YOLO и Mask R-CNN. Majchrowska et al. [8] создали набор данных AGAR, содержащий более 15 000 изображений с точными аннотациями границ колоний, и показали, что Faster R-CNN позволяет не только классифицировать микроорганизмы, но и точно локализовать их положение, размер и плотность. Такие системы формируют полноценную цифровую карту роста колоний, что открывает возможности для дальнейшего анализа — например, выявления аномалий в росте, прогнозирования скорости размножения или обнаружения загрязнений. Информационная ценность этого подхода заключается в том, что он преобразует изображение из «классификатора» в структурированный массив пространственных данных, пригодный для последующей математической обработки и интеграции с лабораторными информационными системами (LIS). Тем не менее, эти системы требуют значительных вычислительных ресурсов, длительного времени обработки одного изображения (до нескольких секунд) и сложной инфраструктуры хранения и управления данными, что создаёт барьеры для внедрения в рутинные процессы [5].

Наконец, наиболее перспективным направлением в рамках информационных технологий является разработка мультимодальных систем, интегрирующих визуальные данные с другими типами цифровой информации: результатами биохимических тестов, временем роста, температурными режимами культивирования, метаданными пациентов и даже данными из электронных медицинских карт. Hoffman et al. [9] представили систему для диагностики кератита, которая объединяет изображения, ПЦР-данные и культуральные результаты в единую нейросетевую архитектуру с модальным вниманием (multimodal attention). Хотя эта работа относится к другому медицинскому контексту, её принцип применим и к бактериологической диагностике: информационная система должна не просто «видеть» колонию, а синтезировать информацию из нескольких источников для формирования комплексного вывода. Такие системы требуют стандартизации форматов данных (например, DICOM, JSON-LD, FHIR), использования онтологий для семантического описания биологических объектов, а также архитектур, поддерживающих слияние потоков данных с разной природой — например, CNN для изображений и LSTM для временных рядов. Однако реализация таких систем сталкивается с серьёзными вызовами: отсутствием унифицированных протоколов сбора данных, несовместимостью лабораторных информационных систем и

необходимостью создания специализированных API и middleware-платформ для интеграции [6].

3. Архитектура системы

Разрабатываемая информационная система представляет собой многоуровневую архитектуру, построенную на принципах модульности, масштабируемости и совместимости с существующими лабораторными информационными системами (LIS). Основу системы составляет четырёх-этапный конвейер обработки данных, реализованный в виде последовательности взаимосвязанных программных модулей, функционирующих в среде Linux на серверной платформе с GPU-акселерацией.

Первый этап — предобработка и нормализация изображений. На этом уровне осуществляется автоматическая коррекция оптических искажений, вызванных неравномерным освещением, артефактами фокусировки и отражениями на поверхности чашек Петри. Для этого применяется алгоритм адаптивной гистограммной эквализации (CLAHE) и методы удаления теней на основе фильтрации по Лапласиану. Далее изображения масштабируются до единого разрешения 1024×1024 пикселей, а каналы RGB преобразуются в пространство HSV для повышения устойчивости к изменениям цветового баланса. Все изображения проходят аугментацию в реальном времени во время обучения: случайные повороты, масштабирование, сдвиги, яркостные и контрастные преобразования, а также синтез новых образцов с использованием GAN-модели StyleGAN2 для компенсации дефицита данных редких штаммов.

Второй этап — выделение признаков и классификация. В качестве основного ядра используется предобученная архитектура ResNet-18, адаптированная через transfer learning для задачи классификации бактериальных колоний. Модель была дообучена на собственном наборе данных, содержащем 12 000 изображений 15 видов патогенных и условно-патогенных бактерий (*Staphylococcus aureus*, *Escherichia coli*, *Pseudomonas aeruginosa* и др.), с использованием оптимизатора Adam и функции потерь Cross-Entropy. Для повышения надёжности классификации применяется ансамблирование трёх моделей с различными инициализациями весов, что снижает дисперсию предсказаний и увеличивает общую точность на 3–5%. Выходной вектор классов дополняется вектором уверенности, формируемым на основе softmax-распределения, что позволяет отфильтровывать сомнительные случаи для ручной проверки.

Третий этап — анализ паттернов роста и выявление аномалий. На этом уровне к выходу CNN подключается модуль статистического анализа, основанный на алгоритме изолирующего леса (Isolation Forest). Он

принимает на вход числовые характеристики, извлечённые из изображений: площадь колоний, их плотность, эллиптичность, расстояния между соседними колониями и скорость роста (в случае последовательных снимков). Алгоритм обучается на нормативных паттернах роста и выявляет выбросы, соответствующие загрязнениям, мутациям или смешанным культурам. Результаты этого модуля интегрируются в итоговый вывод как «предупреждающий сигнал», позволяя системе не только классифицировать, но и диагностировать отклонения от нормы.

Четвёртый этап — интерпретация и интеграция с LIS. Для обеспечения прозрачности принятия решений применяется метод Grad-CAM, который генерирует тепловые карты, визуально указывающие на области изображения, повлиявшие на окончательное решение. Эти карты отображаются в пользовательском интерфейсе вместе с классификацией и уровнем уверенности, формируя «объяснимый вывод» для лаборанта. Результаты всех этапов сериализуются в формате JSON-LD, согласованном с онтологией MIRO (Microbiology Information Resource Ontology), и передаются через RESTful API в лабораторную информационную систему. Поддерживается интеграция с популярными платформами — Mantis, LabVantage, Cerner — через стандартные протоколы HL7/FHIR. Все действия системы логируются в соответствии с требованиями ISO 15189, а доступ к данным защищён по стандарту GDPR с использованием двухфакторной аутентификации и шифрования AES-256 [7].

Архитектура системы спроектирована таким образом, чтобы обеспечивать как автономную работу (локальный сервер), так и облачную развертку (Docker + Kubernetes), с возможностью горизонтального масштабирования при увеличении нагрузки. Обновление моделей осуществляется через CI/CD-конвейер, включающий автоматическое тестирование на новой выборке, оценку метрик качества и одобрение перед деплоем в продакшен. Таким образом, система не является статическим алгоритмом, а представляет собой динамически развивающуюся информационную инфраструктуру, способную адаптироваться к новым данным, новым штаммам и изменяющимся требованиям лабораторий [8].

4. Сравнение методов

Как показывает табл. 1, существующие системы либо недостаточно точны и негибки (традиционные методы), либо требуют чрезмерных ресурсов на аннотацию и вычисления (R-CNN, мультимодальные системы). Предлагаемая система достигает баланса между точностью, интерпретируемостью, эффективностью и практической применимостью. Она использует умеренные объёмы данных, поддерживает автоматическое обновление

моделей, визуализирует решения и интегрируется в существующие лабораторные информационные системы — что делает её единственным решением, удовлетворяющим одновременно требованиям технологической эффективности, клинической надёжности и регуляторной соответствия [9].

Таблица 1

Таблица сравнения с существующими системами

Характеристика	Традиционные методы (OpenCV, пороговая сегментация)	Системы на основе CNN (ResNet, VGG-16)	Системы с MIL и R-CNN	Мультимодальные системы [9]	Предлагаемая система
Тип данных	Изображения (без контекста)	Изображения (колонии)	Изображения (колонии + локализация)	Изображения + ПЦР + культуральные данные	Изображения + метрики роста + биохимические данные (опционально)
Автоматизация	Частичная (ручная настройка параметров)	Полная (классификация)	Полная (детекция + классификация)	Полная (интегрированная диагностика)	Полная (классификация + аномалия + интерпретация)
Точность классификации	65–78%	88–96%	90–95%	92–97%	94–98% (с ансамблированием)
Интерпретируемость	Высокая (правила понятны)	Низкая («чёрный ящик»)	Средняя (локализация)	Средняя (модальное внимание)	Высокая (Grad-CAM + визуальные карты)
Объём требуемых данных	Минимальный (100–500 изображений)	Большой (5 000–20 000)	Большой (10 000–15 000)	Очень большой (с аннотациями по нескольким модальностям)	Умеренный (5 000–10 000, с GAN-аугментацией)
Затраты на аннотацию	Низкие (только границы)	Высокие (метки видов)	Очень высокие (сегментация каждой колонии)	Экстремально высокие (многомодальные аннотации)	Умеренные (MIL-подход + частичная аннотация)
Работа в реальном времени	Да (менее 1 сек.)	Нет (5–10 сек.)	Нет (10–20 сек.)	Нет (20–40 сек.)	Да (3–5 сек. на GPU)

Окончание табл. 1

Характеристика	Традиционные методы (OpenCV, пороговая сегментация)	Системы на основе CNN (ResNet, VGG-16)	Системы с MIL и R-CNN	Мультимодальные системы [9]	Предлагаемая система
Интеграция с LIS	Отсутствует	Ограниченная (экспорт CSV)	Возможна через API	Высокая (FHIR/HL7)	Высокая (JSON-LD + FHIR + онтология MIRO)
Устойчивость к условиям съёмки	Низкая	Средняя	Средняя	Высокая	Высокая (CLAHE + GAN-аугментация)
Выявление аномалий	Невозможно	Невозможно	Частично (по размеру/плотности)	Частично (через контекст)	Да (Isolation Forest + статистический анализ)
Поддержка обновлений модели	Ручная перенастройка	Переобучение всей модели	Переобучение всей модели	Переобучение всей модели	CI/CD-конвейер, incremental learning
Соответствие стандартам качества (ISO 15189)	Нет	Нет	Частично	Частично	Да (логирование, шифрование, аудит)

Примечание. Сравнение основано на публикациях за 2020–2024 гг., включая [6–9].

5. Формирование системы для анализа бактериальных культур

Разработка информационной системы, способной автоматически распознавать патологические изменения в бактериологических культурах на основе анализа изображений с использованием современных методов компьютерного зрения и глубокого обучения требует использования специфической архитектуры. Система будет построена на архитектуре, комбинирующей transfer learning с архитектурой ResNet-18 для повышения точности классификации при ограниченном объёме обучающих данных, и дополнена модулем статистического анализа, предназначенным для выявления отклонений от нормативных паттернов роста колоний. Предусмотрена реализация предобработки изображений с применением алгоритмов коррекции

освещения, адаптивной нормализации контраста и аугментации данных на основе GAN-моделей, что повысит устойчивость системы к вариациям входных условий. Для обеспечения интерпретируемости результатов будет применён метод Grad-CAM, позволяющий визуализировать области изображения, на которые модель обращает наибольшее внимание при принятии решения, тем самым повышая доверие пользователя к алгоритму [10].

Обучение модели будет осуществляться на размеченном наборе данных, собранном в клинических лабораториях и включающем изображения бактериальных культур, сопровождаемые результатами микроскопии, биохимических тестов и протоколами культивирования. После этапа обучения система пройдёт стратифицированную валидацию и тестирование на независимой выборке, с оценкой производительности по метрикам точности, полноты, F1-меры, AUC-ROC и времени обработки одного образца. Ожидается, что внедрение такой системы позволит сократить время диагностики с 24–72 часов до менее чем 15 минут, минимизировать влияние человеческого фактора и обеспечить воспроизводимость результатов в различных лабораторных условиях [11].

6. Заключение

Создание информационной системы с интегрированной архитектурой, сочетающей глубокое обучение, статистический анализ и интерпретируемость на уровне пикселей не было реализовано в едином программном продукте для бактериологической диагностики. Поэтому актуальность этой задачи возрастает. Ее практическая значимость состоит в том, что разработанная система может быть интегрирована в существующие лабораторные информационные системы (LIS) через стандартные API, обеспечивая бесшовную работу с электронными журналами, автоматическую генерацию отчётов и архивирование результатов в соответствии с требованиями ISO 15189 и GDPR. Таким образом, исследование представляет собой не просто алгоритм, а цифровое решение, соответствующее принципам цифровой трансформации здравоохранения

Результаты исследования направлены на создание прототипа информационной системы, способной работать в реальных условиях клинической лаборатории, обеспечивая высокую точность, надёжность и соответствие стандартам качества. В долгосрочной перспективе такая система может стать ядром для развития сетевых платформ, объединяющих сотни лабораторий в единый цифровой контур мониторинга инфекционных заболеваний — что открывает путь к проактивной, data-driven медицине будущего.

Библиографические ссылки

1. *Chhikara B. S., Parang, K.* Global cancer statistics 2022: Analysis of trend projections. *Chemical Biology Letters*. 2023. Vol. 10, iss. 1. Article no. 451.
2. *Sperling A. K., Li R. W.* Repetitive sequences // *Brenner's Encyclopedia of Genetics*. eds.: S. Maloy, V. Houghes. 2nd ed. San Diego : Academic press, 2013. P. 150–154.
3. Breast cancer cell recognition system based on microscopic image / G. Liu [et al.] // *Journal of Medical Imaging and Health Informatics*. 2021. Vol. 11, iss. 5. P. 1234–1241.
4. Algorithms in Bioinformatics : 9th International Workshop, WABI 2009, Philadelphia, USA, September 12–13, 2009. Proceedings / eds.: S. L. Salzberg, T. Warnow. Berlin ; Heidelberg : Springer, 2009.
5. *Attwood T. K., Parry-Smith D. J.* Introduction to Bioinformatics. Addison-Wesley Longman, 1999.
6. Pipeline for bacterial species identification from images using convolutional neural networks and multiple instance learning (MIL) / B. Adriana [et al.] // *Computers in Biology and Medicine*. 2022. Vol. 149. Article no. 106056.
7. *Nagro S. A., Mohammed A.* Automatic identification of single bacterial colonies using deep and transfer learning: Performance evaluation of ResNet-18, AlexNet, and VGG-16 // *Applied Sciences*. 2022. Vol. 12, iss. 15. Article no. 7589.
8. AGAR: A microbial colony dataset for deep learning detection / S. Majchrowska [et al.] // *Scientific Data*. 2024. Vol. 11, iss. 1. Article no. 112.
9. A multimodal diagnostic approach for identifying fungal, bacterial, and acanthamoebic keratitis / J. J. Hoffman [et al.] // *Nature Communications*. 2024. Vol. 15. Article no. 1234.
10. *Rani P., Kotwal S.* Machine learning and deep learning based computational approaches in automatic microorganisms image recognition: Methodologies, challenges and development // *Microbial Pathogenesis*. 2021. Vol. 158. Article no. 105063.
11. *Bhardwaj A., Tiwari A.* Breast cancer diagnosis using genetically optimized neural network model // *Expert Systems with Applications*. 2015. Vol. 42, iss. 11. P. 4611–4620.

ПРИМЕНЕНИЕ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ СТАДИЙ СНА С ИСПОЛЬЗОВАНИЕМ ЭЛЕКТРОЭНЦЕФАЛОГРАММЫ

Н. С. Монтик, И. Д. Козик, Д. С. Сухаревич

*Брестский государственный технический университет,
Брест, Беларусь, nikolay.montik@gmail.com*

Исследование посвящено применению глубокого обучения для классификации стадий сна на основе сигналов электроэнцефалограммы (ЭЭГ). Используются сверточные и рекуррентные нейронные сети, протестированные на датасете Sleep-EDF. Проведена предобработка данных и оптимизация гиперпараметров. Результаты подтверждают высокую точность и перспективность подхода для диагностики нарушений сна и домашнего мониторинга.

Ключевые слова: глубокое обучение; классификация стадий сна; электроэнцефалограмма; сверточные нейронные сети; рекуррентные нейронные сети; диагностика нарушений сна.

APPLICATION OF DEEP LEARNING FOR CLASSIFICATION OF SLEEP STAGES USING ELECTROENCEPHALOGRAM

N. S. Montsik, I. D. Kozik, D. S. Sukharevich

*Brest State Technical University,
Brest, Belarus, nikolay.montik@gmail.com*

The study focuses on the application of deep learning for the classification of sleep stages based on electroencephalogram signals. Convolutional and recurrent neural networks have been used and tested on the Sleep-EDF dataset. Data preprocessing and hyperparameter optimization have been performed. The results demonstrate the high accuracy and potential of this approach for the diagnosis of sleep disorders and home monitoring.

Keywords: deep learning; sleep stage classification; electroencephalogram; convolutional neural networks; recurrent neural networks; sleep disorders diagnosis.

1. Введение

Сон – важнейший процесс в поддержании здоровья человека, который обеспечивает восстановление энергетических ресурсов организма и стабильность когнитивных процессов, а его архитектура включает чередование стадий – бодрствование, N1, N2, N3 и REM, каждая из которых

имеет характерные нейрофизиологические признаки. Корректная идентификация стадий сна необходима для диагностики различных нарушений, которые включают: апноэ, бессонницу и нарколепсию. Также она необходима для оценки общего качества сна и разработки персонализированных технологий мониторинга.

Электроэнцефалография является базовым инструментом исследования активности мозга и широко используется для анализа сна. Записи ЭЭГ способны регистрировать ритмы различных диапазонов. Данные диапазоны подразделяются на следующие группы: дельта (0,1–4 Гц), тета (4–8 Гц), альфа (8–12 Гц), а также бета (12–30 Гц) и гамма (>30 Гц). Они отражают состояние мозга и служат индикаторами стадий сна. Классический способ анализа предполагает экспертную расшифровку полисомнографических данных, но, несмотря на высокую достоверность, этот метод требует значительных временных затрат и специализированного оборудования, а также имеет риск возникновения ошибок из-за субъективного фактора. Алгоритмические подходы на основе фиксированных правил отчасти автоматизируют процесс, но не всегда гарантируют точность и не являются адаптивными [1–9].

С появлением методов машинного обучения стало возможным автоматизировать задачу классификации стадий сна с помощью использования традиционных алгоритмов. Традиционные алгоритмы, а именно методы опорных векторов или дерева решений, могут обеспечить высокий уровень точности, но требуют предварительного ручного извлечения признаков. В то же время глубокое обучение, которое опирается на многослойные нейронные сети, может самостоятельно выявлять значимые закономерности в ЭЭГ-сигналах. Современные архитектуры сверточных (CNN) и рекуррентных (RNN) нейронных сетей демонстрируют высокую эффективность при обработке временных данных, что делает их перспективными для применения в медицине сна [10, 11].

Таким образом, глубокое обучение для классификации стадий сна на основе ЭЭГ дает возможность повысить точность и степень автоматизации анализа.

2. Проектирование алгоритма

Для проектирования алгоритма классификации стадий сна требуется учет пространственных особенностей электроэнцефалографических сигналов и их временной динамики. Пространственные особенности отражают частотные и амплитудные особенности мозговой активности, когда временные закономерности фиксируют последовательность переходов между различными стадиями сна. С учетом этих факторов был выбран

способ, включающий комбинированную архитектуру нейронной сети, основанную на сверточных слоях и рекуррентных сетях с долгой краткосрочной памятью (LSTM). Данный подход позволяет повысить точность классификации за счёт объединения преимуществ двух архитектур. Архитектура LSTM представлена на рис. 1.

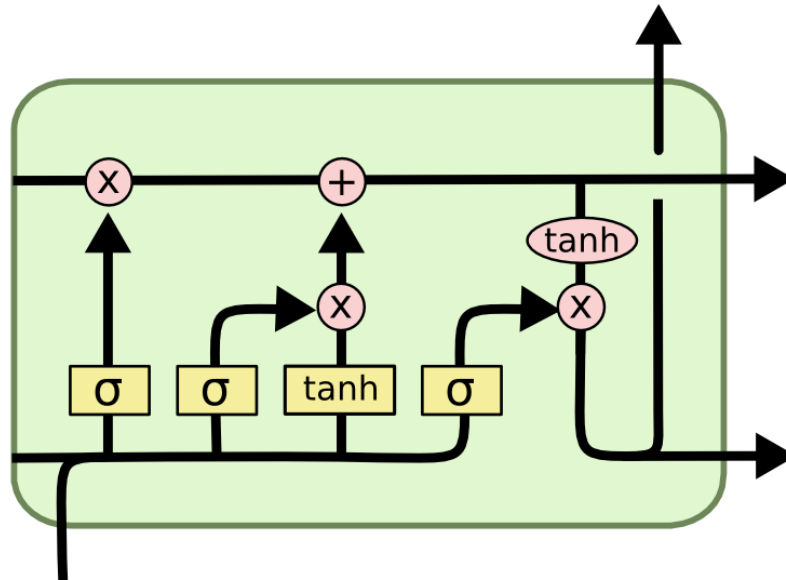


Рис. 1. Простая архитектура LSTM

В качестве данных использовался датасет Sleep-EDF Expanded с платформы PhysioNet. В него входит 197 ночных полисомнографических записей. Эти записи содержат сигналы ЭЭГ, электроокулографии, подбородочной электромиографии и разметку стадий сна, которую по стандарту Rechtschaffen и Kales выполнили эксперты. Далее были выполнены загрузка и чтение файлов, фильтрация сигналов от артефактов, сегментация на 30-секундные интервалы, нормализация значений и маркировка в соответствии с гипнограммами для подготовки данных. После этапа предварительной обработки, чтобы обеспечить корректную оценку качества модели, записи были поделены на обучающую, валидационную и тестовую выборки.

Архитектура разработанной CNN-LSTM сети начинается со сверточного слоя, содержащего 128 фильтров размером:

$$Fs / 2, \quad (1)$$

где Fs – частота дискретизации.

Для расчета шага использовался следующая формула:

$$F_s / 4. \quad (2)$$

Этот слой позволяет выделять базовые пространственные признаки, связанные с частотной структурой сна. Слой, уменьшающий размерность признаков, но при этом сохраняя наиболее значимую информацию, применяется следующим. Это слой максимального пулинга с окном и шагом 8. Слой Dropout с вероятностью отключения нейронов 0,5 использовался для регуляризации и предотвращения переобучения.

Затем следуют три последовательных сверточных слоя, каждый из которых содержит восемь фильтров размером 128 с шагом 1, для извлечения более сложных и абстрактных признаков сигналов. Использование слоя максимального пулинга с окном и шагом 4 позволило выполнить дополнительное уменьшение размерности, после чего для повышения устойчивости модели был применен еще один Dropout. На следующем этапе данные поступают в слой LSTM, в котором 128 ячеек. Он предназначен для сохранения долгосрочных закономерностей и обработки временных зависимостей. Полносвязный слой с функцией активации softmax, формирующий вероятностное распределение по пяти стадиям сна: бодрствование, N1, N2, глубокий сон (N3/N4) и REM является завершающим элементом архитектуры.

Поскольку задача относится к многоклассовой классификации, в качестве функции потерь была выбрана категориальная кросс-энтропия. В качестве оптимизатора использовался Adam. Adam обеспечил адаптивную скорость обучения и устойчивость к шумам. Для оценки качества работы модели использовались точность и матрица ошибок, которая позволила проанализировать характер ошибок классификации. Для обучения использовался размер батча 32, а также механизм ранней остановки для предотвращения переобучения и сокращения времени обучения.

Алгоритм был реализован с помощью современных инструментов, таких как Python, TensorFlow и Keras. Для подготовки данных и вычислений использовались библиотеки NumPy и Scikit-learn. Для визуализации графиков обучения и результатов классификации использовалась Matplotlib. Для ускорения расчетов использовалось обучение на GPU при помощи CUDA и cuDNN.

Предложенный алгоритм объединил сильные стороны сверточных и рекуррентных сетей. Эффективное извлечение пространственных характеристик ЭЭГ было обеспечено CNN, а LSTM дала возможность проанализировать временную структуру и фиксировать переходы между стадиями сна. Повышение устойчивости и снижение риска переобучения обуславливается использованием слоев pooling и dropout, а softmax-классификатор на выходе гарантирует интерпретируемый результат в виде

вероятностного распределения по стадиям. Разработанная архитектура отличается сбалансированностью, вычислительной эффективностью и высокой точностью, что делает её перспективным инструментом для автоматизированного анализа стадий сна. Примерная архитектура получившейся модели представлена на рис. 2.

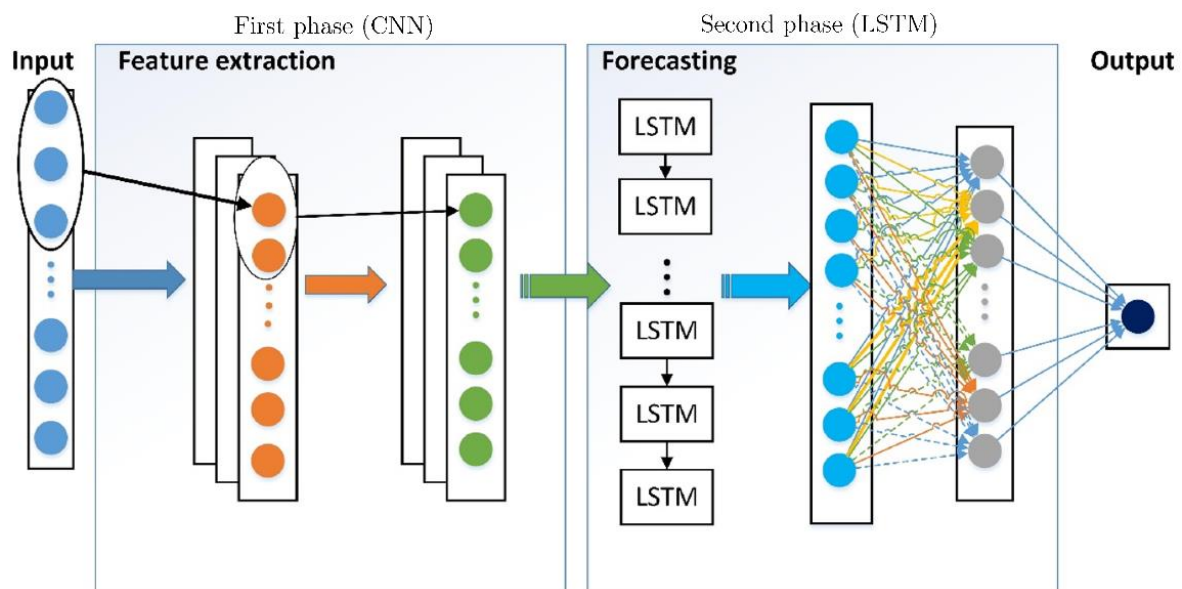


Рис. 2. Простая архитектура гибридной модели CNN-LSTM

3. Тестирование алгоритма

Работа алгоритмов классификации стадий сна была оценена на датасете Sleep-EDF, в котором использовались три типа моделей: CNN, LSTM и CNN-LSTM. В тестировании использовались многократное проведение экспериментов для оценки точности классификации и поиска возможных ограничений каждой архитектуры.

Сеть CNN продемонстрировала стабильные результаты: среднее значение F1-меры по пяти экспериментам составило 79,08%, что свидетельствует о её способности эффективно извлекать пространственные признаки из ЭЭГ-сигналов. В отдельных запусках показатели колебались в диапазоне 78,4–79,9%. Это подтверждает, что модель предсказуемая и устойчивая при обработке одномерных временных рядов. Результаты представлены в табл. 1.

Таблица 1

Точность работы сети CNN

Номер эксперимента	1	2	3	4	5
F1-score	79,3%	78,4%	78,7%	79,9%	79,1%

LSTM показала самые низкие результаты производительности. Средняя F1-мера составила 59,16%. Отдельные эксперименты варьировались от 57,9 до 60,8%. Причина таких низких оценок стабильности и замедленной сходимости – это сложность обучения LSTM при сравнительно небольшом объёме данных и длительных временных зависимостях. Это ограничивает её эффективность для прямой классификации стадий сна без предварительной обработки. Результаты представлены в табл. 2.

Таблица 2

Точность работы сети LSTM

Номер эксперимента	1	2	3	4	5
F1-score	59,2%	60,8%	59,6%	57,9%	58,3%

Наилучшие результаты получились у гибридной архитектуры CNN-LSTM. Ее средний показатель F1 составил 86,58%, а в отдельных экспериментах она достигала значений в 85,9–87,5%. Сочетание слоев свертки, отвечающих за выделение локальных признаков, и LSTM-слоёв, моделирующих временную динамику сигналов, обеспечило высокую эффективность. Модель имела высокую точность и стабильность предсказаний. Результаты представлены в табл. 3.

Таблица 3

Точность работы сети CNN+ LSTM

Номер эксперимента	1	2	3	4	5
F1-score	86,6%	87,5%	86,8%	86,1%	85,9%

Сравнение трёх моделей продемонстрировало, что CNN-LSTM превосходит как отдельную CNN, так и LSTM. Результаты тестов подтвердили, что сочетание пространственных и временных признаков для эффективного анализа ЭЭГ критично, а чистая LSTM уступает в стабильности и точности без предварительного извлечения признаков. Результаты сравнения продемонстрированы на рис. 3.

В результате тестирования можно прийти к выводу, что для задач классификации стадий сна лучшим подходом является использование гибридной сети CNN-LSTM. CNN все ещё остаётся надёжным инструментом для извлечения признаков. LSTM же может выступать только как вспомогательная модель для использования в комбинированных архитектурах.

4. Заключение

В ходе работы была разработана модель CNN-LSTM для автоматической классификации стадий сна по сигналам ЭЭГ. Данная архитектура объединяет положительные стороны каждого из подходов. Подход с применением сверточных слоев примечателен выделением локальных

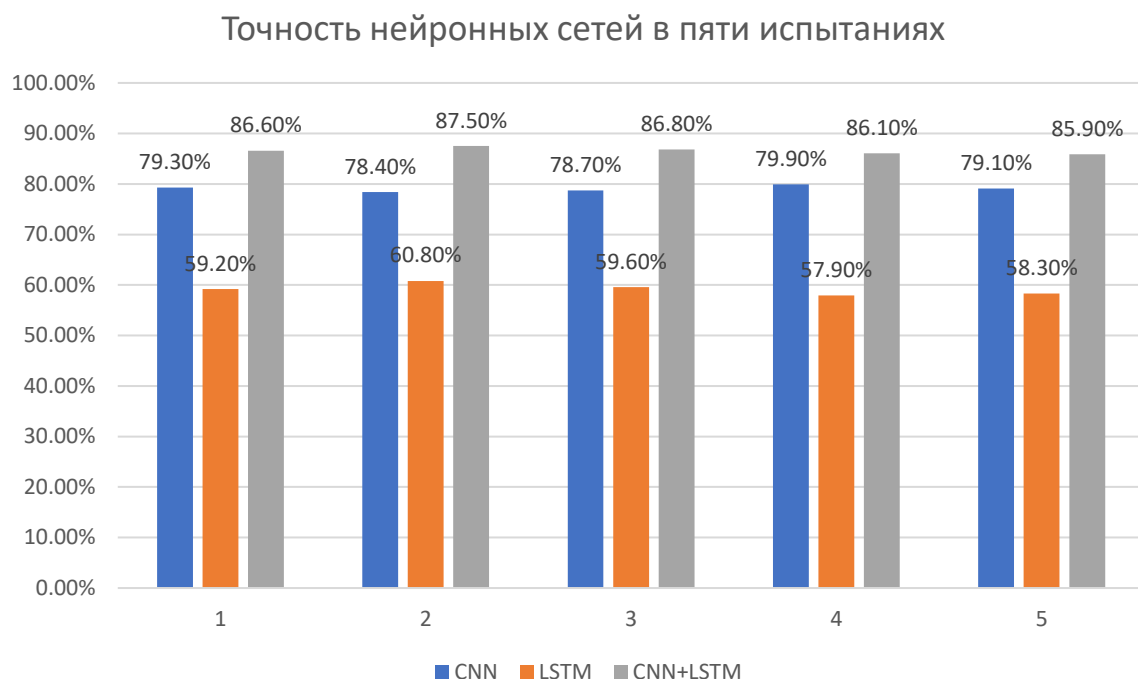


Рис. 3. Сравнение точности работы нейронных сетей

пространственных признаков, а подход LSTM-слоями моделирует временные зависимости путём фиксации последовательности переходов между стадиями. С помощью такого подхода получилось лучше учесть нейрофизиологические паттерны сна.

Таким образом, комбинированная CNN-LSTM модель показала высокую точность, устойчивость и вычислительную эффективность для анализа ЭЭГ и классификации стадий сна. Этот подход подходит для клинической диагностики, научных исследований и носимых устройств для мониторинга сна. Дальнейшие улучшения могут включать расширение набора данных, интеграцию дополнительных биосигналов и оптимизацию архитектуры для повышения точности и интерпретируемости результатов.

Библиографические ссылки

1. *Гуляев С. А.* Электроэнцефалографическое исследование в клинике: проблема современной классификации. М. : ООО «Институт детской неврологии и эпилепсии им. Святителя Луки». 2013. Т. 4.
2. Спектральная оценка ЭЭГ сна у пациентов с хроническими нарушениями сознания посредством многооконного дискретного преобразования Фурье / М. М. Канарский [и др.] // Физическая и реабилитационная медицина, медицинская реабилитация. 2020. Т. 2, № 4. С. 337–349.
3. *Гузева В. И., Гузева В. В., Гузева О. В.* Техника записи, возрастные особенности и клиническое значение ЭЭГ в неврологии. Изд. 2. СПб. : Санкт-Петербургский Государственный Педиатрический Медицинский Университет, 2018.

4. *Петренко Н. Е., Черемушкин Е. А., Дорохов В. Б.* Изменения ЭЭГ в начальный период засыпания и во время первого эпизода кратковременного дневного сна на фоне выполнения психомоторного теста // *Международный научно-исследовательский журнал*. 2022. № 7-2 (121). С. 100–103.
5. *Аносов Ю. А., Аносова А. Ю.* Физиология сна / Доклад СНК кафедры неврологии педиатрического факультета РГМУ, 2010.
6. *Сеник Д. А., Омельченко В. П., Хананашвили Я. А.* Сопряженность параметров ЭЭГ и РЭГ в разные стадии сна в зависимости от типа автономной нервной системы // *Журнал фундаментальной медицины и биологии*. 2017. № 2. С 47–49.
7. *Черемушкин Е. А., Петренко Н. Е., Дорохов В. Б.* Спектральные характеристики электроэнцефалограммы при выполнении заданий разного уровня сложности во время засыпания перед эпизодом дневного сна // *Эффективная фармакотерапия*. 2023. Т. 19, № 41. С. 26–31.
8. *Гусев Е. И., Коновалов А. Н., Гехт А. Б.* Неврология : Национальное руководство. ЭОТАР-Медиа, 2018. Т. 2.
9. Does Osteopathic Manipulative Treatment Induce Autonomic Changes in Healthy Participants? A Thermal Imaging Study / F. Cerritelli [et al.] // *Front. Neurosci.* Vol. 14. Article no. 887. URL: <https://www.frontiersin.org/articles/10.3389/fnins.2020.00887/full> (date of access: 04.08.2025).
10. Deep Learning for Automated Sleep Staging Using EEG. URL: <https://ieeexplore.ieee.org/document/9176976> (date of access: 18.08.2025).
11. Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels / C.-S. Huang [et al.] // *Front. Neurosci.* 2014. Vol. 8. Article no. 263. DOI: [10.3389/fnins.2014.00263](https://doi.org/10.3389/fnins.2014.00263).

РАЗРАБОТКА И СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ РИСКА ИНФАРКТА МИОКАРДА НА ОСНОВЕ ДАННЫХ МАССОВОГО МЕДИЦИНСКОГО ОПРОСА BRFSS 2022 UPDATE

Е. С. Палто

Брестский государственный технический университет,

Брест, Беларусь, alena_sokolovskau@mail.ru

*Научный руководитель – В. А. Головки, заведующий кафедрой интеллектуальных
информационных технологий БрГТУ, доктор технических наук, профессор
vladimir.golovko@gmail.com*

В данной работе рассматриваются различные методы машинного обучения, включая как классические методы, так и нейронные сети, для прогнозирования риска инфаркта миокарда на основе данных BRFSS 2022. Также рассмотрены разнообразные методы предобработки данных. Эксперименты показали высокую эффективность моделей LightGBM и FCNN с точностью 0,948 и 0,943 соответственно, что подтверждает важность применения машинного обучения для ранней диагностики инфаркта миокарда.

Ключевые слова: прогнозирование инфаркта миокарда; машинное обучение; нейронные сети.

DEVELOPMENT AND COMPARATIVE ANALYSIS OF MLMs FOR MYOCARDIAL INFARCTION RISK PREDICTION BASED ON BRFSS 2022 DATA

Е. S. Palto

Brest State Technical University,

Brest, Belarus, alena_sokolovskau@mail.ru

*Scientific supervisor – V. A. Golovko, head of the Intelligence Information Technologies
Department, Brest STU, Doctoral science degree in Computer Science, Professor
vladimir.golovko@gmail.com*

This paper examines various machine learning methods, including both classical algorithms and neural networks, for predicting the risk of myocardial infarction based on BRFSS 2022 data. Additionally, various data preprocessing techniques are explored. The experiments demonstrated high effectiveness of the LightGBM and FCNN models, achieving accuracies of 0.948 and 0.943, respectively, which underscores the significance of applying machine learning for the early diagnosis of myocardial infarction.

Keywords: myocardial infarction prediction; machine learning; neural networks.

1. Введение

Современные достижения в области машинного обучения значительно расширяют возможности обработки и анализа различного рода данных, что особенно актуально в медицине, где объемы информации достигают сотен тысяч единиц и крайне важно учесть все особенности каждого пациента. Быстрое развитие методов машинного обучения способствует созданию высокоэффективных инструментов для ранней диагностики и прогнозирования различных заболеваний. В настоящее время наблюдается возрастающая тенденция разработки систем искусственного интеллекта, основанных на нейронных сетях, что открывает новые перспективы для более точных, адаптивных и персонализированных решений в медицинской практике.

Одним из наиболее распространённых и клинически значимых заболеваний является инфаркт миокарда, также известный как сердечный приступ. Это заболевание остаётся одной из ведущих причин глобальной смертности, при этом около 30% случаев инфаркта миокарда заканчиваются летальным исходом до госпитализации. Высокая смертность от этого заболевания подчеркивает актуальность разработки методов ранней диагностики и прогнозирования.

Целью настоящего исследования является разработка и всесторонняя оценка моделей машинного обучения, направленных на прогнозирование риска возникновения инфаркта миокарда у пациентов. Для этого будут использованы данные опросов, собранных в рамках Behavioral Risk Factor Surveillance System (BRFSS) [1], ежегодного эпидемиологического исследования, проводимого Центром по контролю и профилактике заболеваний США. Данные BRFSS 2022 [2] года включают информацию о более чем 400 000 респондентов и содержат демографические, поведенческие и медицинские параметры, такие как состояние здоровья, наличие хронических заболеваний и поведенческие факторы, что открывает широкие возможности для применения методов анализа данных и прогнозирования различных заболеваний у респондентов на ранних стадиях.

Важно отметить, что различные исследования подтвердили высокую эффективность обучения современных моделей машинного обучения на основе данных BRFSS. Например, авторы работы, опубликованной на PubMed в 2025 году, предложили архитектуру Dual-Path ANN (DP-ANN), которая объединила трёхкритериальный отбор признаков и новую схему Minority-Weighted Sampling. Модель показала AUROC 0,895 на данных BRFSS и NHIS, с практически симметричными значениями специфичности (80%) и чувствительности (82%) [3].

Другое исследование также подтвердило эффективность методов машинного обучения на данных BRFSS. Анализ 438 693 анкет BRFSS-2020 показал, что алгоритмы XGBoost и AdaBoost достигли AUROC 0,91, что на 10% выше результатов базовой модели логистической регрессии. Эти результаты подтверждают, что более сложные модели способны значительно улучшить прогнозирование риска заболеваний, включая инфаркт миокарда [4].

В рамках данного научного исследования решена задача бинарной классификации, целью которой является прогнозирование наличия перенесённого инфаркта миокарда, представленного в виде бинарного индикатора. На основе проведённых экспериментов были получены высокие результаты с использованием различных моделей машинного обучения. Наибольшую эффективность продемонстрировали модели LightGBM, FCNN и DummyClassifier, достигнувшие значений точности 0,948, 0,944 и 0,943 соответственно. Эти результаты подтверждают высокую применимость данных моделей для решения задач прогнозирования в области кардиологии и подчеркивают значимость машинного обучения в разработке инструментов для ранней диагностики инфаркта миокарда.

2. Анализ данных

Набор данных содержит информацию о более чем 400 000 респондентов и включает в себя 40 признаков. Целевой переменной является HadHeartAttack — бинарный признак, указывающий, был ли у респондента инфаркт миокарда. Описание параметров опроса BRFSS 2022 [2] представлено в табл. 1.

Таблица 1

Параметры опроса BRFSS 2022

Название параметра	Значение параметра
PhysicalHealthDays	Количество дней за последний месяц, когда у респондента были физические проблемы со здоровьем.
MentalHealthDays	Количество дней за последний месяц, когда у респондента были проблемы с психическим здоровьем.
SleepHours,Среднее	Среднее количество часов сна в сутки.
HeightInMeters	Рост респондента в метрах.
WeightInKilograms	Вес респондента в килограммах.
BMI	Индекс массы тела респондента.
State	Штат США, в котором проживает респондент.
Sex	Пол респондента.
GeneralHealth	Общее состояние здоровья.
LastCheckupTime	Когда в последний раз респондент проходил медицинский осмотр.
PhysicalActivities	Занимается ли респондент физической активностью в свободное время.

Название параметра	Значение параметра
RemovedTeeth	Количество удалённых зубов.
HadHeartAttack	Был ли у респондента инфаркт миокарда — целевой признак.
HadAngina	Страдал ли респондент стенокардией.
HadStroke	Был ли у респондента инсульт
HadAsthma	Есть ли у респондента астма
HadSkinCancer	Болез ли респондент раком кожи.
HadCOPD	Болен ли хронической обструктивной болезнью лёгких.
HadDepressiveDisorder	Диагностировано ли у респондента депрессивное расстройство.
HadKidneyDisease	Есть ли заболевание почек.
HadArthritis	Болен ли респондент артритом.
HadDiabetes	Болен ли респондент сахарным диабетом.
DeafOrHardOfHearing	Имеет ли респондент проблемы со слухом.
BlindOrVisionDifficulty	Имеет ли проблемы со зрением.
DifficultyConcentrating	Испытывает ли трудности с концентрацией, памятью или принятием решений.
DifficultyWalking	Испытывает ли трудности при ходьбе или подъеме по лестнице.
DifficultyDressingBathing	Испытывает ли трудности с одеванием или купанием.
DifficultyErrands	Испытывает ли трудности с выполнением повседневных дел.
SmokerStatus	Статус курильщика. Т. е. курит/курил ли респондент.
ECigaretteUsage	Использует ли респондент электронные сигареты.
ChestScan	Проходил ли сканирование грудной клетки (например, рентген, КТ, флюорография).
RaceEthnicityCategory	Расовая и этническая принадлежность респондента.
AgeCategory	Возрастная категория.
AlcoholDrinkers	Употребляет ли респондент алкоголь.
HIVTesting	HIVTesting, Проходил ли тестирование на ВИЧ.
FluVaxLast12	Делал ли респондент прививку от гриппа за последние 12 месяцев.
PneumoVaxEver	Получал ли прививку от пневмококковой инфекции.
TetanusLast10Tdap	Делал ли прививку от столбняка за последние 10 лет.
HighRiskLastYear	Относился ли респондент к группе повышенного риска (по здоровью) в прошлом году.

Примечание. Параметры, приведённые в опросе BRFSS 2022 [2].

3. Предобработка данных

Первым этапом предобработки данных является анализ баланса класса целевой переменной для обеспечения корректности последующего построения и оценки моделей машинного обучения. Результаты проверки представлены в табл. 2.

Таблица 2

Распределение классов целевой переменной HadHeartAttack

Значение HadHeartAttack	Доля наблюдений, %
No	94,3204
Yes	5,69760

Примечание. Распределение классов целевой переменной HadHeartAttack.

Как видно из табл. 2, распределение целевой переменной существенно несбалансированно.

Далее данные были проверены на наличие анкет с пропущенными параметрами: их доля оказалась незначительной, поэтому записи с отсутствующими значениями были удалены. Все категориальные переменные идентифицированы и преобразованы в числовой формат посредством one-hot, label-encoding или target-encoding; метод способа преобразование категориальных данных в числовые был выбран индивидуально по типу переменной.

Для устранения выраженного дисбаланса целевой переменной были апробированы два подхода.

1. Upsampling – метод случайного дублирования наблюдений миноритарного класса. В результате объём выборки увеличен до 568 918 записей, что обеспечило паритет между классами.

2. Downsampling – метод случайного удаление наблюдений мажоритарного класса. Итоговый размер выборки составил 34 242 записи, сохраняя равное представление классов при минимальной избыточности данных.

В качестве методов понижения размерности были выбраны два метода.

1. Principal Component Analysis (PCA) — линейное ортогональное преобразование, позволяющее сохранить максимальную дисперсию в главных компонентах и тем самым сократить размер признакового пространства без значимой потери информации.

2. Correlation-based Feature Selection (CFS) — отбор самых информативных признаков посредством построения и анализа тепловой корреляционной матрицы. Корреляционный анализ показал, что для повышения информативности и уменьшения размерности признакового пространства следует оставить лишь переменные с наибольшей связью с целевой переменной HadHeartAttack. К числу наиболее переменных с наибольшей связью с целевой переменной относятся: HadAngina, GeneralHealth, HadStroke, RemovedTeeth, AgeCategory, HadDiabetes, PhysicalHealthDays, HadCOPD, HadArthritis, HadKidneyDisease, ChestScan, DifficultyWalking, PhysicalActivities, SmokerStatus, Sex, LastCheckupTime, DeafOrHardOfHearing, HadSkinCancer, AlcoholDrinkers, PneumoVaxEver и State. Признаки с низкой корреляцией с целевой переменной были

исключены, что снижает риск переобучения и упрощает последующее моделирование без потери существенной информации.

4. Описание исследуемых моделей машинного обучения

Для выбора оптимальной модели был испытан широкий спектр алгоритмов машинного обучения: от классической логистической регрессии, RandomForest и др. до современных решений — LightGBM, FCNN разных архитектур.

Для повышения эффективности моделирования был применён метод Grid Search в сочетании с кросс-валидацией, что обеспечило систематическую оптимизацию гиперпараметров и выбор конфигураций, демонстрирующих наилучшую обобщающую способность на независимых данных.

Представленный комплексный подход обеспечил всестороннюю оценку как традиционных, так и современных архитектур машинного обучения.

Используются следующие архитектуры FCNN.

1. Первая архитектура:

- i. Полносвязный слой на 128 нейронов, функция активации ReLU;
- ii. Полносвязный слой на 64 нейрона, функция активации ReLU;
- iii. Полносвязный слой на 32 нейрона, функция активации ReLU;
- iv. Выходной полносвязный слой на 1 нейрон с функцией активации Sigmoid.

2. Вторая архитектура:

- i. Входной полносвязный слой на 128 нейронов с функцией активации LeakyReLU ($\alpha=0,01$);
- ii. Слой пакетной нормализации (BatchNormalization);
- iii. Слой Dropout с коэффициентом 0,3;
- iv. Первый скрытый полносвязный слой на 64 нейрона с функцией активации LeakyReLU ($\alpha=0,01$);
- v. Слой пакетной нормализации (BatchNormalization);
- vi. Слой Dropout с коэффициентом 0,3;
- vii. Второй скрытый полносвязный слой на 32 нейрона с функцией активации LeakyReLU ($\alpha=0,01$);
- viii. Слой пакетной нормализации (BatchNormalization);
- ix. Слой Dropout с коэффициентом 0,3;
- x. Выходной полносвязный слой на 1 нейрон с функцией активации Sigmoid.

Архитектура SSD-Convolutional:

- i. Свёрточный слой Conv1D на 32 фильтра;
- ii. Слой пакетной нормализации (BatchNormalization);

- iii. Слой максимального подвыборки (MaxPooling1D);
- iv. Свёрточный слой Conv1D на 64 фильтра;
- v. Слой пакетной нормализации BatchNormalization;
- vi. Слой максимального подвыборки MaxPooling1D;
- vii. Свёрточный слой Conv1D на 128 фильтров;
- viii. Слой пакетной нормализации BatchNormalization;
- ix. Слой глобального усреднения по временной оси (GlobalAveragePooling1D);
- x. Полносвязный слой на 64 нейрона, функция активации ReLU;
- xi. Слой Dropout для регуляризации (параметров 0);
- xii. Выходной полносвязный слой на 1 нейрон с функцией активации Sigmoid.

5. Анализ результатов работы моделей машинного обучения

Проведён систематический подбор гиперпараметров комбинированной модели PCA и логистическая регрессия с использованием GridSearch и 5-фолдной кросс-валидации. Проведён сравнительный анализ 81 конфигурации модели, включавших различные значения доли объяснённой дисперсии в методе главных компонент (PCA) — 0,85, 0,90 и 0,9; решателей — lbfgs, newton-cg, liblinear, saga; типов регуляризации — L1, L2, elasticnet; и значений параметра регуляризации C — 0,1, 1, 10, включая перебор l1_ratio — 0,3, 0,5, 0,7 для elasticnet. Важно отметить, что всего 81 валидная конфигурация, т. к. не все комбинации решателей и типов регуляризации совместимы.

Таким образом была найдена оптимальная конфигурация из исследуемых конфигураций:

```
pca__n_components: 0,95;
clf__solver: newton-cg;
clf__penalty: l2;
clf__C: 10.
```

Модель с данной конфигурацией на тестовой выборке достигла точности 0,805. Дополнительно были протестированы модели с данной конфигурацией, но с различными значениями параметра pca__n_components: сохраняя 95% дисперсии, сохраняя 50% дисперсии, сохраняя 20 параметров.

Также для моделей Logistic Regression, CatBoostClassifier, LightGBM, RandomForestClassifier с помощью алгоритма Grid Search были подобраны оптимальные гиперпараметры.

Модели с наилучшими показателями точности на независимой тестовой выборке представлены в табл. 3.

Таблица 3

**Сравнительные показатели точности моделей машинного обучения
при различных схемах балансировки классов и сокращения размерности**

Название модели машинного обучения	Метод сокращения размерности	Метод балансировки классов	Accuracy
LightGBM	CFS	Downsampling	0,948
DummyClassifier	CFS	Upsampling	0,943
DummyClassifier	CFS	Downsampling	0,943
RandomForestClassifier	CFS	Downsampling	0,940
FCNN с первой архитектурой	CFS	Downsampling	0,930
FCNN с первой архитектурой	CFS	Upsampling	0,925
RandomForestClassifier	CFS	Upsampling	0,850
LightGBM	CFS	Upsampling	0,843
CatBoostClassifier	CFS	Upsampling	0,841
CatBoostClassifier	CFS	Downsampling	0,841
Logistic Regression	CFS	Upsampling	0,835
Logistic Regression	CFS	Downsampling	0,835
FCNN со второй архитектурой	CFS	Downsampling	0,831
Logistic Regression	PCA, сохраняя 95% дисперсии	Downsampling	0,805
Logistic Regression	PCA, сохраняя 20 параметров	Downsampling	0,801
FCNN с первой архитектурой	PCA, сохраняя 95% дисперсии	Downsampling	0,794
Logistic Regression	PCA, сохраняя 50% дисперсии	Downsampling	0,782
SSD-Convolutional	PCA, сохраняя 95% дисперсии	Downsampling	0,537

Примечание. Accuracy рассчитана на независимой тестовой выборке. PCA — *Principal Component Analysis* (метод главных компонент); CFS — *Correlation-based Feature Selection* (корреляционный отбор признаков). NN — полносвязная нейронная сеть.

Изучив табл. 3, можно сделать вывод о важности роли тщательной предобработки данных для повышения эффективности работы моделей машинного обучения.

С точки зрения алгоритмических архитектур моделей машинного обучения, представленных в данной работе, то LightGBM, нейронные сети различных архитектур и Dummy Classifier продемонстрировали высокую способность извлекать релевантные закономерности из многомерных медицинских данных и довольно точно предсказывать наличия инфаркта миокарда у респондентов. Их устойчивые показатели точности

подтверждают релевантность данных моделей для надёжного прогнозирования клинических исходов.

Таким образом, изучив результаты работы моделей машинного обучения, можно сделать вывод, что современный инструментарий машинного обучения способен не только дополнять, но и значительно усиливать традиционные методы анализа, открывая путь к более точным и персонализированным решениям в медицине.

Библиографические ссылки

1. Annual Survey Data. URL: https://www.cdc.gov/brfss/annual_data/annual_data.htm (date of access: 12.08.2025).
2. 2022 BRFSS Survey Data and Documentation. URL: https://www.cdc.gov/brfss/annual_data/annual_2022.html (date of access: 12.08.2025).
3. Fair and explainable Myocardial Infarction (MI) prediction: Novel strategies for feature selection and class imbalance correction / S. B. Akter [et al.] // Computational Biology and Medicine. 2025. Vol. 184. Article no. 109413. DOI: [10.1016/j.combiomed.2024.109413](https://doi.org/10.1016/j.combiomed.2024.109413).
4. Identification of Myocardial Infarction (MI) Probability from Imbalanced Medical Survey Data: An Artificial Neural Network (ANN) with Explainable AI (XAI) Insights / S. B. Akter [et al.] // medRxiv. 2024. URL: <https://www.medrxiv.org/content/10.1101/2024.02.28.24303497v1> (date of access: 12.08.2025).

ОБРАБОТКА МУЛЬТИМОДАЛЬНЫХ ДАННЫХ БИОМЕДИЦИНСКОГО СИГНАЛА С ИСПОЛЬЗОВАНИЕМ МЕТРИКИ ХАУСДОРФА

И. А. Перова, М. В. Вычугжанин

*Пензенский государственный университет,
Пенза, Россия, ivs@pnzgu.ru*

Рассматривается вопрос измерения электрических сигналов сердца на различных предопределенных частотах на платформе MAX30001. Показана обработка мультимодальных данных биомедицинского сигнала, их графическое представление посредством математического пакета Scilab с использованием метрики Хаусдорфа.

Ключевые слова: информационные технологии; биомедицинские данные; платформа MAX30001; электрокардиограмма; метрика Хаусдорфа.

PROCESSING OF MULTIMODAL BIOMEDICAL SIGNAL DATA USING THE HAUSDORFF METRIC

I. A. Perova, M. V. Vychugzhanin

*Penza State University,
Penza, Russia, ivs@pnzgu.ru*

The article discusses the measurement of cardiac electrical signals at various predefined frequencies on the MAX30001 platform. It shows the processing of multimodal biomedical signal data and their graphical representation using the Scilab mathematical package and the Hausdorff metric.

Keywords: information technology; biomedical data; MAX30001 platform; electrocardiogram; Hausdorff metric.

1. Введение

Сегодня информационные технологии активно используются в медицинской отрасли, что привело к значительному расширению возможностей использования специализированного медицинского оборудования. Пациенты получили доступ к современным стационарным и мобильным устройствам, позволяющим регулярно контролировать работу сердечно-сосудистой системы. С помощью этих приборов можно отслеживать важнейшие показатели: уровень артериального давления, частоту сердечных сокращений и другие характеристики организма. Параллельно

развивается и совершенствуется инфраструктура для работы с данными, получаемыми от различных биомедицинских датчиков, которые человек может носить на себе в повседневной жизни. Одной из таких систем является платформа MAX30001 [1].

Исследование, рассматриваемое в статье, направлено на изучение моделирования электрической активности сердца и создание эффективных инструментов визуализации данных при использовании возможностей устройства MAX30001 для мониторинга и анализа сердечной активности в различных условиях. В ходе изыскания была изучена работа устройства на десяти предопределенных частотах (1 кГц, 2 кГц, 4 кГц, 8 кГц, 18 кГц, 40 кГц, 80 кГц, 128 кГц, 250 кГц, 500 кГц) при нахождении человека в сидячем положении без нагрузки при нормальном дыхании. Данные электрокардиограммы (ЭКГ) собираются и анализируются для каждого состояния одновременно с соответствующей частота дискретизации канала ЭКГ.

Для визуального отображения мультимодальных данных биомедицинского сигнала разработана программа в математическом пакете Scilab. Входными данными для приложения являются массивы данных в виде файлов в формате csv, сформированных при работе устройства MAX30001. Фрагмент данных биомедицинского сигнала представлен ниже. Размеры входного массива данных биомедицинского сигнала не являются постоянными и зависят от времени проводимого измерения.

Фрагмент формата данных электрокардиограммы на платформе MAX30001 представлен ниже:

```
% Time (s), ECG_DATA[17:0], ETAG[2:0], PTAG[2:0], ECG (mV), ECG  
Filtered (mV), Filter Type, LDOFF_PH, LDOFF_PL, LDOFF_NH, LDOFF_NL  
0, 0, 0, 7, 0, 0, None,  
0.0078125, 0, 0, 7, 0, 0, None,  
0.015625, -1, 0, 7, -0.0003814697265625, -0.0003814697265625, None,  
0.0234375, -3, 0, 7, -0.0011444091796875, -0.0011444091796875, None,  
0.03125, -3, 0, 7, -0.0011444091796875, -0.0011444091796875, None,
```

2. Основные результаты

В математическом пакете Scilab разработана программа, предназначенная для построения графиков, визуализирующих сигналы ЭКГ по мультимодальным данным, которые загружаются из файла csv, сформированном на платформе MAX30001. Данные из .csv файлов содержат временные метки и значения ЭКГ.

Для сглаживания сигналов был применен метод двусторонней полосовой фильтрации, с помощью которого был удален дрейф и высокочастотный шум без фазового сдвига. На рис. 1–4 представлены отфильтрованные графики сигнала ЭКГ с соответствующей частотой дискретизации – 1 кГц, 2 кГц, 250 кГц и 500 кГц.

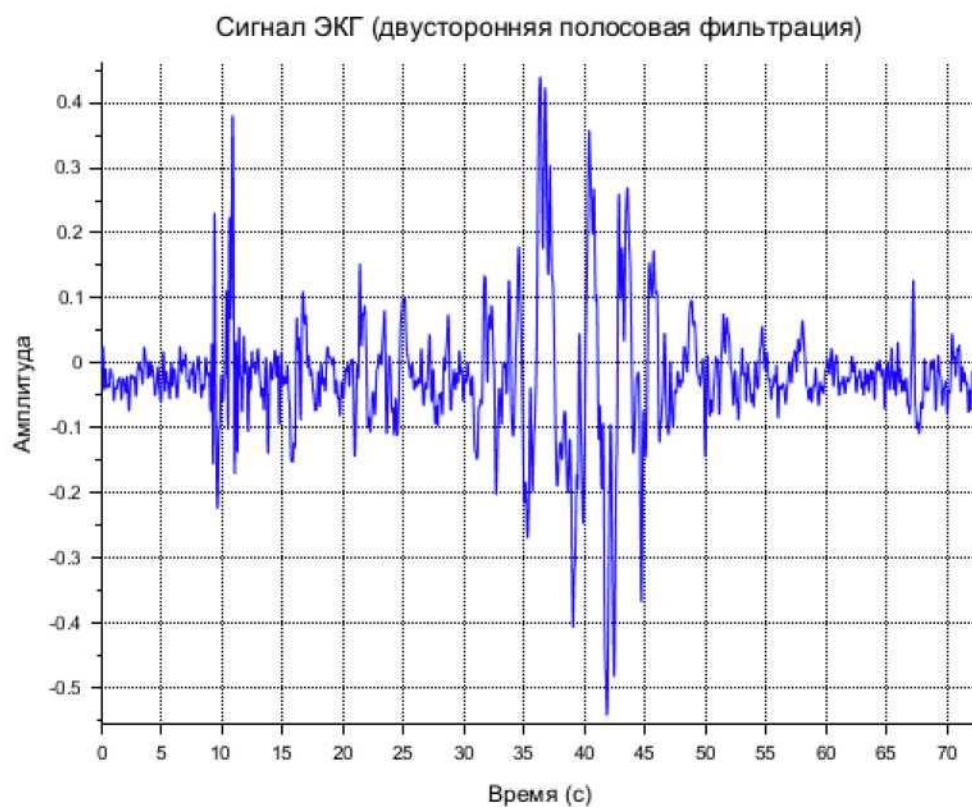


Рис. 1. Отфильтрованный график сигнала ЭКГ с частотой дискретизации 1 кГц



Рис. 2. Отфильтрованный график сигнала ЭКГ с частотой дискретизации 2 кГц

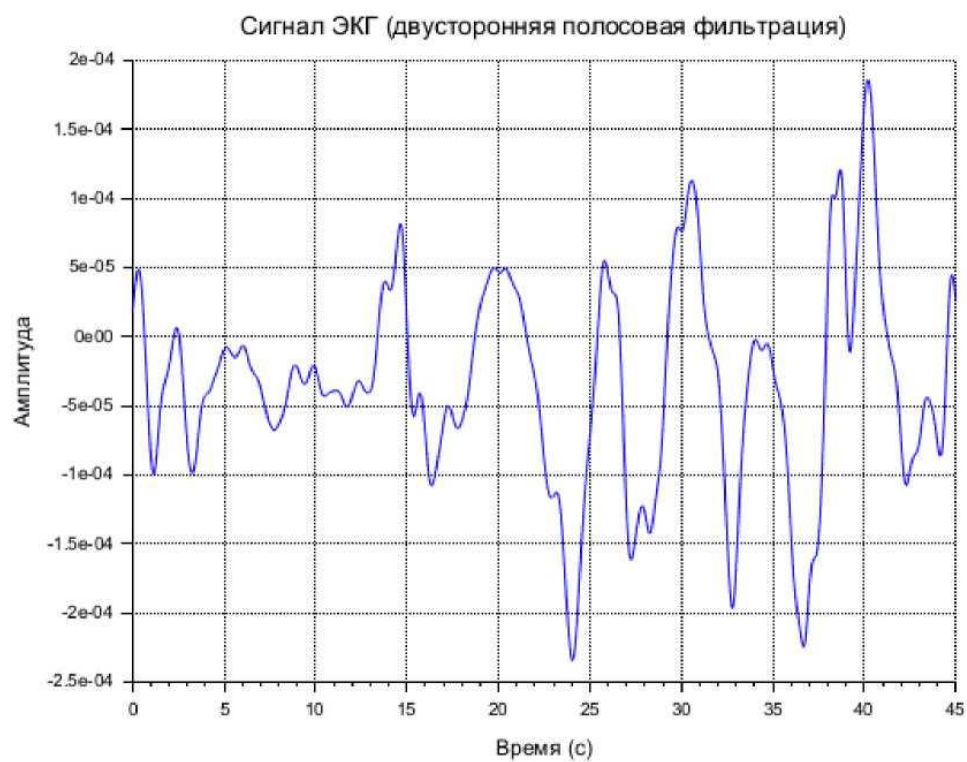


Рис. 3. Отфильтрованный график сигнала ЭКГ с частотой дискретизации 250 кГц

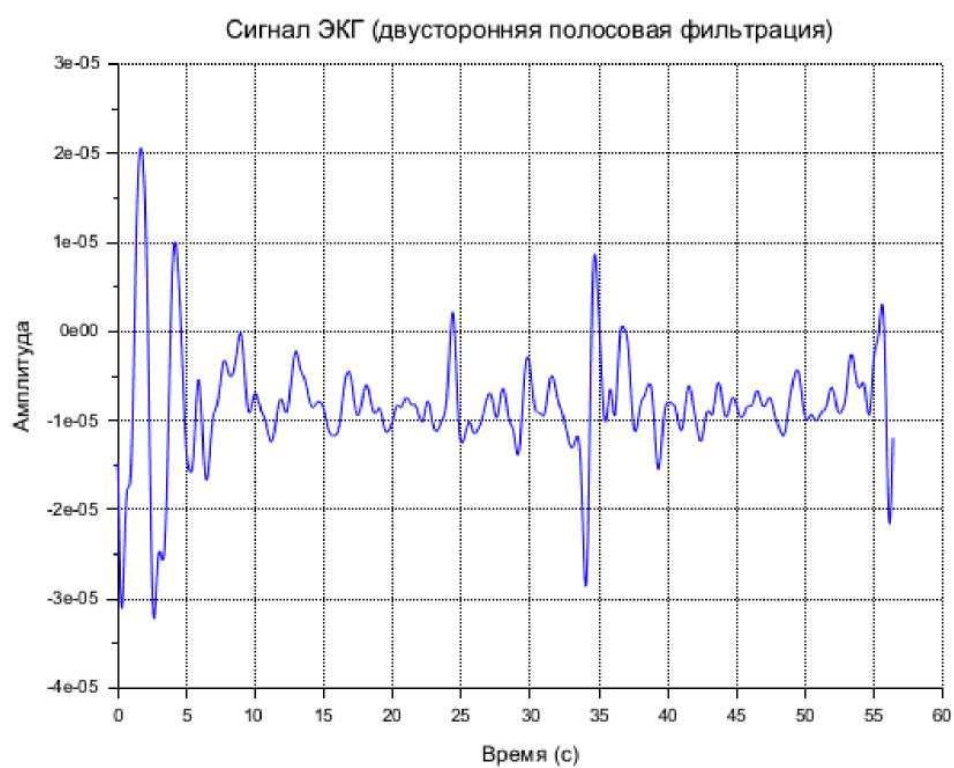


Рис. 4. Отфильтрованный график сигнала ЭКГ с частотой дискретизации 500 кГц

Метрика Хаусдорфа — мера «расстояния» между двумя множествами точек. В контексте сигналов расстояние Хаусдорфа показывает, насколько далеко максимально удалённая точка одного сигнала находится от ближайшей точки другого. Чем меньше расстояние, тем более похожи сигналы. В ЭКГ такую метрику можно использовать для поиска аномалий (например, фибрилляции предсердий), для сравнения сердечных циклов, для кластеризации сигналов или шаблонов [2].

На рис. 5 показано сравнение сигналов разной частотной дискретизации (1 kHz и 2 kHz). Для корректного результата программа определяет общую временную зону. Так сравнение происходит только там, где оба сигнала существуют.



Рис. 5. Сравнение сигналов с частотой дискретизации 1 кГц и 2 кГц

На рис. 5 видно, что сигналы почти совпадают. Но у сигнала 2 в районе 2,2–2,3 секунд наблюдается большой выброс (амплитуда $-0,4955$ и $0,2747$ соответственно). Максимальное отклонение между двумя множествами точек равно $0,437$. В ходе исследования также проведено сравнение сигналов 2 kHz и 4 kHz, 4 kHz и 8 kHz, 8 kHz и 18 kHz, 40 kHz и 80 kHz. На рис. 6 показано сравнение сигналов разной частотной дискретизации (250 kHz и 500 kHz).

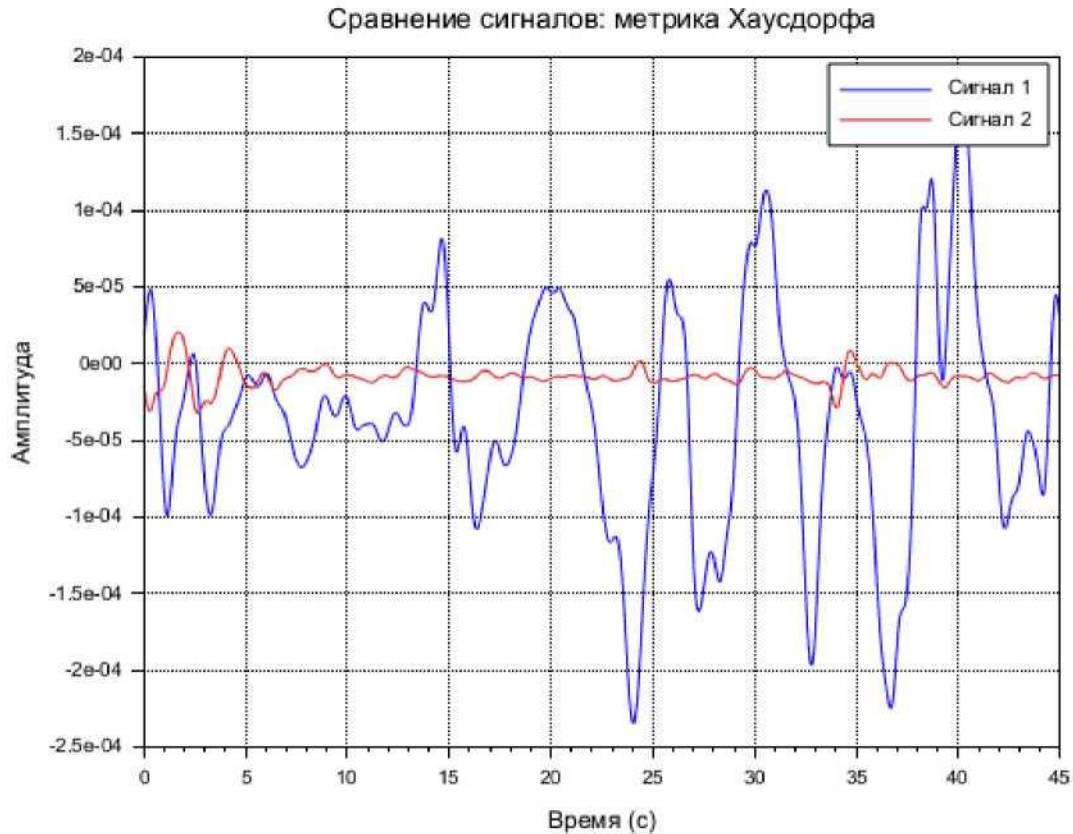


Рис. 6. Сравнение сигналов с частотой дискретизации 250 kHz и 500 kHz

На рис. 6 видно, что сигналы почти совпадают. Но у сигнала 1 наблюдаются значительные выбросы. Максимальное отклонение между двумя множествами точек равно 0,00023. Таким образом, несмотря на различия сигналов на графике, наименьшее расстояние между точками наблюдается у сигналов с частотой дискретизации 250 kHz и 500 kHz.

3. Заключение

В ходе исследования реализована обработка мультимодальных биомедицинских сигналов на примере ЭКГ с применением метрики Хаусдорфа в среде Scilab. Особое внимание уделено сравнению различных участков сигнала и визуализации ключевых точек, отражающих максимальное и минимальное расстояние между ними. Такой подход позволяет не только количественно оценить различие между сегментами сигнала, но и наглядно выделить области наибольшего расхождения или совпадения [3].

Использование Scilab обеспечило удобную и эффективную платформу для математического анализа сигналов, включая загрузку и корректировку исходных данных, построение графиков, реализацию алгоритмов поиска расстояний и интерпретацию полученных результатов. Реализация

метрики Хаусдорфа в Scilab подтвердила её применимость для задач анализа биомедицинских данных, где важна оценка схожести или различия между временными структурами сигналов.

Таким образом, проделанная работа подтверждает актуальность использования метрик расстояния в анализе ЭКГ, а также демонстрирует потенциал Scilab как универсального инструмента в задачах биомедицинской инженерии. В дальнейшем возможно расширение подхода на интеграцию с методами машинного обучения для автоматического выявления аномалий и классификации патологий.

Библиографические ссылки

1. *Вычугжанин М. В., Перова И. А.* Одновременная регистрация и первичная обработка биомедицинских сигналов на платформе MAX30001 // Актуальные проблемы медицинской науки и образования (АПМНО-2024) : сборник статей по материалам X Международной научной конференции, посвященной 25-летию Медицинского института ПГУ / Пенза, 2024. С. 128–131.

2. Способ и аппаратно-программные средства анализа биоимпеданса для систем мобильного мониторинга ЭКГ / М. И. Сафонов [и др.] // Модели, системы, сети в экономике, технике, природе и обществе. 2020. № 3 (35). С. 118–128. DOI: [10.21685/2227-8486-2020-3-10](https://doi.org/10.21685/2227-8486-2020-3-10).

3. *Волчихин В. И., Митрохин М. А., Кузьмин А. В.* Особенности обработки электрокардиосигналов в системах мобильного мониторинга // Известия высших учебных заведений. Поволжский регион. Технические науки. 2018. № 1 (45). С. 54–63.

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ДЛЯ АНАЛИЗА МЕДИЦИНСКИХ ПОКАЗАТЕЛЕЙ ПАЦИЕНТОВ

А. П. Преображенский¹⁾, В. Н. Ряснянская¹⁾, А. А. Буракова¹⁾,
О. Н. Чопоров²⁾

¹⁾ Воронежский институт высоких технологий,
Воронеж, Россия, app@vivt.ru

²⁾ Воронежский государственный медицинский университет,
Воронеж, Россия, choporov_oleg@mail.ru

В работе рассматриваются особенности формирования интеллектуальной системы для анализа медицинских показателей пациентов. Представлена структурная схема компонентов системы. Показана диаграмма активности обработки неструктурированного текста. Обсуждаются основные особенности обработки данных. Точность прогноза с использованием такой системы составила 87%.

Ключевые слова: искусственный интеллект; система; медицинский показатель; прогноз; модель.

DEVELOPMENT OF AN INTELLIGENT SYSTEM FOR ANALYZING PATIENTS' MEDICAL INDICATORS

A. P. Preobrazhenskiy^{a)}, V. N. Ryasnyanskaya^{a)}, A. A. Burakova^{a)},
O. N. Choporov^{b)}

^{a)} Voronezh Institute of High Technologies,
Voronezh, Russia, app@vivt.ru

^{b)} Voronezh State Medical University,
Voronezh, Russia, choporov_oleg@mail.ru

The paper discusses the features of the formation of an intellectual system for the analysis of patients' medical indicators. A structural diagram of the system components is presented. A diagram of unstructured text processing activity is shown. The main features of data processing are discussed. The accuracy of the forecast using such a system was 87%.

Keywords: artificial intelligence; system; medical indicator; forecast; model.

1. Введение

В настоящее время в медицинских системах для автоматизации анализа больших объёмов данных необходимо использовать методы искусственного интеллекта (ИИ). Распределение ответов на вопрос о причинах

недовольства медицинской помощью в 2023 году среди пациентов показало, что около 40–45% назначенных лечений не помогло у участковых терапевтов и врачей специалистов [1], что указывает на наличие проблемы с верно определённым клиническим диагнозом или назначенным лечением. В текущей ситуации ИИ способен будет улучшить качество принятия различных клинических решений. Помимо этой проблемы, ИИ также может помочь замечать сложные и незаметные закономерности в множестве неоднородных данных для выявления более точных диагнозов.

Цель работы заключается в проведении анализа предметной области и существующих методов, проектировании системы, которая будет способна решить поставленные задачи и разработка прототипа интеллектуальной системы анализа медицинских показателей пациентов на основе методов машинного обучения (ML).

2. Основные характеристики системы

Проектирование системы направлено на создание программно-аппаратного комплекса, который автоматизирует анализ медицинских данных, обеспечивает высокую точность анализа, удобство использования и соответствие нормативным требованиям. Проектирование осуществляется поэтапно, с учетом функциональных и нефункциональных требований с использованием различных инструментов проектирования, в том числе и UML, чтобы наглядно и подробно рассмотреть все аспекты системы.

Система состоит из пяти основных модулей, это ввод данных, предобработка, анализ, результат и сопровождение. Далее будет рассмотрен каждый модуль чуть подробнее.

Модуль ввода данных отвечает за загрузку данных из какой-нибудь медицинской информационной системы или через пользовательский интерфейс, проверку их форматов и аутентификацию пользователей. Он включает веб-интерфейс, который позволяет врачам загружать файлы в форматах CSV для различных анализов или других структурированных данных, HDF5 для ЭКГ или других временных рядов [2], DICOM с изображением и метаданными и JSON с неструктурируемыми данными, а также проверяет целостность данных. Аутентификация обеспечивает доступ только для авторизованных пользователей с ролями врача или администратора. Модуль предобработки выполняет обезличивание, очистку, нормализацию, извлечение признаков и векторизацию данных, подготавливая их для анализа. Обезличивание использует механизм распознавания именованных сущностей NLP под названием Named Entity Recognition для удаления персональных данных заменяя их на уникальные

идентификаторы. Предобработанные данные сохраняются в базе данных и временно на локальном устройстве.

Модуль анализа является тем самым модулем, который выполняет задачи классификации, прогнозирования, сегментации и интерпретации результатов с использованием методов ML. Он включает индивидуальные модели для анализа структурированных данных, обработки изображений, анализа временных рядов и обработки текстов. Ансамблевая модель объединяет признаки от этих моделей для получения итоговых вероятностей если это требуется. Интерпретация результатов осуществляется через метод SHAP, который предоставляет объяснения отдельным прогнозам, сделанным моделями ML, определяя вклад признаков. Модели хранятся в формате ONNX, а их метаданные – в отдельной таблице. Модуль вывода результатов формирует отчеты, визуализирует данные и позволяет сохранить их. Веб-интерфейс может отображать таблицы с вероятностями, различные графики и тепловые карты. В модуле сопровождения логирование фиксирует запросы, операции и ошибки в отдельной таблице, обеспечивая аудит. Взаимосвязи между модулями организованы через API и хранилища данных. Более подробно рассмотреть взаимосвязи между компонентами системы можно на структурной схеме (рис. 1).



Рис. 1. Структурная схема компонентов системы

Процесс ввода данных является начальным этапом работы системы, обеспечивающим загрузку медицинских данных от врачей для последующей обработки и анализа. Этот процесс интегрирует функциональность модуля ввода данных, описанного ранее, который включает аутентификацию пользователя, загрузку данных в различных форматах и их предварительную проверку. Обработка неструктурированного текста в системе является одним из самых сложных задач предобработки данных, направленным на преобразование медицинских текстов анамнеза пациента, различных заключений врачей или ответы на опросы, в структурированную форму, пригодную для последующего анализа. Для более подробного представления этого процесса была также построена диаграмма активности на рис. 2.

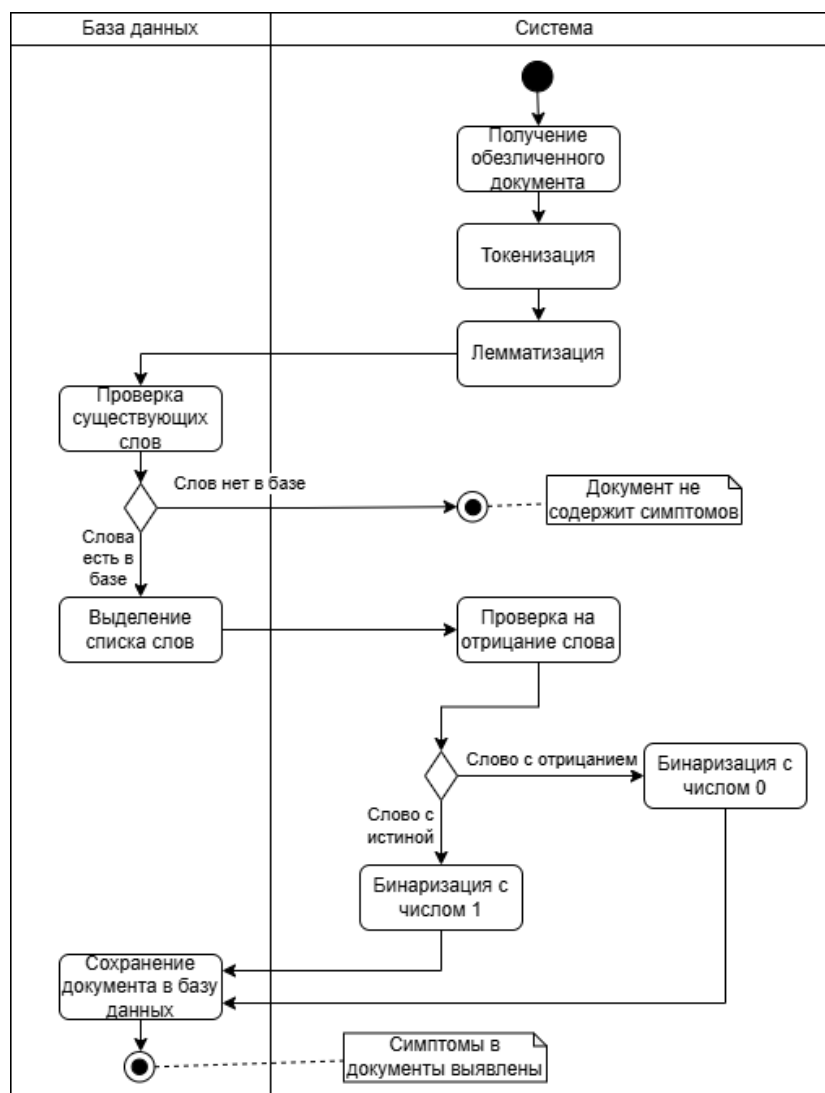


Рис. 2. Диаграмма активности обработки неструктурированного текста

После прохождения компонента, где проводилось обезличивание, а также проверка формата документа, он попадает в компонент предобработки, где, после указания вначале врачом, что это целостный текст, он начинает проходить обработку. Сначала выполняются операции по токенизации и лемматизации, где текст разбивается на отдельные слова или фразы, а затем приводится к начальной форме. Как пример, слово «начались» станет словом «начать», а слово «жалуется» – «жаловаться». Это упрощает последующий анализ, устраняя вариации форм слов. Для этого применяется библиотека spaCy, настроенная на медицинский словарь, что позволяет учитывать специфические термины тех же заключений «гипертония» или «аритмия».

Далее система, выделив весь список подготовленных слов обращается к базе данных в отдельную таблицу со списком всех необходимых

слов для анализа. Если слово в базе данных имеется, то оно сохраняется. Таким образом проверяются все слова и база данных возвращает только список тех слов, которые существуют в базе данных.

Следующим этапом является выделение бинарных признаков симптомов, где системе необходимо преобразовать их в бинарные значения, где 1 – истинно, а 0 – ложно. Но перед этим необходима дополнительная проверка. Система находит эти слова в тексте снова и анализирует ещё пару слов до него и после него на наличие отрицания. Например «кашля нет» должно определиться как 0, поэтому если система обнаружит слово «нет» перед или после слова «кашель», то будет присвоено 0. После этого затем уже выполняется векторизация текста, преобразующая его в числовую форму для анализа моделями ML.

3. Особенности обработки данных

Процесс анализа данных начинается с применения индивидуальных моделей. Каждая модель адаптирована к специфическому типу данных. XGBoost, то есть градиентный бустинг, обрабатывает структурированные данные в системе. В том числе и неструктурированные, которые проходят тщательную предобработку и приобретают структурированность. Модель данного метода предварительно должна быть обучена на тестовых наборах данных вроде MIMIC-III, которая была предложена при анализе. Для анализа изображений был выбран метод ResNet-50, который будет анализировать все возможные медицинские изображения в формате как DICOM, так и PNG или JPG, так как предварительно все эти форматы приводятся к единому, а для анализа возможных временных рядов выбран метод LSTM.

Далее выполняется объединение результатов в ансамблевую модель, если для выявления диагноза потребуется результаты сразу нескольких документов. Признаки от индивидуальных моделей объединяются в единый вектор. Для снижения размерности применяется PCA, оставляя некоторое меньшее количество главных компонентов, сохраняющих, как минимум 95% вариации данных. Для интерпретации результатов работы моделей был выбран метод SHAP. Оценивая вклад каждого признака в предсказание, данный метод генерирует итоговые значения, которые визуализируются как графики.

Общий принцип работы с имеющимися данными в базе и наборах данных с последующим сохранением обученной ансамблевой модели происходит на обучении одной модели на определённых типах данных, требующихся той или иной болезни. То есть, пользуясь МКБ-10, каждая болезнь будет записана в отдельные записи в общей таблице всех болезней системы с указанием тех медицинских показателей, которые требуются

для обнаружения болезни. Соответственно есть два варианта использования моделей ML, где для каждой отдельной болезни будет создана модель, обученная на том необходимом списке данных, или будет одна модель на несколько болезней, имеющих схожие показатели и необходимые данные для обучения.

Система реализована на языке программирования Python. В качестве инструмента для реализации реляционной базы данных была выбрана система управления базами данных SQLite, которая обеспечит эффективное хранение и управление структурированными данными в процессе разработки и тестирования. Для создания веб-интерфейса использовалась библиотека Flask, фреймворк для Python, позволяющий быстро разворачивать веб-сервисы с минимальной настройкой, что отлично подходит для реализации текущего прототипа, а HTML и CSS применялись для структурирования и стилизации контента, обеспечивая адаптивный дизайн и визуальную согласованность интерфейса. Динамическое поведение страниц реализовывалось с использованием JavaScript, интегрированного непосредственно в HTML код.

Был выбран открытый доступный набор данных под названием “Chronic Kidney Disease” на вебсайте Kaggle. Как указано в описании, этот набор данных представляет собой коллекцию медицинских показателей, предназначенных для диагностики хронической болезни почек. Он содержит 25 атрибутов, включая как количественные, так и категориальные переменные, а также целевую переменную, определяющую наличие или отсутствие диагноза. Данные были собраны в больнице в течение примерно 2 месяцев. В данном тестировании набор данных используется для обучения модели градиентного бустинга с целью классификации наличия данной болезни у пациентов [3]. Рассмотрев данный набор данных, можно заметить, что он включает пропущенные значения. Общее количество записей составляет 400, что хоть и является небольшим количеством, но для диагностики наличия лишь одного диагноза этого вполне будет достаточно. После подготовки набора данных, включающей загрузку файла, анализ атрибутов и их соотнесение с признаками, и другими подготовками перед обучением следующим этапом является процесс обучения модели машинного обучения. Пока происходит обучение, пользователь может свернуть или закрыть это окно, если обучение будет происходить долго. Но в текущем примере набора данных это происходит быстро, так как данных немного.

Процесс начинается с агрегации данных, собранных на предыдущих этапах. Важным шагом является валидация, где проверяется наличие пропусков, несоответствий в записях набора данных, а также корректность типов данных. Для обработки пропусков применяется заполнение

медианными значениями. Для обучения таких типов данных используется XGBoost, которая использует градиентный бустинг для последовательного улучшения предсказаний через добавление слабых моделей, как было выбрано при проектировании. Данные разбиваются на обучающую и тестовую выборки для объективной оценки модели с соотношением 80 на 20 процентов. Это делается с использованием стратифицированного разбиения «stratify=y», чтобы сохранить пропорции классов. Обучение проводится с помощью метода «fit», а для мониторинга процесса может использоваться валидационная выборка. Для повышения точности модели проводится оптимизация гиперпараметров с использованием техники Grid Search. После обучения модель сохраняется в файл формата pkl с названием `диагноз_xgboost_model`, а путь к данной модели сохраняется в базу данных.

4. Заключение

Точность модели составила 0,87, что означает, что 87% всех прогнозов оказались правильными. Это высокое значение отражает общую надёжность модели на тестовом наборе. Значение AUC-ROC, равное 0,88, подтверждает её отличную способность различать классы, что согласуется с визуальной интерпретацией ROC-кривой. F1-Score, достигший 0,86, представляет собой гармоническое среднее между точностью и полнотой и указывает на сбалансированность модели. Эти метрики в совокупности демонстрируют, что обученная модель на выбранном наборе данных успешно справляется с задачей классификации.

Библиографические ссылки

1. *Прорехин Э. А.* Перспективы развития российского здравоохранения после эпидемии коронавируса // Экономика и управление: проблемы, тенденции, перспективы : материалы XIII научной конференции студентов и аспирантов. Саратов, 2023. С. 204–209.
2. Исследование LSTM-нейросетевого подхода при моделировании временных рядов / Г. Г. Рапаков [и др.] // Вестник Череповецкого государственного университета. 2023. № 3 (114). С. 47–54.
3. *Хузяханов Е. И., Шарко С. С., Поляков П. М.* Эффективные методы предобработки данных: как подготовить данные для машинного обучения // Современные научные исследования и инновации. 2024. № 11. URL: <https://web.snauka.ru/issues/2024/11/102801> (дата обращения: 25.07.2025).

**РАЗРАБОТКА R-ПАКЕТА HUGSVCALLER,
ПРЕДНАЗНАЧЕННОГО ДЛЯ ИДЕНТИФИКАЦИИ
И АННОТАЦИИ САЙТОВ ОДНОНУКЛЕОТИДНОГО
ПОЛИМОРФИЗМА В ГЕНОМЕ ЧЕЛОВЕКА**

**Д. Д. Сарнацкий, И. Н. Ильюшенко, Н. А. Боева, Л. С. Воронюк,
Н. Н. Яцков, В. В. Скакун, В. В. Гринев**

*Белорусский государственный университет,
Минск, Беларусь, grinev_vv@bsu.by*

В работе описана архитектура и функциональные возможности нового R-пакета huGSVcallerR, предназначенного для идентификации и аннотации сайтов однонуклеотидного полиморфизма в геноме человека. Пакет включает более 40 оригинальных функций, написанных на языках программирования R и C++. Функции собраны в шесть модулей, которые в свою очередь интегрированы в единый конвейер обработки данных геномного секвенирования.

Ключевые слова: геномное секвенирование; конвейер данных; генетический полиморфизм; идентификация и аннотация; R-пакет.

**DEVELOPMENT OF THE R-PACKAGE huGSVcallerR
FOR IDENTIFICATION AND ANNOTATION OF SINGLE
NUCLEOTIDE POLYMORPHISM SITES IN THE HUMAN GENOME**

**D. D. Sarnatski, I. M. Ilyushonak, N. A. Boyeva, L. S. Varaniuk,
M. M. Yatskou, V. V. Skakun, V. V. Grinev**

*Belarusian State University,
Minsk, Belarus, grinev_vv@bsu.by*

The paper describes the architecture and functionality of the new R-package huGSVcallerR, designed to identify and annotate single-nucleotide polymorphism sites in the human genome. The package includes more than 40 original functions written in the R and C++ programming languages. The functions are grouped into six modules, which in turn are integrated into a single pipeline for processing genomic sequencing data.

Keywords: genomic sequencing; pipeline; genetic polymorphism; identification and annotation; R-package.

1. Введение

Геном человека можно представить как траекторию в вероятностном пространстве, где каждая позиция с некой вероятностью может быть полиморфной в индивидуальных геномах. Самым распространенным

вариантом вариативности нуклеотидной последовательности генома является однонуклеотидный полиморфизм, или SNVs. Большинство однонуклеотидных замен не имеют фенотипического проявления, однако в некоторых случаях эти замены значимы и могут приводить не только к фенотипическим изменениям в пределах нормы, но и к развитию заболеваний. Идентификация сайтов таких SNVs по данным геномного секвенирования до сих пор не имеет стандартизированного решения. В связи с этим мы задались целью разработать интегрированный пакет компьютерных программ, выполняющий все этапы поиска сайтов SNVs по данным геномного секвенирования в единой R-среде.

2. Дизайн исследования

Разрабатываемый пакет, согласно цели и задачам исследования, должен был объединить порядка 40 оригинальных функций и оболочек функций, написанных на языках программирования R и C++. Количество разрабатываемых функций определялось вычислительными задачами, на решение которых был нацелен пакет. Выбор языка программирования R в качестве основного был обусловлен теми преимуществами, которыми он обладает, являясь лидирующим языком в статистике и биоинформатике. В тех же случаях, когда требуются быстрые вычисления, было решено воспользоваться возможностями языка C++.

Все функции пакета было решено сгруппировать по шести модулям, взаимосвязанным между собой. Первый модуль должен объединять функции, обеспечивающие оценку и улучшение качества исходных ридов, предоставляя пользователю развернутые отчеты. Второй модуль должен отвечать за локальное картирование ридов, оценку качества такого картирования и подготовку соответствующих отчетов. Третий модуль может быть комбинированным модулем, функции которого должны быть нацелены на подготовку данных для обнаружения сайтов SNVs либо с помощью модели GATK, либо с помощью классических статистических тестов, либо через машинное обучение. Соответственно, четвертый модуль должен предоставлять возможность пользователю проводить первичную идентификацию сайтов SNVs либо в рамках подхода GATK, либо с помощью одного из пяти классических статистических тестов, либо с помощью алгоритмов машинного обучения. В зависимости от того, какой подход выбирается, пользователю должна предоставляться гибкая система фильтрации списка первичных кандидатур SNVs с помощью функций пятого модуля. И, наконец, шестой модуль должен содержать множество функций, позволяющих осуществлять как базовую, так и специальную аннотацию сайтов SNVs, включая поиск ассоциаций со сплайсингом и анализ аминокислотных замен в соответствующих белках.

Таким образом, разрабатывая пакет, мы придерживались строгой логики в пошаговом развитии архитектуры такого пакета, учитывая, что в конечном итоге пакет должен обеспечивать конвейерную обработку данных геномного секвенирования. Кроме того, как обязательный компонент разработки, мы включили два разных инструмента валидации результатов, получаемых с помощью пакета: экспериментальную валидацию и биоинформатический бенчмаркинг.

3. Результаты исследования

Архитектура пакета huGSVcalleR представлена на рис. 1. Как уже отмечалось, первый модуль пакета объединил функции, главным образом обеспечивающих оценку качества исходных данных. Результаты работы таких функций предоставляются пользователю в виде обобщений трех типов: таблица с обобщенной статистикой качества, набор обобщающих графиков качества и мульти-FASTA запись сверхпредставленных ридов. Первые два отчета записываются в виде *.PDF файлов, третий – как файл стандартного формата FASTA. Таблица, помимо сведений об образце, дате/времени и методике оценки качества, включает 47 показателей качества, часть из которых более подробно развернута и графически проиллюстрирована во втором отчете.

Исходя из своего собственного опыта работы с данными геномного секвенирования, а также опыта, накопленного другими исследователями, мы включили в перечень рассчитываемых характеристик как общепринятые показатели качества (например, распределение длины ридов), лишь исключив из конечного списка низкоинформативные характеристики, так и некоторые дополнительные показатели. Из наиболее значимых инноваций, реализованных в первом модуле пакета, хотим обратить внимание на следующие показатели: 1) сделана замена распределения ридов по качеству на распределение доли ридов, преодолевающих пороговое значение качества (рис. 2, *а*); 2) график Q-Q частоты встречаемости GC нуклеотидов в ридов был заменен на график распределения эмпирической и теоретически ожидаемой (при условии нормальности распределения данных) плотности вероятности с добавлением результатов теста Колмогорова-Смирнова; 3) анализ распределения частоты встречаемости k-меров был заменен на оценку композиционной сложности ридов (рис. 2, *б*); 4) под управление пользователя был отдан список адаптеров и чужеродных последовательностей, которые могут детектироваться в ридов.

Ориентируясь на показатели качества, пользователь может, используя функцию *cleanRawReads()* из первого модуля, настроить и провести фильтрацию и улучшить качество ридов по семи основным показателям: от

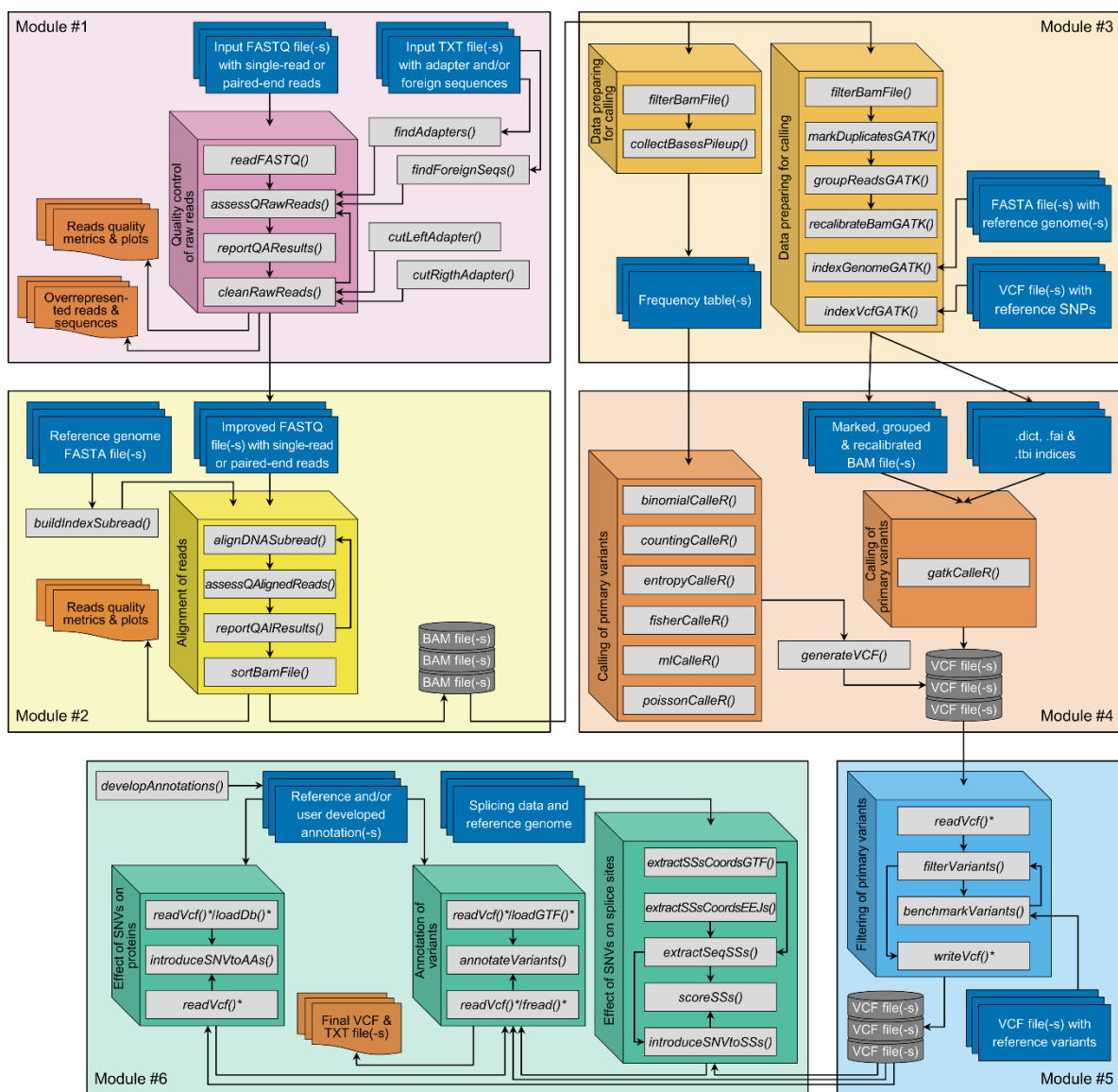


Рис. 1. Архитектура пакета huGSVcaller

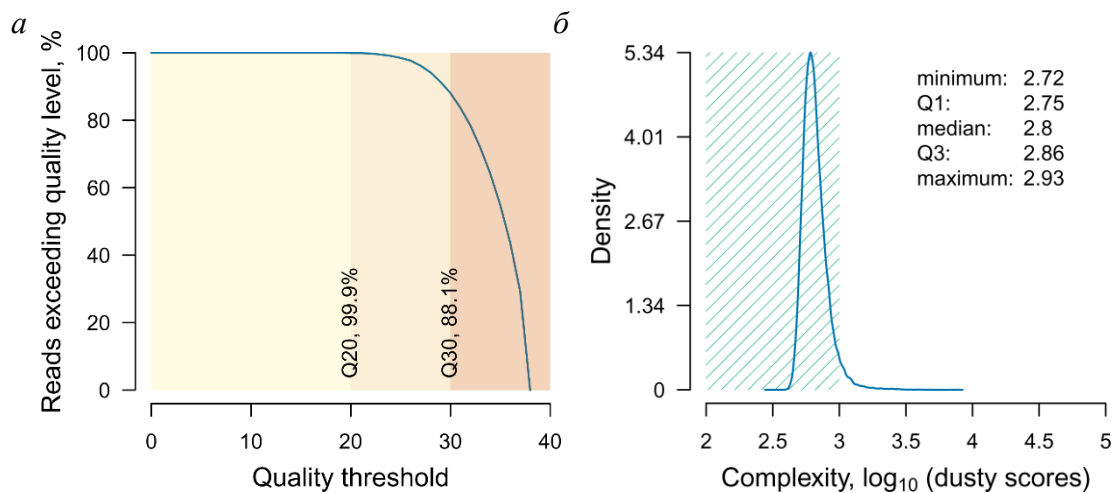


Рис. 2. Некоторые показатели качества первичных данных (пояснения в тексте)

тримминга адаптерных и других нежелательных последовательностей до удаления дубликатов. При необходимости прошедшие фильтрацию и улучшенные риды могут подвергаться повторной оценке качества и только потом передаваться на следующий модуль пакета.

Во втором модуле мы собрали функции, обеспечивающие индексацию выбранного эталонного генома, картирование ридов относительно такого генома, оценку качества картирования (с созданием соответствующих отчетов), а также сортировку и индексацию полученного *.BAM файла. В текущей версии пакета проведение картирования возможно только с помощью картировщика Subread. Однако мы планируем расширить перечень доступных алгоритмов, подключив к пакету такие картировщики, как Bowtie 2, BWA и HISAT2. В любом случае, результативность картирования с помощью функции *assessQAlignedReads()* оценивается по 23 показателям, включая уникальность картирования ридов, корректность картирования парных ридов, подсчет синглтонов, качество картирования, размер секвенированного фрагмента и т.д. Рассчитанные показатели с помощью функции *reportQAlResults()* предоставляются пользователю в двух *.PDF файлах в табличном и графическом виде.

Третий модуль пакета объединяет семь служебных функций, позволяющих подготовить прошедшие через второй модуль данные к процедуре поиска сайтов SNVs. Характер подготовки зависит от того, какой именно вычислительный подход будет использован для поиска полиморфных сайтов. Если классические статистические тесты, то подготовка сводится к сбору с помощью функции *collectBasesPileup()* данных по частоте встречаемости четырех типов нуклеотидов по каждой из изучаемых геномных позиций. Если же запланирован к применению метод GATK, то в *.BAM файле необходимо маркировать дубликаты ридов, сгруппировать их, сделать ре-калибровку качества секвенирования, а также провести GATK-индексации эталонного генома и *.VCF файла с известными сайтами полиморфизма. В обоих случаях предварительно можно с помощью функции *filterBamFile()* провести фильтрацию *.BAM файла по показателям качества картирования, рассчитываемых благодаря функционалу предыдущего модуля.

В четвертом модуле собраны функции, позволяющие провести поиск сайтов полиморфизма либо с помощью метода GATK, либо с помощью классических статистических тестов, либо через машинное обучение. В пакете реализовано пять классических статистических тестов: от энтропийного подхода до точного теста Фишера (см. рис. 1). При выборе машинного обучения пользователю предлагается пять предобученных моделей в формате .rds, доступных после загрузки пакета. Для осуществления выбора алгоритма и контроля его реализации в четвертый модуль добавлена функция *callSNVs()*, которая интегрирует работу всех

низкоуровневых функций обнаружения сайтов SNVs. На выходе четвертый модуль возвращает файл формата *.VCF версии 4.0 с первичными кандидатами на сайты SNVs.

Фильтрация списка первичных кандидатов на сайты SNVs может быть осуществлена с помощью функционала пятого модуля пакета. В текущей версии пакета реализована только жесткая фильтрация с помощью функции *filterVariants()* по следующим показателям: геномная локализация, надежность обнаружения сайта, глубина секвенирования, общая гетерозиготность сайта, а также гетерозиготность первой альтернативной аллели (если сайт представлен не одной, а несколькими альтернативами).

Этот же модуль включает функцию *benchmarkVariants()* для бенчмаркинга, которая позволяет провести сравнение полученных результатов либо с эталонными результатами (если таковые имеются для изучаемых первичных данных), либо между разными подходами, и использовать итог сравнения в настройках процедуры фильтрации. Бенчмаркинг – это способ проверить, насколько точно и надёжно работают алгоритмы или методы, сравнив их результаты с известным эталоном или симулированными (искусственно созданными) данными. Для объективной оценки используются три ключевые метрики: 1) точность (precision), отражающую долю истинных положительных вариантов среди всех случаев, предсказанных как положительные; 2) чувствительность (recall), показывающую долю найденных истинных вариантов среди всех реально существующих; 3) балы F1 (F1-score), которые являются гармоническим средним между точностью и чувствительностью. Эти показатели позволяют оценить способность алгоритма корректно идентифицировать истинные варианты и избегать ложноположительных результатов.

Функция *benchmarkVariants()* принимает на вход тестируемый файл *.VCF и файл *.VCF, именуемый эталонным (или «золотым стандартом»), который содержит информацию о сайтах полиморфизма, идентифицированных с высокой степенью достоверности. Информация о вариантах сравнивается по ключевым параметрам: хромосомной принадлежности, позиции на хромосоме, а также соответствию референсных и альтернативных аллелей. В результате все варианты классифицируются как истинные положительные, ложноположительные или ложноотрицательные. При необходимости пользователь может использовать аргумент *featureTable* выше указанной функции и сформировать сводную таблицу, где каждому варианту присвоена соответствующая метка. На основании подсчитанных категорий функция вычисляет метрики точности, чувствительности и F1-score. В консоли выводятся значения этих метрик для однонуклеотидных вариантов, инделов, а также для всех записей в совокупности. Такое разделение позволяет оценить, насколько хорошо алгоритм обрабатывает на разных типах нуклеотидных вариаций.

Наконец, в шестом модуле собраны функции для базовой и специальных аннотаций идентифицированных сайтов SNVs, успешно прошедших фильтрацию через пятый модуль. Базовая аннотация осуществляется с помощью функции *annotateVariants()*. Эта функция генерирует две группы показателей, характеризующих каждый из обнаруженных сайтов SNVs. Первой группой таких показателей является геномный контекст сайта: его геновая принадлежность, попадание сайта на тот или иной структурно-функциональный элемент гена, является ли сайт уже известным или же обнаружен новый вариант, и т.д. Ко второй же группе показателей мы отнесли ожидаемые клинические последствия от появления такого сайта полиморфизма.

Что касается специальных аннотаций, то в текущей версии пакета реализовано два варианта таких описаний: влияние однонуклеотидных замен на силу донорных и акцепторных сайтов сплайсинга, а также идентификация сайтов несинонимичных SNVs, приводящих к аминокислотным заменам в соответствующих белках. При локализации сайта SNVs в границах сайта сплайсинга через функцию *introduceSNVtoSSs()* проводится соответствующая корректировка нуклеотидной последовательности последнего. Далее с помощью функции *scoreSSs()* рассчитывается сила скорректированного сайта сплайсинга по моделям максимальной энтропии и сравнивается с таким же показателем, но для эталонной последовательности сайта сплайсинга.

Для изучения аминокислотных замен пользователю предлагаются две опции: либо использовать эталонные аннотации генома (например, предлагаемые Ensembl) и по ним проводить, задействовав функцию *introduceSNVtoAAs()*, вначале корректировку нуклеотидной последовательности CDS-областей генов, а затем корректировку аминокислотной последовательности соответствующих белков (если нуклеотидная замена оказалась несинонимичной), либо использовать свои собственные аннотации для этих же целей. Для подготовки корректных пользовательских аннотаций, основанных на сборках экспериментальных транскриптомов, в модуль добавлена специализированная функция *developAnnotations()*.

В конечном итоге все функции описанных модулей мы организовали в единый пакет под названием huGSVcalleR. Архитектура этого пакета соответствует традиционной организации R-пакетов. При этом пакет оформлен под требования репозитория Bioconductor: в пакете использованы core-классы Bioconductor в качестве основы, семантика функций и всего пакета следует принципам S4-ориентированного программирования, он независим от эталонных аннотаций, пакет поддерживает работу с большими данными через использование эффективных структур данных, функции пакета четко документированы и обеспечены помощью.

Для обеспечения совместимости и управляемости зависимостями пакет был оформлен с использованием современной инфраструктуры

devtools и roxygen2. В файле DESCRIPTION явно указаны все импортируемые внешние пакеты (более 30 зависимостей, включая ключевые пакеты Bioconductor Biostrings, GenomicRanges, VariantAnnotation и др.), необходимые для корректной работы разрабатываемого пакета. Функции, требующие больших вычислительных мощностей (например, статистические тесты для идентификации вариантов, извлечение признаков для методов машинного обучения), были реализованы на C++ с использованием пакета Rcpp и многопоточности с использованием пакета RcppParallel. Главная страница собранного пакета представлена на рис. 3.

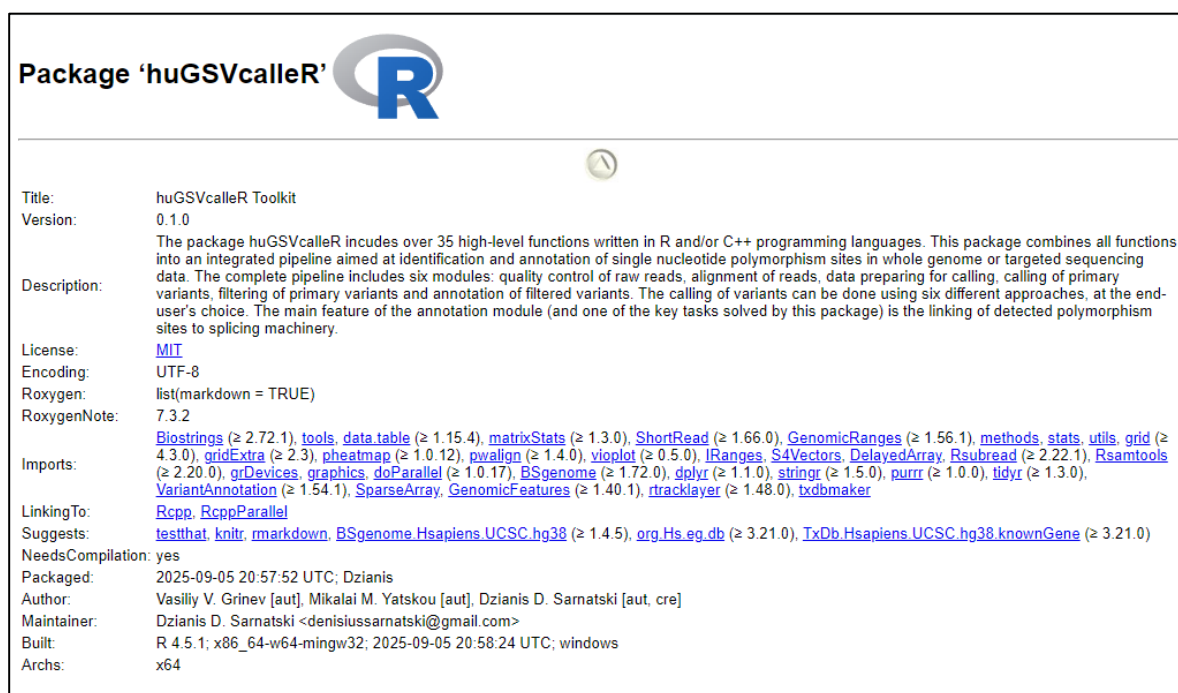


Рис. 3. Главная страница пакета huGSVcallerR

В папке R, предназначенной для расположения R функций пакета, файлы с функциями организованы в соответствии с рис. 1 в шесть последовательно связанных модулей: 1) rawReadsQC.R – контроль качества исходных данных; 2) readsAlignent.R – картирование ридов; 3) preparingCallingData.R – подготовка данных к коллингу; 4) calling.R – сам коллинг; 5) filteringAndBenchmarkingVariants.R – фильтрация и бенчмаркинг SNVs; 6) annotatingAndMeasuringEffects.R – аннотация и оценка эффекта SNVs. В папке src, предназначенной для размещения C++ реализаций, находятся функции пяти статистических тестов: *snp_binomial_cpp_parallel.(h/cpp)*, *snp_counting_cpp_parallel.(h/cpp)*, *snp_entropy_cpp_parallel.(h/cpp)*, *snp_fisher_cpp_parallel.(h/cpp)*, *snp_poisson_cpp_parallel.(h/cpp)*, которые вызываются через функцию *callSNVs()*, доступную для пользователя. Также в папке src расположен

набор вспомогательных C++ функций, предназначенных для ускорения выполнения в сравнении с высокоуровневыми R реализациями (*utils.(h/cpp)*) и функция для быстрого извлечения признаков нуклеотидных сайтов для применения методов машинного обучения *vectorise_data_cpp_parallel.(h/cpp)*.

Интеграция модулей в единый конвейер обеспечивается последовательной передачей выходных данных одной функции на вход другой в соответствии с общей архитектурой пакета (см. рис. 1). Например, выход функции *cleanRawReads()* (очищенные FASTQ-файлы) является прямым входом для функции *alignDNASubread()*, а результат ее работы (файл *.BAM) используется функциями *assessQAlignedReads()* и *collectBasesPileup()*.

Для проверки корректности интеграции всех компонентов и работоспособности пакета в целом разработан комплексный набор тестов с использованием фреймворка *testthat*. Все тестовые наборы симулированных и публичных данных реальных геномов размещены в папке *extdata*. В этой же папке размещены предобученные модели машинного обучения в формате *.rds*, доступные пользователю после загрузки пакета. Для пакета разработан ряд модульных тестов, проверяющих корректность выполнения отдельных функций на специально подготовленных тестовых данных (например, искусственные FASTQ и BAM-файлы небольшого объема, миниатюрные эталонные геномы).

4. Заключение

Таким образом, нами разработан новый R-пакет *huGSVcaller*, интегрирующий множество оригинальных функций в единый конвейер обработки данных геномного секвенирования. Пакет имеет модульную организацию с гибкой инфраструктурой, которая легко адаптируется под разные задачи, связанные с выявлением и характеристикой сайтов SNVs. Представленная разработка может быть полезной как для фундаментальных исследований, так и для решения прикладных задач в клинической диагностике и прогнозировании.

БИОИНФОРМАТИЧЕСКИЙ ПОДХОД НА ОСНОВЕ МЕТАМОДЕЛЕЙ ДЛЯ УЛУЧШЕНИЯ ПРОГНОЗИРОВАНИЯ ИММУНОГЕННОСТИ ОПУХОЛЕВЫХ ЭПИТОПОВ

И. С. Трусав, В. В. Гринев

*Белорусский государственный университет,
Минск, Беларусь, ivan.trusau@yandex.by*

В данной работе предложен и апробирован биоинформатический подход, основанный на метамоделях, комплексно оценивающих весь каскад событий от внутриклеточной обработки антигена до его распознавания иммунной системой. Проведённый в работе сравнительный анализ продемонстрировал преимущества метамоделей по сравнению с существующими аналогами в точности предсказания презентации и иммуногенности эпитопов.

Ключевые слова: неоантигены; прогнозирование иммуногенности; машинное обучение; мета модель.

BIOINFORMATIC APPROACH BASED ON METAMODELS FOR IMPROVING THE PREDICTION OF THE IMMUNOGENICITY OF TUMOUR EPITOPES

I. S. Trusau, V. V. Grinev

*Belarusian State University,
Minsk, Belarus, ivan.trusau@yandex.by*

In this paper, we propose and test a bioinformatics approach based on metamodels that comprehensively evaluate the entire cascade of events from intracellular processing of an antigen to its recognition by the immune system. A comparative analysis conducted in this work demonstrated the advantages of metamodels over existing analogues in terms of the accuracy of predicting immunogenicity and epitope presentation.

Keywords: neoantigens; immunogenicity prediction; machine learning; metamodel.

1. Введение

В современной онкологии иммунотерапия признана одним из наиболее перспективных и быстро развивающихся. Иммунотерапия открывает перспективы для пациентов с различными формами рака, которые раньше считались неизлечимыми. Клинический успех данного подхода во многом

обусловлен способностью нашей иммунной системы самостоятельно распознавать и уничтожать опухолевые клетки. Наиболее значимым элементом в этом механизме являются неоантигены, представляющие собой уникальные пептидные фрагменты, возникающие в результате соматических мутаций и отсутствующие в нормальных тканях организма. Их специфичность позволяет использовать их в качестве идеальных мишеней при лечении, минимизируя риск побочных эффектов и аутоиммунных реакций на здоровые клетки [5].

Для распознавания неоантигена иммунными клетками необходим целый ряд сложных и многоступенчатых реакций, называемых процессингом. Инициация процессинга антигена происходит на стадии протеолитического расщеплением белка-предшественника специализированными ферментными комплексами – протеасомами. Образующиеся пептидные фрагменты затем транспортируются в эндоплазматический ретикулум при участии транспортёра ТАР, ассоциированного с процессингом антигена. В эндоплазматическом ретикулуме происходит укорочение пептидов специальными ферментами и их связывание с молекулами главного комплекса гистосовместимости (МНС) класса I. Пептид, связанный с МНС, также называют эпитопом. Такой комплекс впоследствии выносится на клеточную поверхность для презентации Т-лимфоцитам. Последние, благодаря своему Т-клеточному рецептору, способны распознать комплекс МНС-эпитоп и запустить иммунный ответ. Прохождение каждого этапа зависит от физико-химических свойств пептидов, поэтому комплексная оценка процесса крайне важна для определения способности неоантигена вызывать эффективный противоопухолевый иммунитет.

Несмотря на бурное развитие биоинформатических инструментов для прогнозирования потенциальных иммуногенных неоантигенов, их точность остаётся недостаточной, что ограничивает их широкое клиническое применение. Существующие модели зачастую фокусируются лишь на одном из этапов обработки антигена. Подобный узкоспециализированный подход не учитывает всю сложность процесса, что обуславливает появление множества ложных результатов, существенно снижающих предсказательную силу моделей и создавая серьёзное препятствие перед разработкой действительно персонализированных вакцин и иммунотерапии онкологических заболеваний.

В связи с вышеизложенным мы задались целью разработать комплексный биоинформатический подход, учитывающий все этапы процессинга неоантигенов и их распознавания иммунной системой. Данный подход, основанный на использовании мета-классификаторов, направлен на существенное повышение точности прогнозирования иммуногенности эпитопов неоантигенов, что, по нашему мнению, может повысить

эффективность и надёжность идентификации молекулярных мишеней для персонализированной иммунотерапии рака.

2. Материалы и методы

При выполнении данной работы был использован язык программирования Python 3.8. Для эффективной обработки данных и реализации алгоритмов машинного обучения применялись библиотеки NumPy, Pandas и Scikit-learn. Вся работа проводилась на компьютере с 16 Гб оперативной памяти, 8-ядерным 16-поточным процессором, видеокартой Nvidia GeForce GTX 1650 4 Гб и SSD.

Основные данные были получены из открытых баз IEDB и CEDAR. Данные содержали информация о пептидах, их связывании с молекулами МНС класса I и иммуногенности. Дополнительные данные были взяты из публикаций, где описывались программные пакеты ConvNeXt, TranspMHC, ImmuneApp, PRIME, BigMHC, ICERFIRE и IEPAPI. Собранные данные прошли оценку качества и фильтрацию, после чего они были сгруппированы в два класса – данные об иммуногенности и данные о презентации. Данные об иммуногенности мы разделили на два набора: 1) набор Neo, куда попали только данные о неоантигенах, и 2) набор IE, который включил пептиды инфекционной природы.

Для прогнозирования процессинга эпитопа использовались модели предсказания протеасомного расщепления (pepsickle), транспорта пептидов через TAP (DeepTAP), N-концевого тримминга ферментом ERAP1 (ERAMER) и связывания эпитопа с молекулой МНС (TranspMHC). Для оценки иммуногенности применялись модели, оценивающие секрецию интерлейкинов: IL2 (IL2pred), IL4 (IL4pred), IL6 (IL6pred), IL10 (IL10pred), IL13 (IL13pred), интерферон гамма (IFNepitope 2) и фактор некроза опухолей (TNFepitope).

В качестве физико-химических свойств были выбраны гидрофобность, алифатический индекс, индекс Бомана, изоэлектрическая точка и длина пептида, которые вычислялись при помощи библиотеки peptides [7]. Также проводилась оценка чужеродности пептида по отношению к человеческому протеому при помощи программы antigen.garnish [10].

Для создания метамоделей использовались алгоритмы случайного леса, реализованные при помощи библиотеки Scikit-learn. Оценка моделей осуществлялась с использованием стандартных метрик AUC-ROC и AUPRC. Для интерпретации результатов и анализа важности признаков применялся метод SHAP (SHapley Additive exPlanations) из библиотеки shap, позволяющий оценить вклад отдельных признаков в конечный результат прогнозов модели.

3. Результаты

3.1. Метамодель предсказания презентации эпитопа

Точная оценка презентации эпитопа является необходимым предварительным условием для последующего анализа иммуногенности, поскольку пептид может быть распознан иммунной системой только в случае его успешной презентации в комплексе с МНС на поверхности клетки. Процесс презентации антигена включает несколько последовательных внутриклеточных этапов процессинга и для его предсказания была разработана метамодель, объединяющая результаты работы нескольких моделей, каждая из которых предсказывает свой этап процессинга. Краеугольным камнем этой системы является модель глубокого обучения TranspМНС [2], которая оценивает финальный этап процессинга – аффинное связывание пептида с молекулой МНС. Данная модель, обученная на наборе данных из 191975 связанных и 191621 несвязанных пар эпитоп-МНС, продемонстрировала исключительно высокую производительность на тестовой выборке (ROC AUC 99,06%).

Для агрегации результатов моделей, предсказывающих этапы процессинга, и построения итоговой метамодели был выбран алгоритм случайного леса (Random Forest). Его выбор обусловлен лучшими показателями производительности на тестовом наборе данных (AUC 98,66% и F1 96,02%) в сравнении с такими альтернативными подходами, как градиентный бустинг (98,59% и 94,16%) и XGBoost (98,62% и 94,34%).

Разработанная метамодель прошла проверку путём сравнения с широко используемыми и авторитетными программами NetМНСpan [8], МНСflurry [9] и DeepМНСI [4]. Результаты сравнительного анализа, представленные на рис. 1, демонстрируют превосходство подхода на основе метамодели. Следует отметить, что размер тестового набора (N на графике) варьировал для каждой программы, так как из него были исключены все данные, пересекающиеся с их обучающими наборами.

Для сравнительной оценки вклада каждого компонента в итоговый результат был применён метод SHAP, результаты которого представлены на рис. 2. Анализ показал, что модель TranspМНС вносит наибольший вклад в предсказание. Вместе с тем, модели, описывающие другие этапы процессинга, также являются значимыми, повышая общую прогностическую способность. Таким образом, интеграция моделей всех этапов процессинга антигена позволяет получить более полную и надёжную оценку потенциала пептида к презентации, что является критически важным для дальнейшего прогнозирования его иммуногенности.

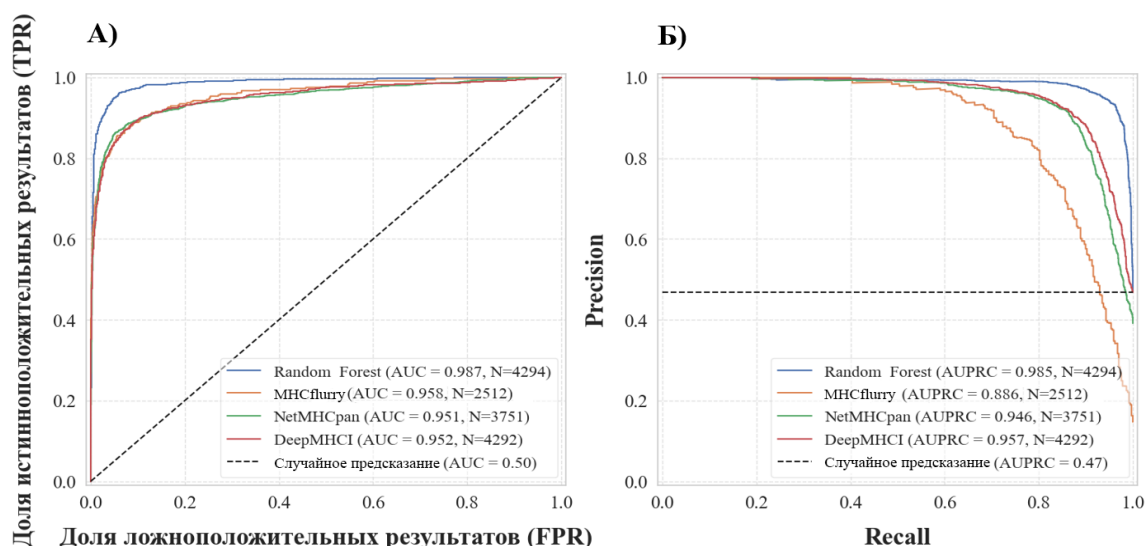


Рис. 1. ROC (А) и PR (Б) кривые сравнения производительности разработанной модели презентации на основе случайного леса с NetMHCpan, mhcflurry и DeepMHC

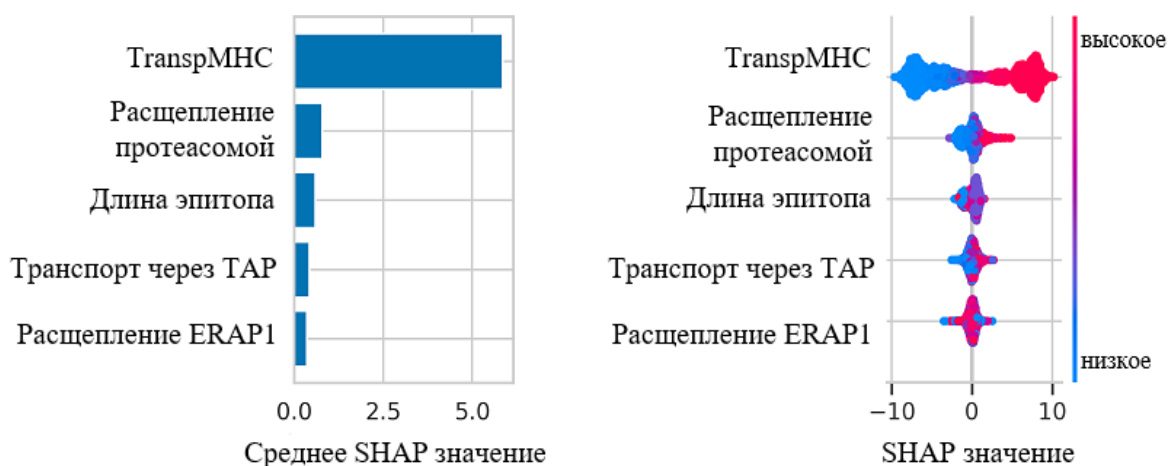


Рис. 2. SHAP-анализ для входных признаков метамодели презентации

3.2. Комплексная метамодель предсказания иммуногенности

Точная оценка презентации эпитопа является необходимым, но недостаточным условием для определения его иммуногенности. Далеко не каждый пептид, представленный на поверхности клетки, способен вызвать иммунный ответ. Поэтому следующим шагом стала разработка отдельной метамодели, нацеленной на предсказание иммуногенности.

Учитывая, что модель TranspMHC показала себя как ключевой компонент в прогнозировании презентации, было принято решение проверить её потенциал и для оценки иммуногенности. Для этого TranspMHC была дополнительно обучена на двух наборах данных: Neo и IE, структура которых представлена на рис. 3. Обе полученные модели продемонстрировали высокую и сопоставимую производительность, с показателем AUC

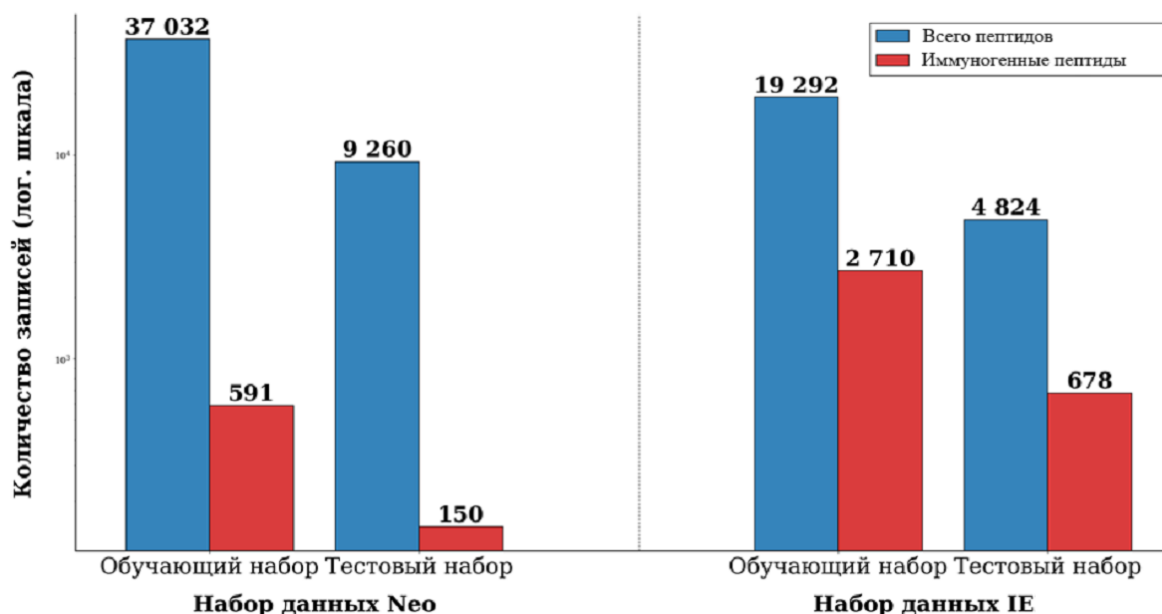


Рис. 3. Структура наборов для предсказания иммуногенности моделью TranspMHC

94,59% на наборе Neo и 94,11% на IE. Однако модель, обученная на данных IE, показала вдвое большую чувствительность (83,43%) по сравнению с Neo (44,67%), что отражает различия в этих наборах данных.

Затем для создания итоговой комплексной метамодели, названной IMM (ImmunoMetaModel), был применён подход, аналогичный построению метамодели презентации. В качестве входных признаков были объединены: предсказание метамодели презентации, предсказания обеих версий TranspMHC (Neo и IE), предсказания моделей, оценивающих секрецию цитокинов, и физико-химические свойства пептидов.

Для определения наиболее значимых признаков был проведён анализ с использованием метода SHAP, результаты которого представлены на рис. 4. Исходя из полученных результатов, наибольший вклад в прогнозирование вносят предсказание метамодели презентации, предсказания обеих моделей TranspMHC, индекс чужеродности пептида, а также прогнозы индукции секреции интерферона-гамма и фактора некроза опухолей альфа.

На основе вклада отобранных признаков было создано три варианта итоговой IMM для определения оптимального соотношения производительности и количества признаков: *imm_plus* – использует все 16 признаков, *imm* – использует топ-6 признаков, *imm_mini* – использует топ-5 признаков без индекса чужеродности. По результатам даже самая компактная модель IMM продемонстрировала высокую производительность – AUC 92,7%, что всего на 2,5% меньше, чем у самой полной версии модели.

Для подтверждения эффективности разработанного подхода все три варианта IMM прошли сравнительный анализ с существующими

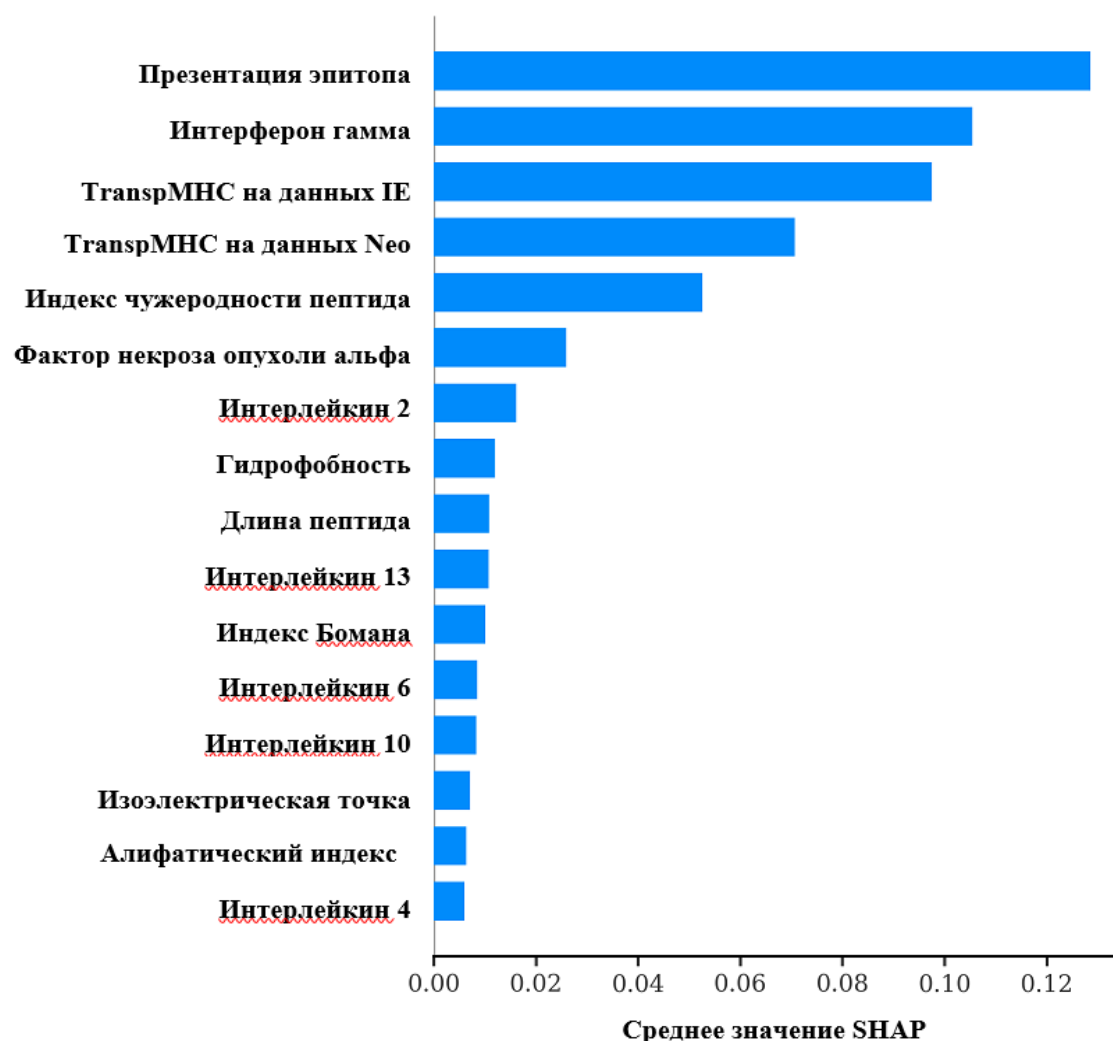


Рис. 4. SHAP-анализ для входных признаков метамодели иммуногенности

программными решениями DeepHLApan [1], DeepImmuno [3] и IEPAPI [6]. Результаты, представленные на рис. 5, демонстрируют значительное превосходство разработанных ИММ. Разница в производительности между imm_plus и моделью IEPAPI составила 21,1% по метрике AUC и 56,1% по AUCPR. Высокий показатель AUCPR особенно важен, так как он свидетельствует о способности метамодели эффективно идентифицировать иммуногенные пептиды даже в условиях сильного дисбаланса классов, что является критически важным для решения задач, связанных с поиском неоантигенов в персонализированной иммунотерапии.

4. Заключение

В ходе работы были созданы две метамодели: для предсказания презентации эпитопа на клеточной поверхности и для определения его иммуногенности. Полученные метамодели были сравнены с известными

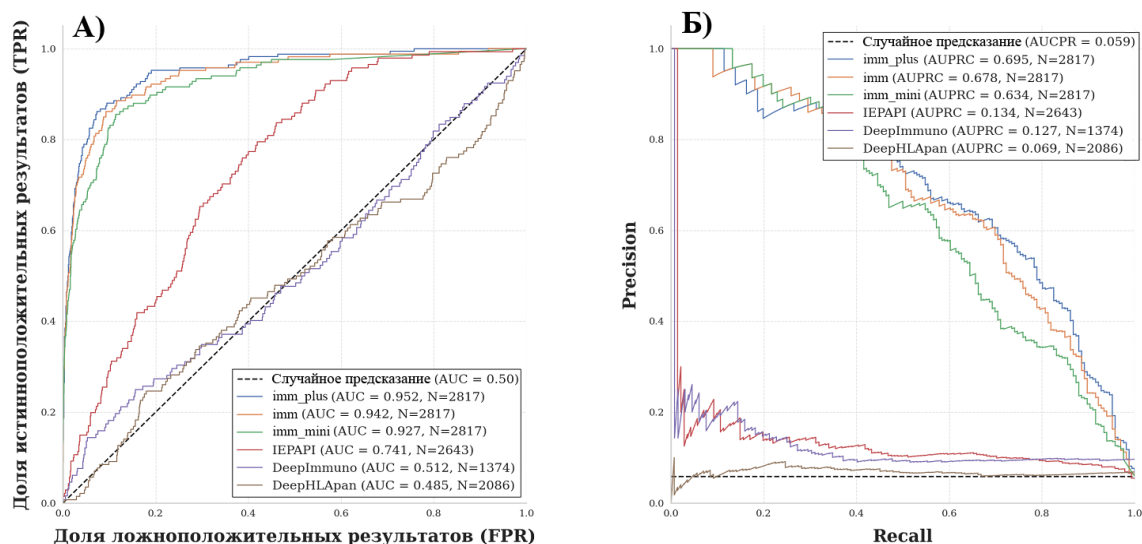


Рис. 5. ROC (А) и PR (Б) кривые сравнения производительности модели иммуногенности на основе случайного леса с IEPAPI, DeepHLApan и DeepImmuno

аналогами. По результатам сравнения на независимых наборах данных обе метамодели продемонстрировали наилучшую производительность в своих категориях, превзойдя ближайших конкурентов в оценке иммуногенности более чем на 21% по метрике AUC и на 56% по AUCPR.

Таким образом, результаты данной работы демонстрируют, что преодоление существующих ограничений в прогнозировании иммуногенности неоантигенов требует перехода от узкоспециализированных моделей к комплексному и многоуровневому анализу. Разработанный подход доказывает, что объединение признаков, связанных с презентацией и узнаванием иммунной системой неоантигена, при помощи метамodelей является оправданным и позволяет достигать высокой предсказательной точности.

Дальнейшее развитие этого подхода может основываться на включении дополнительных факторов, таких как особенности опухолевого микроокружения, что будет способствовать открытию прямых перспектив для ускорения перехода персонализированной иммунотерапии из области научных изысканий в стандартную клиническую практику.

Библиографические ссылки

1. A deep learning approach for neoantigen prediction considering both HLA-peptide binding and immunogenicity / J. Wu [et al.] // Front. Immunol. 2019. Vol. 10. Article no. 2559.
2. Attention-aware differential learning for predicting peptide-MHC class I binding and T cell receptor recognition / R. Niu [et al.] // Brief. Bioinform. 2024. Vol. 26. Article no. bbaf038.
3. DeepImmuno: deep learning-empowered prediction and generation of immunogenic peptides for T-cell immunity / G. Li [et al.] // Brief. Bioinform. 2021. Vol. 22. Article no. bbab160.

4. DeepMHCI: an anchor position-aware deep interaction model for accurate MHC-I peptide binding affinity prediction / W. Qu [et al.] // *Bioinformatics*. 2023. Vol. 39. Article no. btad551.
5. Driver mutation landscape of acute myeloid leukemia provides insights for neoantigen-based immunotherapy / P. Jin [et al.] // *Cancer Lett.* 2025. Vol. 611. Article no. 217427.
6. IEPAPI: a method for immune epitope prediction by incorporating antigen presentation and immunogenicity / J. Deng [et al.] // *Brief. Bioinform.* 2023. Vol. 24. Article no. bbad171.
7. Meta learning for mutant HLA class I epitope immunogenicity prediction to accelerate cancer clinical immunotherapy / L. Xu [et al.] // *Brief. Bioinform.* 2024. Vol. 26. Article no. bbae625.
8. NetMHCpan-4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data / V. Jurtz [et al.] // *J. Immunol. Baltim. Md 1950*. 2017. Vol. 199. P. 3360–3368.
9. O'Donnell T. J., Rubinsteyn A., Laserson U. MHCflurry 2.0: Improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing // *Cell Syst.* 2020. Vol. 11. P. 42–48.
10. Richman L. P., Vonderheide R. H., Rech A. J. Neoantigen dissimilarity to the self-proteome predicts immunogenicity and response to immune checkpoint blockade // *Cell Syst.* 2019. Vol. 9. P. 375–382.

ОЦЕНОЧНЫЕ ФУНКЦИИ МОЛЕКУЛЯРНОГО ДОКИНГА: ТЕСТИРОВАНИЕ И СРАВНИТЕЛЬНЫЙ АНАЛИЗ

К. В. Фурс¹⁾, Я. В. Лайков¹⁾, А. В. Тузиков¹⁾, А. М. Андрианов²⁾

¹⁾ Объединённый институт проблем информатики НАН Беларуси,
Минск, Беларусь, kvfurs@gmail.com

²⁾ Институт биоорганической химии НАН Беларуси,
Минск, Беларусь, alexande.andriano@yandex.ru

На базе набора экспериментальных данных CASF-2016 в молекулярном докинге проведено сравнение оценочных функций AutoDock Vina, NNScore2, RF-Score-4, CENsible, HGScore, OnionNet-2, PIGNet2 и PLANET. Показано, что функции глубокого обучения PLANET и OnionNet-2 обеспечивают наибольшую точность, улучшая прогноз сродства лиганда и достоверность отбора потенциальных ингибиторов.

Ключевые слова: оценочные функции; прогнозирование аффинности связывания; CASF-2016; молекулярный докинг; взаимодействия белок-лиганд; поиск новых лекарств.

SCORING FUNCTIONS OF MOLECULAR DOCKING: TESTING AND COMPARATIVE ANALYSIS

K. V. Furs^{a)}, Y. V. Laikou^{a)}, A. V. Tuzikov^{a)}, A. M. Andrianov^{b)}

^{a)} The United Institute of Informatics Problems,
Minsk, Belarus, kvfurs@gmail.com

^{b)} Institute of Bioorganic Chemistry,
Minsk, Belarus, alexande.andriano@yandex.ru

Based on the benchmark dataset CASF-2016, a comparison of scoring functions in molecular docking was carried out, including AutoDock Vina, NNScore2, RF-Score-4, CENsible, HGScore, OnionNet-2, PIGNet2, and PLANET. The results show that the deep learning functions PLANET and OnionNet-2 provide the highest accuracy, improving ligand affinity prediction and the reliability of selecting potential inhibitors.

Keywords: scoring functions; binding affinity prediction; CASF-2016; molecular docking; protein-ligand interactions; drug discovery.

1. Введение

В последние годы было предложено много новых оценочных функций на основе моделей машинного обучения, предлагающих альтернативы использованию эмпирических или математических функций для предсказания энергии связывания комплексов лиганд-белок [1]. В то же время

точность оценочных функций обычно проверяют на кристаллических комплексах с уже адаптированным активным центром, а в докинге обычно используют жесткий рецептор и гибкий лиганд. Это оставляет открытым вопрос о том, какие из них более целесообразны для практического применения в условиях молекулярного докинга.

В данной работе осуществлен сравнительный анализ точности трех традиционно используемых оценочных функций AutoDock Vina (<https://vina.scripps.edu>), NNScore2 (<https://github.com/durrantlab/nnscore2>) и RF-Score-4 (<https://pjballester.wordpress.com/software/>) с новыми CENsible [2], HGScore [3], OnionNet-2 [4], PIGNet2 [5] и PLANET [6]. Для этой цели использован набор экспериментальных данных CASF-2016 [7], и рассмотрены два варианта тестирования оценочных функций, в одном из которых проведено их сравнение для кристаллических структур комплексов лиганд-белок, а в другом – по данным молекулярного докинга.

2. Материалы и методы

Тестовый набор данных. CASF-2016 представляет собой набор данных для проверки точности оценочных функций, используемых в виртуальном скрининге потенциальных лекарств [7]. Он содержит набор кристаллических структур комплексов белок-лиганд и соответствующие им экспериментальные данные об аффинностях связывания, выраженные в величинах $\log K_a$ (K_a – константа ассоциации). Набор CASF-2016 включает 57 молекулярных мишеней, каждая из которых представлена пятью кристаллическими структурами комплексов белок-лиганд.

Тестирование на основе данных рентгеноструктурного анализа. Подготовку структур белков в формате .pdbqt осуществляли с помощью программного пакета MGLTools 1.5.6 (<https://ccsb.scripps.edu/mgltools/>). В качестве исходного формата лигандов был выбран формат .mol2, который последовательно преобразовывали в представления .sdf, .pdb и .pdbqt с помощью пакетов PyMOL (<https://www.pymol.org>), RDKit (<https://www.rdkit.org>) и MGLTools 1.5.6. PLANET не рассматривалась в первом варианте сравнительного анализа из-за чувствительности к качеству исходных данных в формате .sdf.

Тестирование на основе данных молекулярного докинга. Необходимо учесть, что в кристаллической структуре комплексов активный центр белка-рецептора адаптирован к лиганду, что не соответствует реалиям молекулярного докинга. Поскольку CASF-2016 включает для каждого белка по 5 комплексов с различными лигандами, было решено «перетасовать» структуры белков для всех 57 молекулярных мишеней. Для этого были случайно составлены пары структур одного белка с разными

PDB ID. Вторую структуру белка выравнивали на первую с помощью PyMol align по цепям, контактирующим с лигандом на расстоянии 10 Å. Затем генерировали 3D-конформеры лигандов в RDKit с помощью метода ETKD Gv3 и последующей оптимизацией в силовом поле MMFF до сходимости. При конструировании ячейки для докинга использовали структуру лиганда в кристалле, взятую с отступом в 6 Å по координатам x, y и z. Молекулярный докинг сгенерированных 3D-конформеров с выровненным белком проводили с использованием программы AutoDock Vina 1.2.5 с параметром полноты, равным 100.

3. Результаты и обсуждение

В таблице приведены значения коэффициента корреляции Пирсона (R) и среднеквадратичной ошибки (RMSE) для сравниваемых оценочных функций, рассчитанные между экспериментальными и предсказанными значениями logKa для первого варианта сравнения. Оценочные функции OnionNet-2 и HGScore демонстрируют высокие значения коэффициента корреляции ($R > 0,8$) и низкие величины RMSE ($< 1,5$) (рисунок).

Значения коэффициента корреляции Пирсона и среднеквадратичной ошибки между предсказанными и экспериментальными значениями logKa

Оценочная функция	Структура комплекса в кристалле		Структура комплекса, построенная методом молекулярного докинга	
	Коэффициент корреляции Пирсона (R) ¹	Среднеквадратичная ошибка (RMSE) ¹	Коэффициент корреляции Пирсона (R) ¹	Среднеквадратичная ошибка (RMSE) ¹
AutoDock Vina	0,61	1,86	0,50	1,97
NNScore2	0,68	1,65	0,50	2,25
RF-Score-4 ²	0,75	1,64	0,79	1,73
CENsible	0,71	1,54	0,56	1,91
HGScore	0,81	1,31	0,70	1,57
OnionNet-2	0,83	1,27	0,77	1,48
PIGNet2	0,75	1,68	0,49	2,87
PLANET	—	—	0,76	1,51

Примечания. 1. Жирным шрифтом выделены значения для лучших оценочных функций. 2. Для оценочной функции RF-Score-4 указаны значения, рассчитанные для выборки из 34 комплексов, которые не входили в её обучающую выборку.

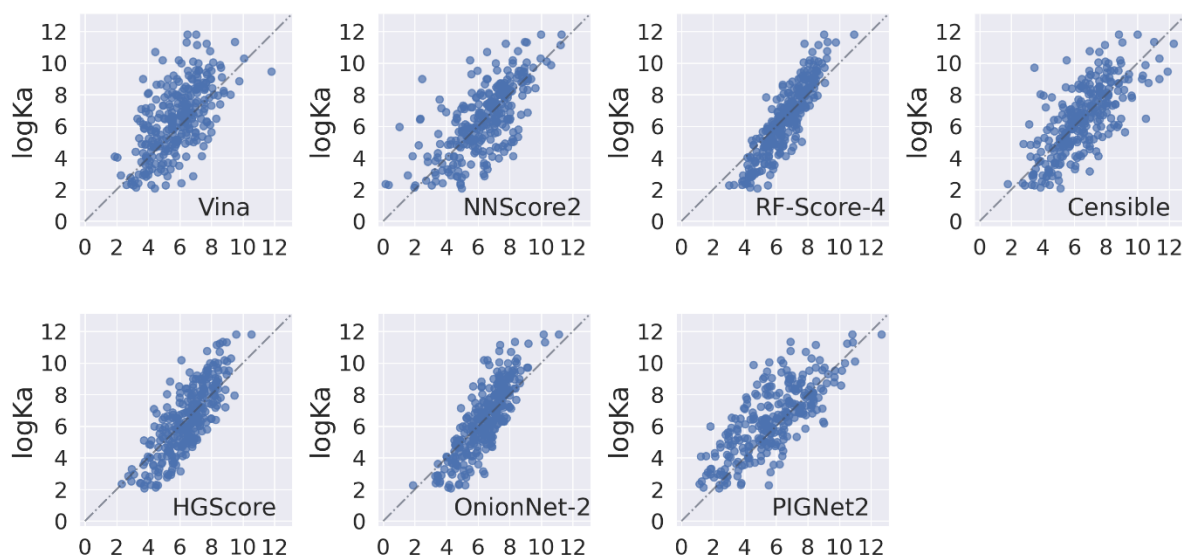


График сравнения предсказанных оценочными функциями (ось x) и экспериментальных (ось y) значений logKa для кристаллических комплексов лиганд-белок

В результате второго варианта тестирования оценочные функции AutoDock Vina, NNScore2, CENSible и PIGNet2 показывают низкую точность предсказания аффинности связывания и умеренную корреляцию с экспериментальными значениями logKa (см. таблицу). OnionNet-2 и PLANET обладают высокой предсказательной эффективностью и высокой корреляцией. HGScore проявляет умеренное снижение точности и имеет сопоставимые характеристики с PLANET и OnionNet-2. Оценочная функция RF-Score-4 демонстрирует высокую корреляцию для комплексов, не входивших в её обучающую выборку, однако размер выборки ($n = 34$) недостаточен для надежной оценки точности RF-Score-4.

4. Заключение

Оценочные функции PLANET и OnionNet-2 демонстрируют наиболее высокую точность, эффективно предсказывая сродство лиганда к белку. Данные оценочные функции могут быть использованы в вычислительных протоколах молекулярного докинга для последующего надежного отбора наиболее вероятных ингибиторов заданной терапевтической мишени.

Работа выполнена при поддержке грантов Белорусского республиканского фонда фундаментальных исследований (проект Ф24КИ-001), Международного научно-технического центра (МНТЦ, проект PR150) и Консорциума и Портала программы «Лекарственно-устойчивый туберкулез» (<https://tbportals.niaid.nih.gov>).

Библиографические ссылки

1. *Meli R., Morris G. M., Biggin P. C.* Scoring functions for protein-ligand binding affinity prediction using structure-based deep learning: a review // *Frontiers in Bioinformatics*. 2022. Vol. 2. Article no. 885983. DOI: [10.3389/fbinf.2022.885983](https://doi.org/10.3389/fbinf.2022.885983).
2. *Bhatt R., Koes D., Durrant J.* CENSible: interpretable insights into small-molecule binding with context explanation networks // *Journal of Chemical Information and Modeling*. 2024. Vol. 64, iss. 12. P. 4651–4660. DOI: [10.1021/acs.jcim.4c00825](https://doi.org/10.1021/acs.jcim.4c00825).
3. Heterogeneous graph convolutional neural network for protein-ligand scoring / K. Crampon [et al.] // *Exploration of Drug Science*. 2023. Vol. 1. P. 126–139. DOI: [10.37349/eds.2023.00010](https://doi.org/10.37349/eds.2023.00010).
4. OnionNet-2: a convolutional neural network model for predicting protein-ligand binding affinity based on residue-atom contacting shells / Z. Wang [et al.] // *Frontiers in Chemistry*. 2021. Vol. 9. Article no. 753002. DOI: [10.3389/fchem.2021.753002](https://doi.org/10.3389/fchem.2021.753002).
5. PIGNet2: a versatile deep learning-based protein-ligand interaction prediction model for binding affinity scoring and virtual screening / S. Moon [et al.] // *Digital Discovery*. 2023. Vol. 3, iss. 2. P. 287–299. DOI: [10.1039/d3dd00149k](https://doi.org/10.1039/d3dd00149k).
6. PLANET: a multi-objective graph neural network model for protein-ligand binding affinity prediction / X. Zhang [et al.] // *Journal of Chemical Information and Modeling*. 2024. Vol. 64, iss. 7. P. 2205–2220. DOI: [10.1021/acs.jcim.3c00253](https://doi.org/10.1021/acs.jcim.3c00253).
7. Comparative Assessment of Scoring Functions: The CASF-2016 Update / M. Su [et al.] // *Journal of Chemical Information and Modeling*. 2019. Vol. 59. P. 895–913. DOI: [10.1021/acs.jcim.8b00545](https://doi.org/10.1021/acs.jcim.8b00545).

ПОИСК ПЕТЛИ В НЕСОВЕРШЕННОМ ПАЛИНДРОМЕ

Г. А. Хазиев, О. А. Зверков, Л. И. Рубанов, А. В. Селиверстов

*Институт проблем передачи информации имени А. А. Харкевича РАН,
Москва, Россия, khaziev@iitp.ru*

Палиндромы часто встречаются при анализе нуклеотидных последовательностей. Вопрос автоматического обнаружения несовершенных палиндромов до сих пор остаётся открытым. Мы предлагаем алгоритм `de_shapker` выделения петли в несовершенном палиндроме. Авторы протестировали работу алгоритма на нескольких множествах высококонсервативных элементов (ВКЭ).

Ключевые слова: несовершенные палиндромы; шпильки; биоинформатика.

SEARCHING FOR THE LOOP IN IMPERFECT PALINDROME

G. A. Khaziev, O. A. Zverkov, L. I. Rubanov, A. V. Seliverstov

*Institute for Information Transmission Problems of the Russian Academy of Sciences
(Kharkevich Institute),
Moscow, Russia, khaziev@iitp.ru*

Palindromes are often encountered in analysis of nucleotide sequences. The question of automatic detection of imperfect palindromes is still open. We propose the `de_shapker` algorithm for identifying a loop in an imperfect palindrome. The authors tested the algorithm on several sets of highly conserved elements (HCE).

Keywords: imperfect palindromes; hairpins; bioinformatics.

1. Введение

Последовательность нуклеотидов называется совершенным палиндромом, если она комплементарна сама себе. Последовательность, отличающаяся от совершенного палиндрома, называется несовершенным палиндромом. Ранее авторами были предложены квадратичный алгоритм `palindrome_self_alignment` поиска наименьшего редакционного расстояния между последовательностью x и совершенным палиндромом, который был получен конкатенацией префикса x и последовательности, комплементарной этому префиксу [1]. Также авторами была предложена функция

$$\text{imp}(x) = \frac{\min \{ \text{dist}(x, wc(w) \mid x = wz \} }{|x|},$$

где $\text{dist}(\cdot)$ – редакционное расстояние, $c(\cdot)$ – комплементарная последовательность, а $|\cdot|$ – длина последовательности. Данная функция характеризует близость последовательности x к совершенному палиндрому. Чем ближе строка к совершенному палиндрому, тем ближе к нулю значение функции $\text{imp}(x)$. Сложность поиска близких к палиндрому последовательностей в молекулярной биологии состоит в наличие длинных подпоследовательностей внутри близких к палиндрому строк, нарушающих общую палиндромность. Для частичного решения данной проблемы, ранее авторами были предложены алгоритмы усечения, целью которых является выделение длинной подстроки x , более близкой к совершенному палиндрому, чем x [2].

2. Алгоритм удаления петли

Авторы предлагают алгоритм `de_sharker`, целью которого является выделение подпоследовательности с более низким значением функции $\text{imp}(x)$ с помощью нахождения некомплементарного себе участка внутри последовательности – петли. В отличие от алгоритмов усечения, которые находили подпоследовательность с более высоким значением $\text{imp}(x)$ с помощью удаления нуклеотидов на краях последовательности, алгоритм `de_sharker` удаляет участок внутри последовательности. Такой участок характерен для несовершенных палиндромов, встречающихся в биологических задачах. Алгоритм получает на вход строку x , значение $\text{imp\_given} = \text{imp}(x)$, матрицу H , вычисленную в алгоритме `palindrome_self_alignment` для строки x , число итераций алгоритма `iteration_counter`, первоначальный размер окна `window_size`, приращение к размеру окна между итерациями `window_delta`, а также стратегию выбора координат начала петли `strategy`. Строка x записывается в строку `result`. Далее, для всех непрерывных подматриц матрицы H размера `window_size` $\times$ `window_size` вычисляется l_1 норма по формуле

$$l_1(A) = \sum_{i=1}^{\text{window\_size}} \sum_{j=1}^{\text{window\_size}} A_{ij},$$

где A – подматрица H . Из индексов (u, v) в матрице H левого верхнего элемента наименьшей по норме подматрицы выбирается индекс s начала найденного участка петли. За выбор индекса отвечает параметр `strategy`. Затем, в x удаляется участок начиная с s , заканчивая $s + \text{window\_size}$ и

записывается в строку x_new . Для полученной строки вычисляется значение функции $\text{imp}(x\_new)$. Если данное значение меньше, чем imp_given , то x_new записывается в строку result и начинается новая итерация, иначе, алгоритм завершает работу и возвращает строку result .

3. Результаты тестирования

В некодировующих областях геномов *Homo sapiens*, *Macaca fascicularis*, *Mus musculus*, *Sus scrofa* были выделены высококонсервативные элементы (ВКЭ), находящиеся на первой хромосоме человека с использованием модификации метода, описанного в [3].

Для двух множеств этих ВКЭ был проведён поиск палиндромов с использованием алгоритмов усечения. Затем, к каждой последовательности была применена функция de_shapker с четырьмя различными значениями параметра strategy : $\max(u, v)$, $\min(u, v)$, $\lfloor \text{mean}(u, v) \rfloor$, $\lceil \text{mean}(u, v) \rceil$. Кроме того, варьировались значения числа итераций: для каждой последовательности выполнялось до 3, 5 и 10 операций. Стартовый размер окна был равен 2, приращение window_delta также был равен 2. Последовательность x считалась близкой к палиндрому, если для одной из её подпоследовательностей x_s , полученной в результате анализа, выполнялось $\text{imp}(x_s) \leq 0.2$.

В первом множестве из 5689 последовательностей - blk01-12.37.8 с помощью усечения был найден 471 палиндром. Количество палиндромов, найденных с помощью функции de_shapker , указано на рис. 1.

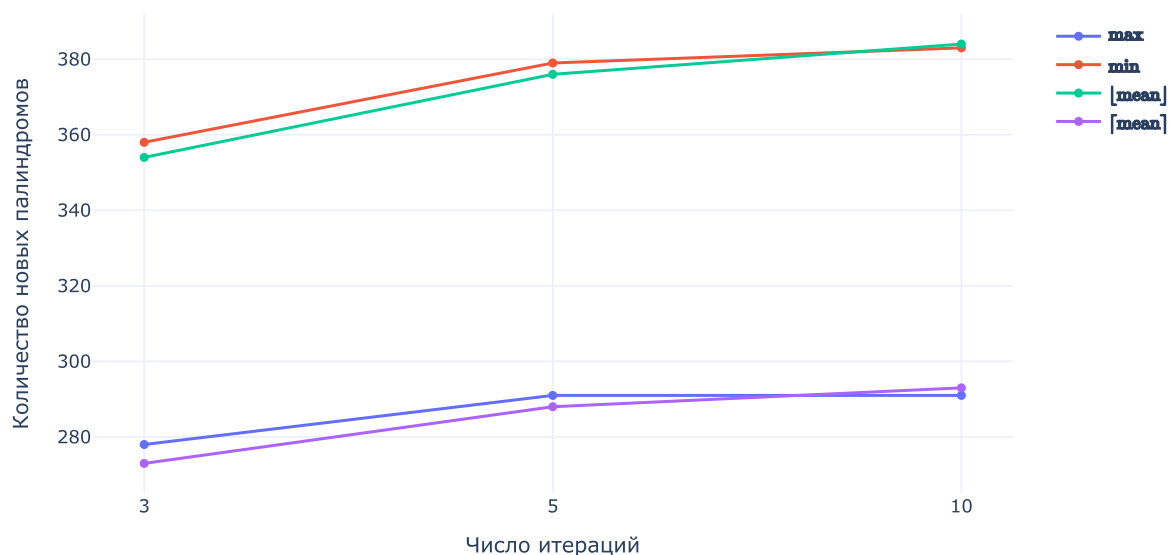


Рис. 1. Зависимость между числом итераций алгоритма de_shapker и количеством найденных палиндромов для множества blk01-12.37.8

Аналогичные эксперименты были проведены для множества blk01-10.39.8. В данном множестве содержится 5862 последовательности, в них с помощью усечения было найдено 475 палиндромов. Количество палиндромов, найденных с помощью функции `de_shapker`, указано на рис. 2.

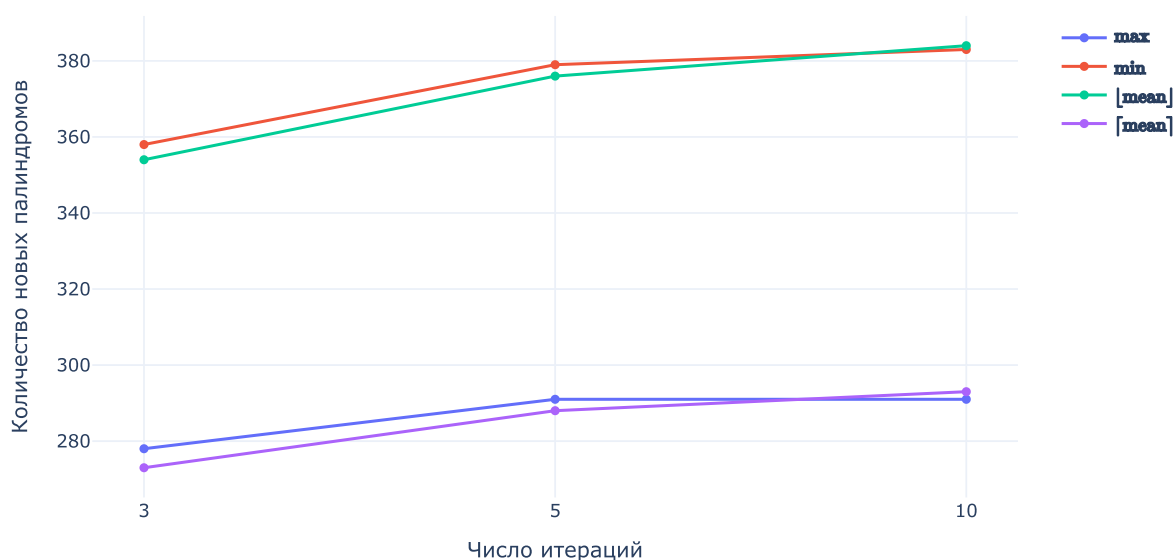


Рис. 2. Зависимость между числом итераций алгоритма `de_shapker` и количеством найденных палиндромов для множества blk01-10.39.8

Из рисунков выше можно сделать два основных вывода. Первый вывод заключается в том, что выбор в качестве параметра `strategy` функции, смещённую к меньшему из значений (u , v) приводит к большему числу обнаруженных палиндромов. Второй вывод состоит в том, что между 3 и 5 итерациями прирост числа палиндромов сильно больше, чем между 5 и 10, следовательно, при анализе большого количества последовательностей можно ограничиться значением `iteration_counter=5` для экономии времени вычисления.

Кроме того, для найденных палиндромов было вычислен процент оставшейся строки после усечения и применения функции `de_shapker` (табл. 1–2). Из таблиц видно, что даже несмотря на небольшое уменьшение средней доли строки после усечения и удаления петли для каждой отдельной стратегии, в среднем, строки потеряли небольшую долю от своей длины. Таким образом, алгоритмы не выдают в результате анализа слишком короткие последовательности, близкие к палиндромам в силу своей длины, а позволяют определить те последовательности, которые действительно могут содержать длинные подпоследовательности, близкие к палиндромам.

Таблица 1

Средняя доля строки после усечения в blk01-12.37.8

strategy	3 итерации	5 итераций	10 итераций
min	81,6%	80,5%	80,5%
max	81,5%	80,5%	80,2%
$\lfloor \text{mean} \rfloor$	81,5%	80,5%	80%
$\lceil \text{mean} \rceil$	81,6%	80,5%	80%

Примечание. Средний процент от первоначальной строки, оставшийся в результате всех преобразований.

Таблица 2

Средняя доля строки после усечения в blk01-10.39.8

strategy	3 итерации	5 итераций	10 итераций
min	81,3%	80,2%	80%
max	81%	80,3%	80%
$\lfloor \text{mean} \rfloor$	81%	80,2%	79,9%
$\lceil \text{mean} \rceil$	81,2%	80,3%	80%

Примечание. Средний процент от первоначальной строки, оставшийся в результате всех преобразований.

4. Заключение

Рассмотренный алгоритм позволяет точнее определять близость строки к совершенному палиндрому. На экспериментальных данных, при выборе стратегии лучше всего себя показали функции минимума и округлённого вниз среднего.

Библиографические ссылки

1. Зверков О. А., Селиверстов А. В., Шиловский Г. А. Выравнивание скрытого палиндрома // Математическая биология и биоинформатика. 2024. Т. 19. № 2. С. 427–438. DOI: [10.17537/2024.19.427](https://doi.org/10.17537/2024.19.427).
2. Khaziev G. A., Seliverstov A. V., Zverkov O. A. Searching for an Imperfect Palindrome // Computer algebra : 6th International Conference Materials, Moscow, June 23–25, 2025 / eds.: A. A. Ryabenko, D. S. Kulyabov. Moscow : RUDN University, 2025. P. 62–65.
3. Новый алгоритм поиска несовершенных палиндромов в ДНК / Г. А. Хазиев [и др.] // MCCMB 2025 : Сборник тезисов 12-й Московской конференции по вычислительной молекулярной биологии (MCCMB). С. 621–624. URL: <https://www.mccmb.info/> (дата обращения: 10.09.2025).

РАЗДЕЛ III

ИНТЕЛЛЕКТУАЛЬНЫЙ И СТАТИСТИЧЕСКИЙ
АНАЛИЗ ДАННЫХ, ПРИНЯТИЕ РЕШЕНИЙ

ПРОГНОЗИРОВАНИЕ ЭФФЕКТИВНОСТИ КАРДИОПРОТЕКТИВНОЙ ТЕРАПИИ У БОЛЬНЫХ РАКОМ МОЛОЧНОЙ ЖЕЛЕЗЫ

М. С. Абрамович¹⁾, Л. В. Гарматная²⁾, Г. С. Дорофеев¹⁾

¹⁾ НИИ прикладных проблем математики и информатики,
Белорусский государственный университет,
Минск, Беларусь, abramovichms@bsu.by, h.lpfrbh@mail.ru

²⁾ Белорусский государственный университет,
Минск, Беларусь, lolita.harmatnaya@gmail.com

Рассматривается проблема классификации больных раком молочной железы с применением и без применения кардиопротективной терапии. Проведена предобработка данных и отобраны информативные признаки для классификации. Для классификации были применены методы машинного обучения и для них получены оценки точности классификации.

Ключевые слова: рак молочной железы; кардиопротективная терапия; методы машинного обучения; эффективность.

THE PREDICTION OF EFFICIENCY OF CARDIOPROTECTIVE THERAPY FOR BREAST CANCER PATIENTS

M. S. Abramovich^{a)}, L. V. Garmatnaya^{b)}, G. S. Dorofeev^{a)}

^{a)} Research Institute for Applied Problems of Mathematics and Informatics,
Belarusian State University,
Minsk, Belarus, abramovichms@bsu.by, h.lpfrbh@mail.ru

^{b)} Belarusian State University,
Minsk, Belarus, lolita.harmatnaya@gmail.com

The article deals with the problem of classification of breast cancer patients with and without using of cardioprotective therapy. Data preprocessing and feature selection were carried out. To perform the task, machine learning methods were applied and their performance was estimated using specific quality metrics.

Keywords: breast cancer; cardioprotective therapy; machine learning methods; efficiency.

1. Введение

В структуре онкологических заболеваний злокачественные новообразования женской репродуктивной системы в Республике Беларусь на протяжении многих лет занимают первое место. В представленном

исследовании рассматривалась проблема классификации больных раком молочной железы пациентов на группы принимающих или не принимающих кардиопротективную терапию в зависимости от изменения фракции выброса левого желудочка. Уменьшение фракции выброса потенциально может привести к развитию сердечной недостаточности [1].

Поэтому существует необходимость в выявлении пациентов с предрасположенностью к сердечно-сосудистым заболеваниям.

До применения классификации пациентов с использованием методов машинного обучения необходимо провести предобработку данных и отбор информативных признаков для классификации.

2. Предобработка данных

У больных раком молочной железы (80 пациентов) исследовалось изменение фракции выброса левого желудочка в зависимости от того применялась кардиопротективная терапия (25 пациентов) или не применялась (55 пациентов). Всего у пациентов были измерены 156 признаков, из них 141 количественных и 14 категориальных.

В качестве признаков были рассмотрены клинико-функциональные показатели, параметры внутрисердечной гемодинамики, антропометрические характеристики, а также генетические полиморфизмы.

На этапе предобработки из всех признаков были исключены 15 «шумовых» признаков, которые не соответствуют цели исследования.

Проведена замена пропущенных значений: пропущенные количественные данные заменялись средним значением, а категориальные наиболее часто используемым элементом. Одновременно исключены ошибочные записи в данных.

Выполнена нормализация данных по методу **минимакса**: линейное преобразование данных в диапазоне от 0 до 1.

3. Отбор информативных признаков

Для построения решающих правил классификации пациенток первоначально среди категориальных и количественных признаков отобраны информативные признаки.

Проблема отбора признаков часто возникает из-за того, что на этапах постановки задачи и формирования данных ещё не ясно, какие признаки бесполезны или дублируют друг друга. Естественное стремление учесть как можно больше потенциально полезной информации приводит к появлению избыточных («шумовых») признаков. Выбор полезной информации (отбор признаков) является важной операцией, поскольку для решения любой классификационной задачи целесообразно отобрать признаки,

несущие не «шум» и не относящуюся к цели исследования, а полезную для данной задачи информацию.

Информативные признаки отбирались отдельно для категориальных и количественных признаков.

Для отбора категориальных признаков применялись фильтрационные критерии хи-квадрат [2] и Крамера [2], а также встроенный алгоритм в методе случайного леса [2]. На рис. 1 представлены категориальные информативные признаки, отобранные в методе случайного леса.

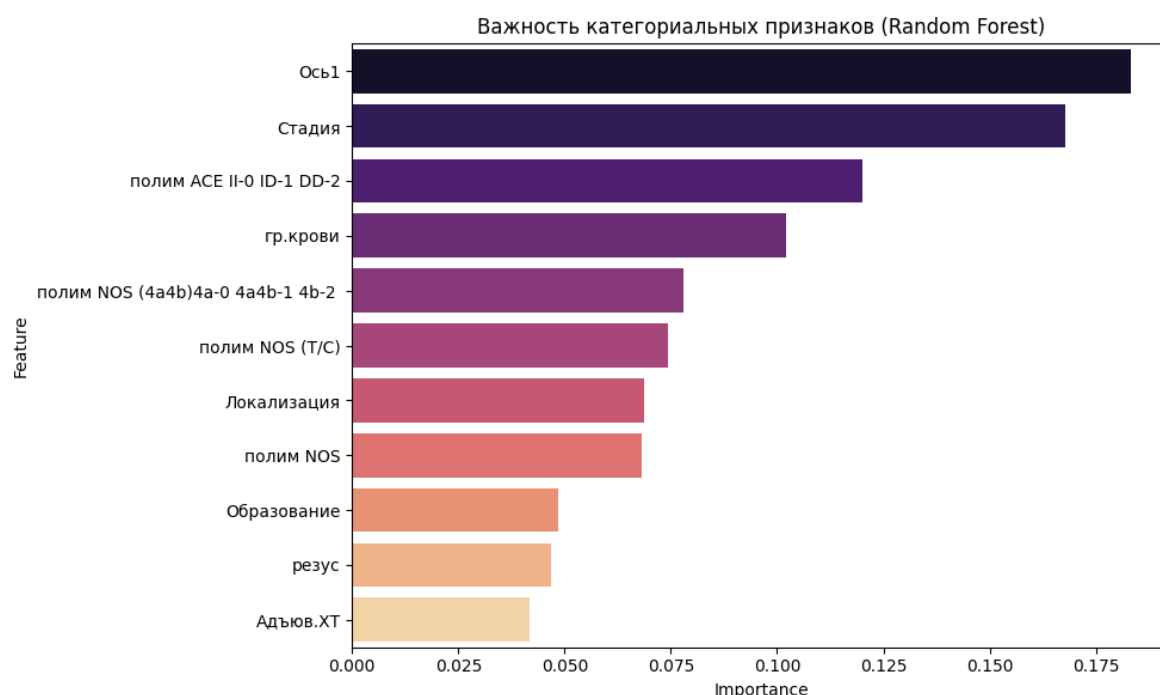


Рис. 1. Категориальные информативные признаки (случайный лес)

Как следует из рис. 1, среди категориальных признаков отобрано 10 информативных.

Для отбора информативных количественных признаков применялся однофакторный дисперсионный анализ (F-критерий). На рис. 2 представлены количественные информативные признаки, отобранные в методе случайного леса.

Как следует из рис. 2, среди количественных признаков отобрано 8 информативных.

4. Классификация пациентов методами машинного обучения

Так как исходная выборка наблюдений состоит из классов, имеющих ограниченный объем (55 и 20 пациентов), то для увеличения объемов классов использовался алгоритм SMOTE [3, 4], предназначенный для

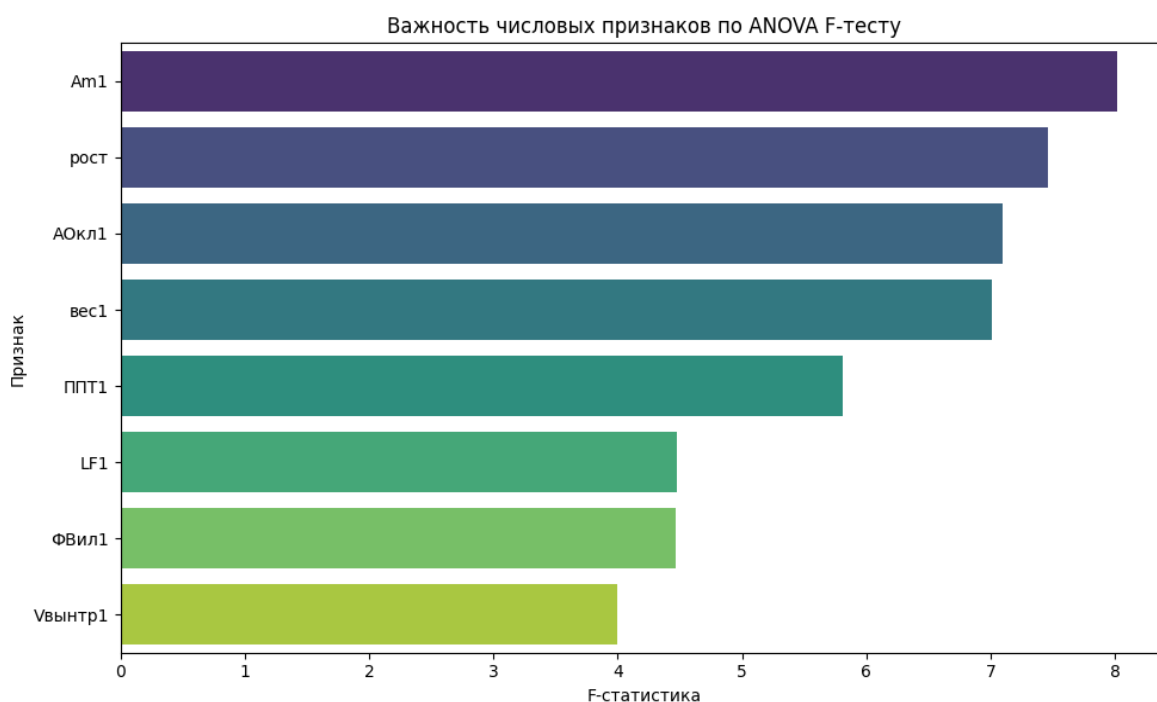


Рис. 2. Количественные информативные признаки (F-критерий)

увеличения числа наблюдений класса меньшего объема. В нашем случае он использовался для увеличения объема обеих классов до 300 наблюдений в каждом, из них 180 наблюдений использовалось для обучающей выборки, 60 – для валидационной и 60 – для экзаменационной.

Для классификации больных раком молочной железы на группы принимающих кардиопротективную терапию и не принимающих ее применялись следующие методы машинного обучения [5]: логистическая регрессия, случайный лес, градиентный бустинг XGBoost и градиентный бустинг CatBoost.

Целевой (группирующий) признак как изменение фракции выброса левого желудочка ФВ создается следующим образом:

$$value = \frac{|[ФВ \text{ без кардиопротективной тер.}] - [ФВ \text{ с кардиопротективной тер.}]|}{data[ФВ \text{ без кардиопротективной тер.}]},$$

где *value* – изменение фракции выброса.

Если $value > 5\%$, то значение группирующего признака полагается равным 1, в противном случае – 0.

В табл. 1 представлены оптимальные гиперпараметры моделей, полученные по валидационной выборке.

Таблица 1

Оптимальные гиперпараметры моделей

Модель	Оптимальные гиперпараметры
Логистическая регрессия	Коэффициент регуляризации: 10
Случайный лес	Количество деревьев: 500, максимальная глубина дерева: 3
Градиентный бустинг XGBoost	Скорость обучения: 0,3, максимальная глубина дерева: 5, количество деревьев: 100
Градиентный бустинг CatBoost	Глубина дерева: 8, количество итераций: 100, скорость обучения: 0,05

Метрики точности классификации, полученные по экзаменационной выборке, представлены в табл. 2.

Таблица 2

Метрики точности классификации моделей

Модель	Метрики точности классификации
Логистическая регрессия	Площадь под ROC-кривой: $0,800 \pm 0,119$, точность: 0,831, полнота: 0,767
Случайный лес	Площадь под ROC-кривой: $0,917 \pm 0,081$, точность: 0,809, полнота: 0,823
Градиентный бустинг XGBoost	Площадь под ROC-кривой: $0,945 \pm 0,072$, точность: 0,905, полнота: 0,903
Градиентный бустинг CatBoost	Площадь под ROC-кривой: $0,928 \pm 0,102$, точность: 0,863, полнота: 0,883

Как следует из табл. 2, наибольшую точность классификации показал градиентный бустинг XGBoost. Высокие значения AUC для этой модели (0,945) свидетельствуют о ее значительном прогностическом потенциале.

5. Заключение

Предложенный в статье подход для прогнозирования больных раком молочной железы, которым показана кардиопротективная терапия, показал высокую эффективность.

Авторы выражают благодарность старшему преподавателю кафедры внутренних болезней и кардиологии БГМУ Н. Б. Конончук за предоставленные для анализа данные.

Библиографические ссылки

1. Применение инновационных технологий в кардиоонкологии / Н. Б. Конончук [и др.] // Вестник Смоленской государственной медицинской академии. 2023. Т. 22, № 2. С. 259–263.
2. *Hastie T., Tibshirani R., Friedman J. H.* The Elements of Statistical Learning. 2nd ed. New York : Springer, 2009.
3. SMOTE: Synthetic Minority Over-sampling Technique / N. Chawla [et al.] // Journal of Artificial Intelligence Research. 2002. № 5. P. 341–378.
4. *Hastie T. H.* Learning from Imbalanced Data // IEEE transactions on knowledge and data engineering. 2009. Vol. 21. № 9. P. 1263–1284.
5. *Harrington P.* Machine Learning in Action. New York : Manning, 2012.

ОПРЕДЕЛЕНИЕ ЗАВИСИМОСТИ ДВУХ СЛУЧАЙНЫХ ВЕЛИЧИН НА ОСНОВЕ СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ

М. С. Баркетов

*Объединённый институт проблем информатики НАН Беларуси,
Минск, Беларусь, barketau@mail.ru*

Рассматривается задача определения зависимости двух ограниченных случайных величин. Дана ограниченная случайная величина X , $x_{\min} \leq X \leq x_{\max}$, и ограниченная случайная величина Y , $y_{\min} \leq Y \leq y_{\max}$. Известно, что есть некоторая функциональная зависимость $Y = f(X) + b^T r$, где r – вектор дополнительных факторов, а b – вектор неизвестных коэффициентов. Делаются дополнительные предположения о зависимости между X и Y . На их основании и на основании статистической выборки значений случайных переменных X и Y для каждого вектора факторов некоторым образом определяется зависимость между X и Y , а также вектор неизвестных коэффициентов b .

Ключевые слова: случайная величина; функциональная зависимость; линейная регрессия; математическая статистика.

OPTIMIZATION OF THE DEPENDANCE OF THE TWO RANDOM VARIABLES BASED ON THE STATISTICAL INFORMATION

M. S. Barketau

*The United Institute of Informatics Problems,
Minsk, Belarus, barketau@mail.ru*

We consider the problem of finding the linear dependence of the two random variables. There is a random variable X , $x_{\min} \leq X \leq x_{\max}$, and there is a random variable Y , $y_{\min} \leq Y \leq y_{\max}$. It is known that there is a functional dependence $Y = f(X) + b^T r$, where r is a vector of additional factors, and b is a vector of unknown coefficients. We make the additional assumptions regarding the dependence between X and Y . Based on these assumptions and based on the statistical values of each variable for each vector of factors we find the functional dependence between X and Y .

Keywords: random variable; functional dependence; linear regression; mathematical statistics.

1. Введение

Рассматривается пара случайных величин X и Y . Известно, что $x_{\min} \leq X \leq x_{\max}$ и $y_{\min} \leq Y \leq y_{\max}$. Для каждого вектора факторов $r_j = (r_1^j, r_2^j, \dots, r_n^j)$, $1 \leq j \leq k$, дана выборка независимых значений переменной X : $x_1^j, x_2^j, \dots, x_{m_j}^j$ и выборка независимых значений переменной Y : $y_1^j, y_2^j, \dots, y_{l_j}^j$. Известно, что переменные X и Y связаны между собой: $Y = f(X) + b^T r$, где r – это вектор факторов, а b – это вектор неизвестных переменных.

Задача состоит в том, чтобы с помощью дополнительных предположений формализовать зависимость между X и Y и определить все неизвестные коэффициенты, определяющие эту зависимость.

Отметим, что можно, например, предполагать линейную зависимость, т.е. $Y = aX + b^T r + W$. a – это некоторый коэффициент, который необходимо определить наряду с вектором коэффициентов b . W – это случайное возмущение с нулевым математическим ожиданием. Однако при таком подходе появляется некоторые трудности с тем, чтобы использовать статистическую выборку из переменных X и Y . Далее предлагается преобразовать эти статистические выборки в гистограммы и строить регрессию вектора гистограммы для переменной X на вектор гистограмму переменной Y .

2. Гистограмма

Сначала построим по выборке независимых значений переменной X , $x_1^j, x_2^j, \dots, x_{m_j}^j$ гистограмму. Для этого разделим интервал $[x_{\min}, x_{\max}]$ на g равных по длине полуинтервалов $[a_1, a_2), [a_2, a_3), \dots, [a_{g-1}, a_g), [a_g, a_{g+1}]$. И пусть n_i^j – это количество значений из выборки, попадающее в i -ый полуинтервал. И пусть $p_i^j = \frac{n_i^j}{\sum_{p=1}^g n_p^j}$ это вероятность попадания в соответствующий интервал.

Аналогичным образом разобьем интервал $[y_{\min}, y_{\max}]$ на s равных полуинтервалов и построим вероятности гистограммы для ряда $y_1^j, y_2^j, \dots, y_{l_j}^j$: q_i^j , $1 \leq i \leq s$.

3. Мультилинейная регрессия

Предполагается мультилинейная зависимость вектора-гистограммы $(q_1^j, q_2^j, \dots, q_s^j)$ от вектора $(p_1^j, p_2^j, \dots, p_g^j, r_1^j, \dots, r_n^j)$.

Пусть Q – это матрица размера $k \times s$, составленная из строк $(q_1^j, q_2^j, \dots, q_s^j)$, $1 \leq j \leq k$, R – это матрица размера $k \times (g + n)$, составленная из строк $(p_1^j, p_2^j, \dots, p_g^j, r_1^j, \dots, r_n^j)$, $1 \leq j \leq k$, а Θ – это матрица коэффициентов размера $(g + n) \times s$.

Тогда мультилинейная зависимость выглядит как

$$Q = R\Theta + W,$$

где W – матрица случайных ошибок с нулевым математическим ожиданием.

Из [2, с. 265, пример 9.6] известно, что можно оценивать матрицу неизвестных коэффициентов как

$$\hat{\Theta} = \arg \min_{\Theta \in R^{(g+n),s}} \left\{ \frac{1}{2k} \|Q - R\Theta\|_F^2 + \lambda_k \sum_{j=1}^{g+n} \|\Theta_{j,\cdot}\|_2 \right\}$$

где $\|\cdot\|_F$ – это норма Фробениуса матрицы, λ_k – нормализующий коэффициент, который можно положить равным $1/k$, а $\Theta_{j,\cdot}$ – j -я строка матрицы Θ .

В общем, это задача выпуклого программирования, подходы к задачам этого класса обсуждаются в книге [3].

Отметим, что возможны более простые оценки матрицы неизвестных коэффициентов на основе нормы l_1 и линейного программирования.

Допустим, мы нашли матрицу $\hat{\Theta}$ и с ее помощью по вектору $(p_1, p_2, \dots, p_g, r_1, \dots, r_n)$ находим вектор $(q_1, q_2, \dots, q_s) = (p_1, p_2, \dots, p_g, r_1, \dots, r_n) \hat{\Theta}$. Спроецируем этот вектор на следующее множество $\{(q_1^*, q_2^*, \dots, q_s^*) | q_i^* \geq 0, 1 \leq i \leq s, \sum_{i=1}^s q_i^* = 1\}$ с помощью следующей выпуклой программы с переменными $q_1^*, q_2^*, \dots, q_s^*$:

$$\begin{aligned} \|(q_1^* - q_1, q_2^* - q_2, \dots, q_s^* - q_s)\|_2 &\rightarrow \min, \\ \sum_{i=1}^s q_i^* &= 1, \\ q_i^* &\geq 0, 1 \leq i \leq s. \end{aligned} \tag{1}$$

Полиномиальные алгоритмы к задачам этого класса обсуждаются в книге [3]. Отметим, что возможна проекция в смысле нормы l_1 , которая осуществляется с помощью линейного программирования.

4. Задача из области планирования прибыли предприятия

Допустим, прибыль предприятия до вложения суммы r в модернизацию является случайной величиной X , а после внедрения новшества, разработанного на данную сумму, является случайной величиной Y .

Тогда рассмотрим следующую статистическую информацию: сумма денег, вложенных в разработку новшества, прибыль предприятия до внедрения новшества за некоторое количество периодов, прибыль предприятия после внедрения новшества за некоторое количество периодов. С помощью вышеприведенного подхода можно предсказать распределение прибыли предприятия после вложения в модернизацию некоторой суммы денег, на основе разработанной мультилинейной регрессии. Далее, возможно подсчитать среднее и дисперсию этого распределения прибыли предприятия после вложения некоторой определенной суммы денег.

5. Заключение

В данной статье разрабатывается подход к формализации и определению зависимости между двумя случайными переменными на основе статистической информации по каждой переменной. Данный подход может быть интересен на производстве для планирования прибыли в зависимости от вложенных денег и для планирования некоторых других показателей производства.

Библиографические ссылки

1. Боровков А. А. Математическая статистика : учебник. 4-е изд., стер. СПб. : Лань, 2010.
2. Wainwright M. J. High-dimensional statistics: A non-asymptotic viewpoint. Cambridge University Press, 2019.
3. Nemirovski A. Modern convex optimization. Department ISYE, Georgia Institute of Technology, 2005.

АНАЛИЗ ГЕНДЕРНЫХ РАЗЛИЧИЙ ПАРАМЕТРОВ ЭКГ НА ОСНОВЕ МЕТОДА МНОГОУРОВНЕВОЙ АМПЛИТУДНОЙ ДИСКРИМИНАЦИИ

К. М. Буданов

*Пензенский государственный университет,
Пенза, Россия, ko13bud@rambler.ru*

Работа посвящена описанию применения метода многоуровневой амплитудной дискриминации сигнала ЭКГ при выявлении различий между записями ЭКГ мужчин и женщин с нормальным сердечным ритмом на основе статистических критериев. Целью исследования является оценка различий распределений сигналов ЭКГ, полученных по гендерному признаку.

Ключевые слова: многоуровневая амплитудная дискриминация; распределение амплитуд; сигнал ЭКГ; статистический критерий.

ANALYSIS OF GENDER DIFFERENCES ECG PARAMETERS BASED ON THE METHOD OF MULTILEVEL AMPLITUDE DISCRIMINATION

K. M. Budanov

*Penza State University,
Penza, Russia, ko13bud@rambler.ru*

The paper describes the application of the method of multilevel amplitude discrimination of the ECG signal in identifying differences between ECG recordings of men and women with normal heart rate based on statistical criteria. The aim of the study is to evaluate the differences in the distribution of ECG signals obtained by gender

Keywords: multilevel amplitude discrimination; amplitude distribution; ECG signal; statistical criterion.

1. Введение

Гендерный признак является одним из факторов, влияющих на показатели электрокардиограммы (ЭКГ), что подтверждают Тегеранское когортное исследование [1], исследование частоты сердечных сокращений у подростков [2], исследование пациентов, принимающих препараты от аритмии [3].

Этиологические причины различий в ЭКГ между гендерными группами до конца не изучены. Однако существуют исследования, показывающие, что на различия в показателях ЭКГ между мужчинами и женщинами могут влиять такие факторы, как половые гормоны [4], строение сердца [5], состав тела [6] и вид сердечной патологии [7, 8].

При обработке сигналов ЭКГ с целью классификации состояний и патологий применяются различные статистические методы анализа [9].

В частности, при исследовании сигналов ЭКГ применяется метод многоуровневой амплитудной дискриминации, позволяющий получить распределение амплитуд сигнала ЭКГ [10].

2. Основная часть

Метод многоуровневой амплитудной дискриминации используется для обработки сигнала ЭКГ, представленного в виде множества числовых значений u_i , соответствующих дискретным моментам времени (отсчетам), где $i = 1, \dots, n$, n – количество отсчетов.

Постановка задачи сравнения сигналов ЭКГ возможна только в том случае, когда эти сигналы сопоставимы. Следовательно, их основные параметры – частота и длительность – должны быть одинаковыми.

Поскольку исходные сигналы различны по амплитуде, требуется процедура преобразования фактических значений по формуле

$$u0_i = \frac{u_i - \min u_i}{\max u_i - \min u_i}, \quad (1)$$

где $u0_i$ – преобразованное значение исходного сигнала; u_i – фактическое значение исходного сигнала; $\min u_i$ – наименьшее значение исходного сигнала, $\max u_i$ – наибольшее значение исходного сигнала.

В результате описанной процедуры диапазоны фактических значений различных сигналов приводятся к одному и тому же интервалу. Для формирования интервального распределения выбирается количество групп l , определяются границы групп и вычисляются групповые абсолютные частоты.

Критерий Пирсона [11] позволяет проверить значимость различий между абсолютными (или относительными) частотами по двум заданным выборкам.

Применение критерия предполагает формулировку нулевой гипотезы об отсутствии различий между двумя заданными выборками. В данном исследовании предполагается нулевая гипотеза об отсутствии различий между усредненными гистограммами распределений сигналов ЭКГ мужчин и женщин.

Далее вычисляется статистика критерия

$$\chi^2 = \sum_{i=1}^l \frac{(f_{m(i)} - f_{w(i)})^2}{f_{w(i)}}, \quad (2)$$

где $f_{m(i)}$ – абсолютные частоты усредненной гистограммы мужчин; $f_{w(i)}$ – абсолютные частоты усредненной гистограммы женщин; l – число интервалов гистограммы.

Расчетное значения критерия сравнивается с критическим, выбранным для принятого уровня значимости критерия α и числа степеней свободы $df = l - 1$, где l – число интервалов гистограммы. Если расчетное значение превысит критическое, то нулевая гипотеза должна быть отвергнута. В противном случае гипотеза об отсутствии различий между гистограммами мужчин и женщин принимается.

Для определения различий между двумя выборочными распределениями можно также воспользоваться критерием Колмогорова-Смирнова. Использование критерия предполагает расчет накопленных относительных частот по каждому из сравниваемых распределений. Расчетное значение критерия по вычисляется формуле:

$$D = \sqrt{\frac{n_m n_w}{n_m + n_w}} \max_i |c_{m(i)} - c_{w(i)}|, \quad (3)$$

где $c_{m(i)}$ – накопленные относительные частоты усредненной гистограммы мужчин; $c_{w(i)}$ – накопленные относительные частоты усредненной гистограммы женщин; n_m – сумма абсолютных частот усредненной гистограммы мужчин; n_w – сумма абсолютных частот усредненной гистограммы женщин.

Полученное расчетное значение критерия сравнивается с критическим, соответствующим выбранному уровню значимости. Если расчетное значение не превышает критического, то на выбранном уровне значимости нулевая гипотеза об отсутствии различий между распределениями сигналов ЭКГ у мужчин и женщин принимается. Если расчетное значение превысит критическое, то нулевая гипотеза должна быть отвергнута.

В ходе экспериментального исследования изучались сигналы ЭКГ, размещённые на Интернет-ресурсе Physionet в базе Mit-bih-normal-sinus-rhythm-database-1.0.0, полученные от людей с нормальным сердечным ритмом. В базу включены сигналы 5 мужчин в возрасте от 26 до 45 лет и 13 женщин в возрасте от 20 до 50 лет, имеющие частоту 128 Гц и измеренные в 2-х отведениях [12]. Возможность проведения сравнительного

анализа обеспечена выбором фрагментов сигналов одинаковой длительности, включающих 38400 отсчетов.

В результате исследования получены амплитудные распределения для каждого сигнала, включающие $l = 32$ интервала. На их основе сформированы усредненные распределения отдельно по мужчинам и женщинам и построены гистограммы этих распределений. Частоты усредненных распределений вычислены как средние значения из частот фактических распределений мужчин и женщин соответственно.

При использовании статистических критериев был выбран уровень значимости $\alpha = 0,05$. Это означает, что статистически достоверные (значимые) различия могут быть обнаружены с вероятностью 0,95.

Графическая интерпретация статистических критериев представлена на рис. 1–4.

Расчетное значение критерия Пирсона по первому отведению составило 17391,8, что превышает критическое значение критерия (45,0) на уровне значимости 0,05 для заданного числа степеней свободы 31.

Расчетное значение критерия Пирсона по второму отведению составило 40120,8, что также превышает критическое значение критерия (45,0) на уровне значимости 0,05 для заданного числа степеней свободы 31.

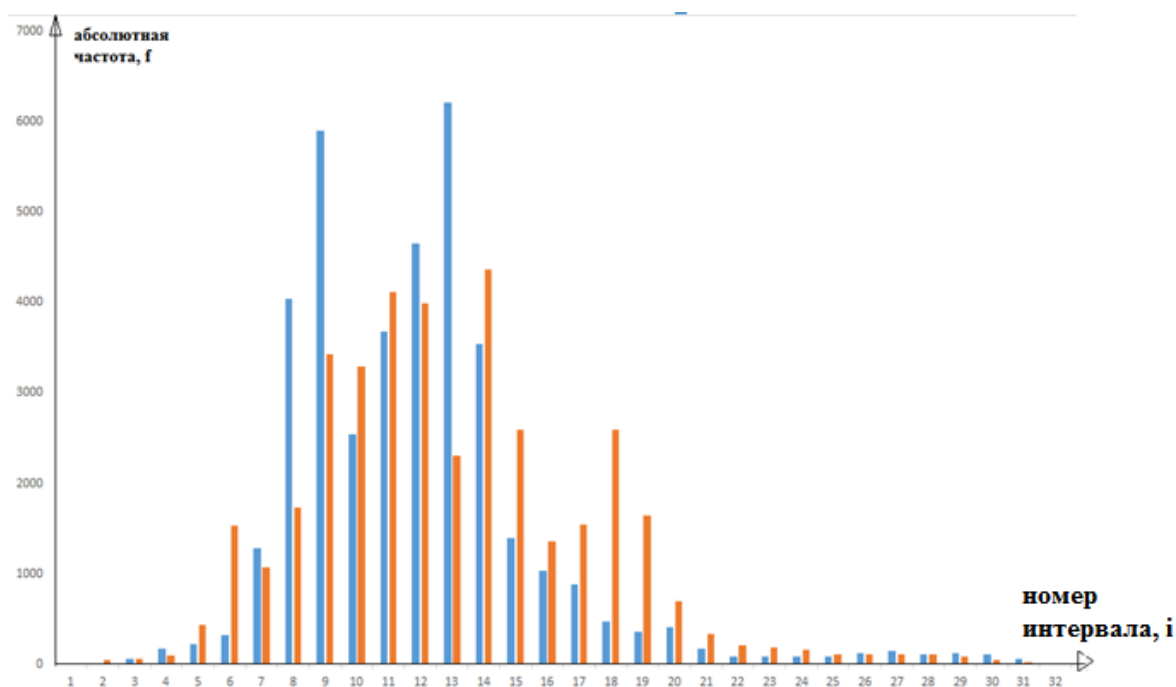


Рис. 1. Гистограммы усредненных распределений сигналов ЭКГ мужчин и женщин в первом отведении (оранжевый – женщины, синий – мужчины)

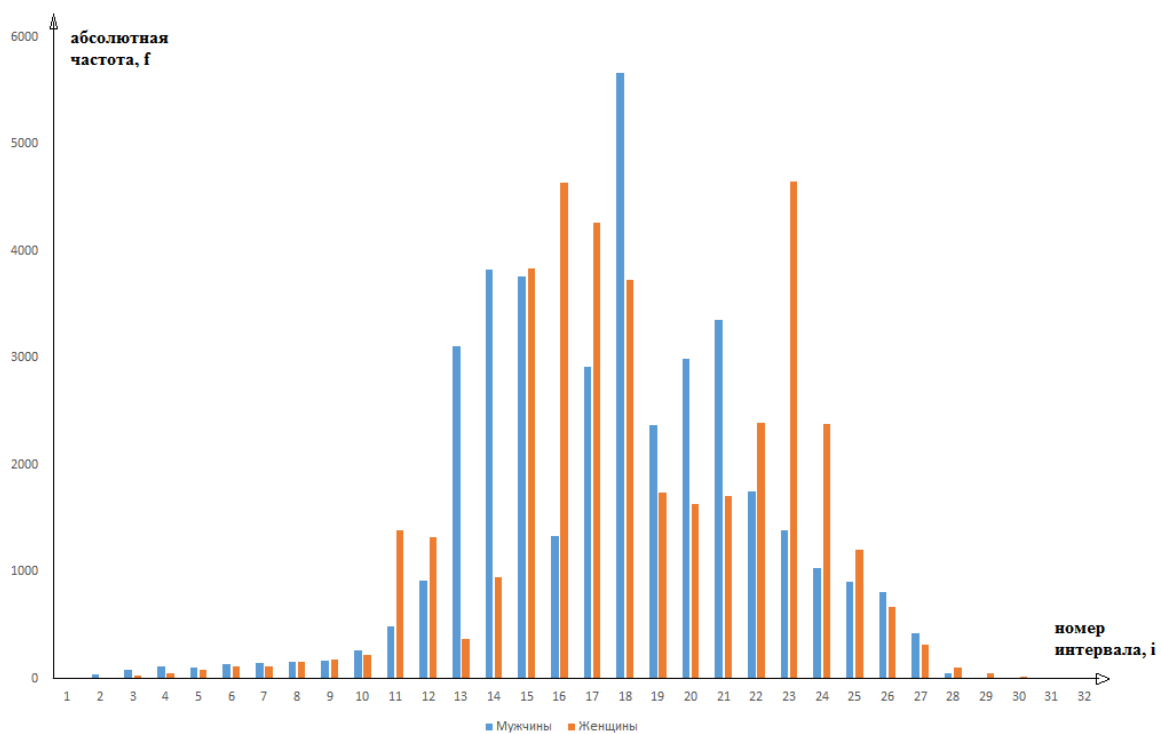


Рис. 2. Гистограммы усредненных распределений сигналов ЭКГ мужчин и женщин во втором отведении (оранжевый – женщины, синий – мужчины)

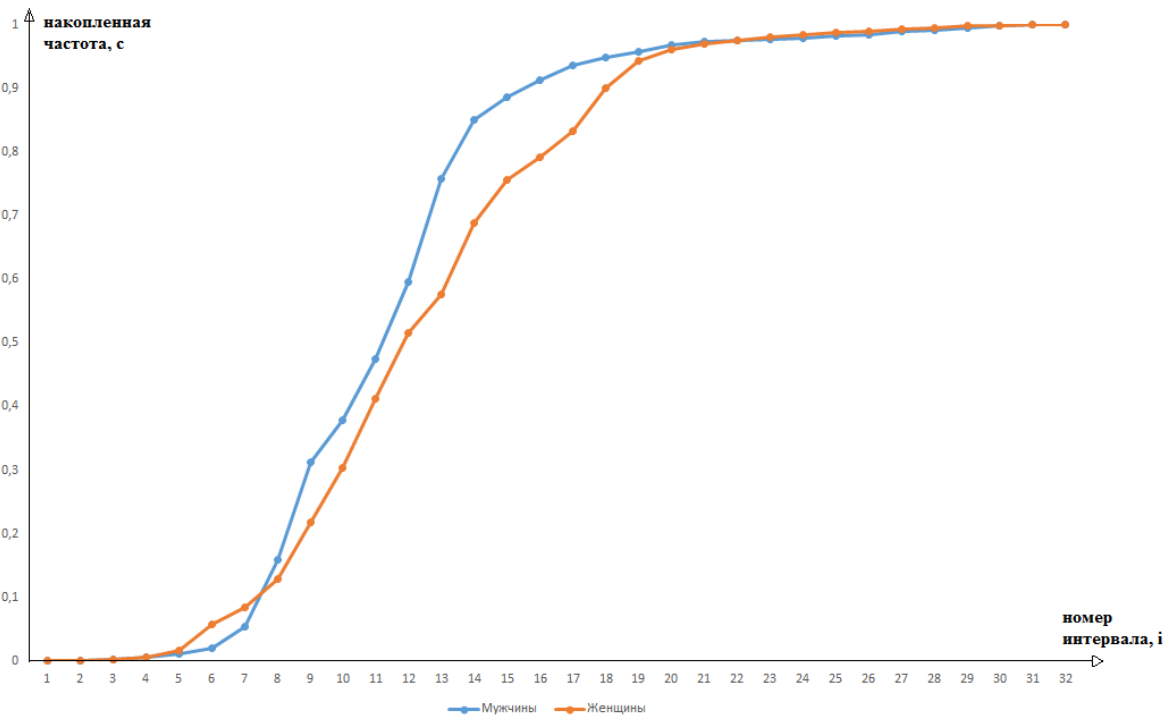


Рис. 3. Графики функций усредненных распределений сигналов ЭКГ мужчин и женщин в первом отведении (оранжевый – женщины, синий – мужчины)

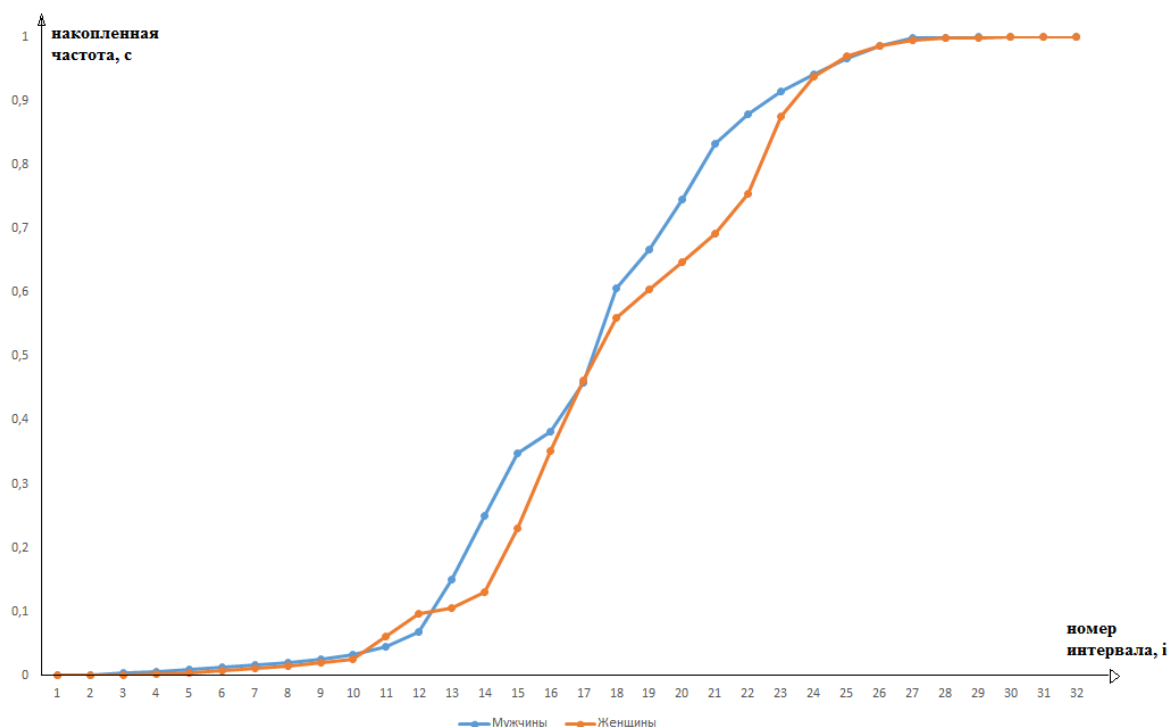


Рис. 4. Графики функций усредненных распределений сигналов ЭКГ мужчин и женщин во втором отведении (оранжевый – женщины, синий – мужчины)

Расчетное значение критерия Колмогорова-Смирнова по первому отведению равно 25,272, что превышает критическое значение критерия (1,385) на уровне значимости 0,05.

Расчетное значение критерия Колмогорова-Смирнова по первому отведению равно 19,530, что превышает критическое значение критерия (1,385) на уровне значимости 0,05.

4. Заключение

Таким образом, метод многоуровневой амплитудной дискриминации подтверждает наличие различий между сигналами ЭКГ у мужчин и женщин с нормальным сердечным ритмом. Результаты применения статистических критериев Пирсона и Колмогорова-Смирнова подтверждают статистическую значимость различий распределений.

При этом, согласно критерию Пирсона, большее различие между сигналами ЭКГ у мужчин и женщин отмечается по второму отведению, критерий Колмогорова-Смирнова показывает наибольшее отличие между сигналами первого отведения.

Следует отметить, что исследование проведено по данным малых выборок, поэтому для получения более репрезентативных результатов требуются дополнительные данные.

Библиографические ссылки

1. Age and gender differences of basic electrocardiographic values and abnormalities in the general adult population; Tehran Cohort Study / P. Ahmadi [et al.] // BMC Cardiovasc Disord. 2023. Vol. 23, Article no. 303.
2. Штина И. Е., Болтачева А. Н., Валина С. Л. Связь антропометрических данных, возраста и пола с показателями электрокардиографии покоя у детей и подростков 6–17 лет // Профилактическая медицина. 2023. Т. 26, № 5. С. 71–77.
3. Колоцей Л. В., Снежицкий В. А. Гендерные особенности электрокардиограммы у пациентов с лекарственно-индуцированным синдромом удлиненного интервала QT на фоне антиаритмической терапии // Известия Национальной академии наук Беларуси. Серия медицинских наук. 2022. Т. 19, № 4. С. 351–363.
4. Effects of sex hormones on ECG parameters and expression of cardiac ion channels in dogs / L. Fülöp [et al.] // Acta Physiol. 2006. Vol. 188, iss. 3–4. P. 163–171.
5. Race-and gender-based differences in cardiac structure and function and risk of heart failure / A. Chandra [et al.] // J Am Coll Cardiol. 2022. Vol. 79, iss. 4. P. 355–368.
6. Different body parts' fat mass and corrected QT interval on the electrocardiogram: The Fasa PERSIAN Cohort Study / M. H. Yazdanpanah [et al.] // BMC Cardiovasc Disord. 2021. Vol. 21. Article no. 277.
7. Wada H., Miyauchi K., Daida H. Gender differences in the clinical features and outcomes of patients with coronary artery disease // Expert Rev Cardiovasc Ther. 2019. Vol. 17, iss. 2. P. 127–133.
8. Gender differences in cardiac hypertrophy / J. Wu [et al.] // J Cardiovasc Transl Res. 2020. Vol. 13. P. 73–84.
9. Modelling and Analysis of Active Biopotential Signals in Healthcare / eds.: B. Varun, G. R. Sinha. IOP Publishing, 2020. Vol. 1. (IPEM–IOP Series in Physics and Engineering in Medicine and Biology.)
10. Способ определения показателей вариативности временных параметров электрокардиографического сигнала: пат. 2827482 РФ / А. В. Кузьмин, В. А. Гасанова, А. В. Пушкарева, Г. К. Олейников, В. А. Баранов ; дата публ.: 27.09.2024.
11. Гланц С. Медико-биологическая статистика. Пер. с англ. М. : Практика, 1998.
12. MIT-BIH Normal Sinus Rhythm Database / A. Goldberger [et al.]. 2000. URL: <https://physionet.org/content/nsrdb/1.0.0/> (date of access: 18.09.2025).

ИНВАРИАНТНОСТЬ СТАЦИОНАРНОГО РАСПРЕДЕЛЕНИЯ ОТКРЫТОЙ СЕТИ ОБСЛУЖИВАНИЯ С ЭКСПОНЕНЦИАЛЬНЫМ ОГРАНИЧЕНИЕМ НА ВРЕМЯ ПРЕБЫВАНИЯ

С. Ю. Евмененко, Ю. В. Малинковский

Гомельский государственный университет им. Ф. Скорины,
Гомель, Беларусь, stas.evmenenko@yandex.ru, Malinkovsky@gsu.by

Рассматривается открытая сеть обслуживания с однолинейными узлами и экспоненциальным ограничением на время пребывания запросов в узлах. Устанавливается инвариантность стационарного распределения по отношению к функциональной форме распределений длительностей обслуживания при фиксированных первых моментах.

Ключевые слова: теория вероятностей; теория массового обслуживания; сети массового обслуживания; стохастические системы; марковские процессы.

INVARIANCE OF THE STATIONARY DISTRIBUTION OF AN OPEN QUEUEING NETWORK WITH EXPONENTIAL LIMIT ON SOJOURN TIME

S. Yu. Evmenenko, Yu. V. Malinkovskiy

Gomel State University named after F. Skorina,
Gomel, Belarus, stas.evmenenko@yandex.ru, Malinkovsky@gsu.by

An open queueing network with single-server nodes and exponential constraints on request sojourn times in the nodes is considered. The invariance of the stationary distribution with respect to the functional form of service time distributions is established for fixed first moments.

Keywords: probability theory; queueing theory; queueing networks; stochastic systems; Markov processes.

1. Введение

Важную роль при исследовании сетей массового обслуживания играет доказательство инвариантности стационарного распределения вероятностей состояний по отношению к форме распределения длительностей обслуживания запросов в узлах. Это связано с тем, что на практике

распределение продолжительности обслуживания обычно отличается от показательного.

В ряде работ, например в [1, 2], инвариантность установлена в предположении, что распределения длительностей обслуживания имеют рациональные преобразования Лапласа-Стилтьеса. В [3] была доказана инвариантность стационарного распределения для открытых и замкнутых сетей обслуживания с обходами узлов запросами. Результаты по инвариантности стационарного распределения сетей при более общих предположениях получены, например, в [4–7]. В [8, 9] показаны примеры применения аппарата сетей обслуживания к современным инфокоммуникационным системам и сетям. В [10, 11] была обоснована необходимость исследования сетей обслуживания, в которых запрос может покинуть узел, не дожидаясь окончания обслуживания. Однако в указанных работах считалось, что продолжительности обслуживания запросов имеют показательное распределение.

В настоящей работе рассматривается сеть обслуживания с экспоненциальным ограничением на время пребывания запросов в узлах, в которой продолжительности обслуживания имеют произвольный закон распределения, а дисциплина обслуживания запросов в узлах имеет следующую специфику. Запрос, поступающий в некоторый узел, обладает абсолютным приоритетом по отношению ко всем остальным запросам в узле, а вытесненный с прибора запрос при повторном его попадании на прибор дообслуживается оставшееся время. Доказывается, что стационарное распределение вероятностей состояний указанных сетей не зависит от функционального вида законов распределения длительностей обслуживания запросов в узлах при фиксированных первых моментах.

2. Марковский случай

В сеть массового обслуживания, состоящую из N однолинейных узлов (систем), поступает стационарный пуассоновский поток запросов с интенсивностью λ . Каждый поступающий запрос независимо от других запросов с вероятностью p_{0i} направляется в i -й узел ($i = \overline{1, N}$, $\sum_{i=1}^N p_{0i} = 1$). Число мест для ожидания в узле неограниченно.

Состояние сети будем описывать вектором $n(t) = (n_1(t), \dots, n_N(t))$, где $n_i(t)$ – число запросов в i -м узле в момент времени t .

Время пребывания запроса в i -м узле является случайной величиной, условное распределение которой (если в i -м узле находится n_i запросов) показательное с параметром v_i/n_i . Другими словами, условная вероятность того, что пребывание каждого запроса в i -м узле закончится в

промежутке $[t, t + h)$, если в момент t в узле находилось n_i запросов, равна $\frac{v_i}{n_i}h + o(h)$ при $h \rightarrow 0$, а условная вероятность завершения пребывания хотя бы одного из этих запросов равна $v_i h + o(h)$. Запрос, обслуженный в i -м узле, мгновенно и независимо от других запросов с вероятностью p_{ij} направляется в j -й узел сети, а с вероятностью p_{i0} покидает сеть $\left(\sum_{j=0}^N p_{ij} = 1\right)$. Запрос, время пребывания которого в i -м узле закончилось, мгновенно и независимо от других запросов с вероятностью r_{ij} направляется в j -й узел сети, а с вероятностью r_{i0} покидает сеть $\left(\sum_{j=0}^N r_{ij} = 1\right)$. Матрицу маршрутизации запросов введем аналогично с [10].

Пусть длительности обслуживания запросов в узлах имеют показательное распределение. Тогда $n(t)$ – однородный марковский процесс с непрерывным временем и не более чем счетным фазовым пространством состояний.

В [10] для сети с ограниченным временем пребывания установлено, что при выполнении условия $\frac{\lambda \varepsilon_i}{v_i + \mu_i} < 1$, $i = \overline{1, N}$, цепь Маркова, описывающая число запросов в сети в момент времени t , эргодична, а ее стационарное распределение имеет вид $p(n) = p_1(n_1) \dots p_N(n_N)$ с множителями

$$p_i(n_i) = \left(\frac{\lambda \varepsilon_i}{v_i + \mu_i} \right)^{n_i} \left(1 - \frac{\lambda \varepsilon_i}{v_i + \mu_i} \right).$$

3. Немарковский случай

Теперь будем предполагать, что времена обслуживания запросов в узлах не зависят от процесса поступления и друг от друга и имеют произвольную функцию распределения $B_i(t)$, причем математическое ожидание фиксированно равенством

$$(\mu_i)^{-1} = \int_0^{\infty} [1 - B_i(t)] dt,$$

где μ_i – скорость обслуживания запросов занятым i -м узлом.

Если запрос поступает в узел, свободный от запросов, он сразу начинает обслуживаться. Если запрос поступает в узел, в котором уже есть запрос, то он выбивает запрос, находящийся на приборе, и сразу же начинает обслуживаться, а выбитый с прибора запрос становится в начало очереди (дисциплина LCFS Preemptive Resume).

В этом случае процесс $n(t)$, вообще говоря, не является марковским. Тем не менее для него справедлива

Теорема 1. Случайный процесс, описывающий число запросов в сети в момент времени t , эргодичен, если выполнено условие $\frac{\lambda \varepsilon_i}{\nu_i + \mu_i} < 1$, $i = \overline{1, N}$, а его финальное распределение имеет вид $p(n) = p_1(n_1) \dots p_N(n_N)$ с множителями

$$p_i(n_i) = \left(\frac{\lambda \varepsilon_i}{\nu_i + \mu_i} \right)^{n_i} \left(1 - \frac{\lambda \varepsilon_i}{\nu_i + \mu_i} \right).$$

Доказательство. Пусть $\tau_{i,k}(t)$ – остаточное время обслуживания запроса в i -м узле с момента t до момента окончания времени обслуживания, а $\tau_i(t) = (\tau_{i,1}(t), \tau_{i,2}(t), \dots, \tau_{i,n_i}(t))$ – вектор, описывающий остаточное время пребывания запросов в i -м узле; где k – номер позиции, на которой находится запрос от «хвоста» к прибору. Поскольку, вообще говоря, $n(t)$ не является марковским процессом, рассмотрим марковский процесс $\zeta(t) = (n(t), \tau(t))$, добавляя к $n(t)$ непрерывную компоненту $\tau(t) = (\tau_1(t); \dots; \tau_N(t))$. Пусть выполнено условие $\frac{\lambda \varepsilon_i}{\nu_i + \mu_i} < 1$, $i = \overline{1, N}$, т.е. в случае, когда $n(t)$ – марковский процесс, существует стационарное эргодическое распределение $n(t)$, а, следовательно, в общем случае и процесса $\zeta(t)$, так как $\zeta(t)$ получается из $n(t)$ добавлением непрерывных компонент. Доказательство этого факта может быть проведено либо по схеме, предложенной в [12]. Введем обозначение

$$\begin{aligned} F(n, x) &= F\left(n, x_{1,1}, \dots, x_{1,n_1}, x_{2,1}, \dots, x_{2,n_2}, \dots, x_{N,1}, \dots, x_{N,n_N}\right) = \\ &= \lim_{t \rightarrow \infty} P\left\{n(t) = n, \tau_{i,1}(t) < x_{i,1}, \dots, \tau_{i,n_i}(t) < x_{i,n_i}, i = \overline{1, N}\right\}. \end{aligned}$$

Введем обозначение: $[\tilde{x}_i]$ – вектор, все элементы которого совпадают с элементами вектора $x_1, \dots, x_N$, а на месте i -го элемента находится элемент $\tilde{x}_i$. Для $F(n, x)$ справедлива следующая система дифференциально-разностных уравнений:

$$\begin{aligned}
& \sum_{i=1}^N \lambda p_{0i} F(n, x) + \sum_{i=1}^N F(n, x) B_i(x_{i,n_i}) v_i = \\
& = \sum_{i=1}^N F(n - e_i, [x_{i,1}, \dots, x_{i,n_i-1}] I_{\{n_i \neq 0\}}) B_i(x_{i,n_i}) \lambda p_{0i} + \\
& + \sum_{i=1}^N \frac{\partial F(n, [x_{i,1}, \dots, x_{i,n_i-1}, 0])}{\partial x_{i,n_i}} B_i(x_{i,n_i}) p_{ii} + \sum_{i=1}^N F(n, x) B_i(x_{i,n_i}) v_i r_{ii} + \\
& + \sum_{i=1}^N \frac{\partial F(n + e_i, [x_{i,1}, \dots, x_{i,n_i}, 0])}{\partial x_{i,n_i+1}} p_{i0} + \sum_{i=1}^N F(n + e_i, [x_{i,1}, \dots, x_{i,n_i+1}]) v_i r_{i0} + \\
& + \sum_{i=1}^N \frac{\partial F(n, x)}{\partial x_{i,n_i}} - \sum_{i=1}^N \frac{\partial F(n, [x_{i,1}, \dots, x_{i,n_i-1}, 0])}{\partial x_{i,n_i}} + \\
& + \sum_{j=1}^N \sum_{i=1, i \neq j}^N \frac{\partial F(n + e_j - e_i, [x_{j,1}, \dots, x_{j,n_j}, 0])}{\partial x_{j,n_j+1}} B_i(x_{i,n_i}) p_{ji} + \\
& + \sum_{j=1}^N \sum_{i=1, i \neq j}^N F(n + e_j - e_i, [x_{j,1}, \dots, x_{j,n_j+1}]) v_j r_{ji}.
\end{aligned}$$

Нетрудно убедиться, что неотрицательным абсолютно непрерывным по x решением является

$$F(n, x) = p(n) \prod_{i=1}^N \prod_{k=1}^{n_i} \mu_i \int_0^{x_{i,k}} [1 - B_i(u)] du,$$

где $p(n)$ – стационарная вероятность состояния n в процессе $n(t)$ в марковском случае. Действительно, подставляя выражение для $F(n, x)$ в уравнение глобального равновесия для произвольной функции распределения $B_i(t)$ времен обслуживания запросов в узлах, получим уравнение глобального равновесия для марковского случая из [10]. Таким образом, теорема 1 доказана.

4. Заключение

В статье для сети массового обслуживания с однолинейными узлами и экспоненциальным ограничением на время пребывания запросов в узлах доказана нечувствительность стационарного распределения по отношению к распределениям длительностей обслуживания запросов при фиксированных первых моментах этих распределений.

Полученные результаты могут применяться при проектировании новых и модернизации уже существующих информационно-вычислительных сетей и сетей передачи данных.

Библиографические ссылки

1. Open, closed and mixed networks of queues with different classes of customers / F. E. Baskett [et al.] // J. Assoc. Comput. Mach. 1975. Vol. 22., no. 2. P. 248–260.
2. Towsley D. Queueing network models with state-dependent routing // J. Assoc. Comput. Mach. 1980. Vol. 27, no. 2. P. 323–337.
3. Малинковский Ю. В. Инвариантность стационарного распределения состояний модифицированных сетей Джексона и Гордона–Ньюэлла // АиТ. 1998. № 9. С. 29–35.
4. Ивницкий В. А. Об условии инвариантности стационарных вероятностей для сетей массового обслуживания // Теория вероятностей и ее применения. 1982. Т. 27, № 1. С. 188–192.
5. Barbour A. D. Networks of queues and the method of stages // Adv. Appl. Probab. 1976. Vol. 8., no. 3. P. 584–591.
6. Chandy K. M., Howard J. H. Jr., Towsley D. F. Product-form and local balance in queueing networks // J. Assoc. Comput. Mach. 1977. Vol. 24, no. 2. P. 250–263.
7. Samelson C. L., Bulgren W. G. A note on product-form solution for queueing networks with Poisson arrivals and general service-time distributions with finite means // J. Assoc. Comput. Mach. 1982. Vol. 29, no. 3. P. 830–840.
8. Скоба А. Н., Состина Е. В. Применение аппарата сетей массового обслуживания для аналитико-численного моделирования работы информационной системы без учета влияния блокировок // Инженерный вестник Дона. 2015. Вып. 3. URL: <http://ivdon.ru/ru/magazine/archive/n3y2015/3130> (дата обращения: 31.08.2025).
9. Кузнецов Н. А., Семенухин К. В. Анализ и оптимизация управляемой модели замкнутой сети массового обслуживания // АиТ. 2020. № 3. С. 67–85.
10. Малинковский Ю. В. Сети Джексона с однолинейными узлами и ограниченным временем пребывания или ожидания // АиТ. 2015. № 4. С. 67–79.
11. Малинковский Ю. В. Стационарное распределение вероятностей состояний G-сетей с ограниченным временем пребывания // АиТ. 2017. № 10. С. 155–167.
12. Коваленко И. Н. Об условии независимости вероятностей состояний системы от вида закона распределения времени обслуживания // Проблемы передачи информации. 1962. Вып. 11. С. 147–151.
13. Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. М.: Наука, 1966.

РАСПОЗНАВАНИЕ ИМЕНОВАННЫХ СУЩНОСТЕЙ КАТЕГОРИИ «ПРОФЕССИЯ» В РУССКОЯЗЫЧНЫХ ТЕКСТАХ

М. Д. Иванова

*Белорусский государственный университет,
Минск, Беларусь, ivanova.mary76@gmail.com*

В работе представлен сравнительный анализ эффективности методов распознавания именованных сущностей категории «Профессия» в русскоязычных текстах. Исследование охватывает словарные подходы, статистические модели (CRF) и современные нейросетевые архитектуры на основе трансформеров (RuBERT). Эксперименты выполнены на специально разработанном размеченном корпусе текстов. Результаты работы применимы для анализа рынка труда и мониторинга вакансий.

Ключевые слова: распознавание именованных сущностей; извлечение профессий; обработка естественного языка; интеллектуальный анализ данных; подготовка данных.

NAMED ENTITY RECOGNITION OF THE “PROFESSION” CATEGORY IN RUSSIAN-LANGUAGE TEXTS

M. D. Ivanova

*Belarusian State University,
Minsk, Belarus, ivanova.mary76@gmail.com*

The paper presents a comparative analysis of the effectiveness of methods for named entity recognition of the “Profession” category in Russian-language texts. The study covers dictionary-based approaches, statistical models (CRF) and modern neural architectures based on transformers (RuBERT). Experiments were conducted on a specially developed annotated text corpus. The work results are applicable to labour market analysis and job vacancy monitoring.

Keywords: named entity recognition; profession extraction; natural language processing; data mining; data preparation.

1. Введение

Распознавание именованных сущностей (Named Entity Recognition, NER) является одной из фундаментальных и критически важных задач в области обработки естественного языка (Natural Language Processing, NLP). Её цель — автоматическое выявление в неструктурированном тексте именованных сущностей и их последующая классификация в заранее

заданные семантические классы, такие как имена людей, названия организаций, географические объекты, денежные выражения, даты и многие другие специфические категории. Выступая краеугольным камнем в процессе извлечения информации, NER находит применение в широком спектре прикладных задач: от совершенствования информационного поиска и машинного перевода до автоматического реферирования и вопросно-ответных платформ.

Актуальность разработки эффективных методов и алгоритмов NER многократно возросла в последнее десятилетие на фоне стремительного роста объёмов цифровых текстовых данных. Это стимулировало активное развитие доменно-ориентированных приложений, где требуется извлечение сущностей не только из универсальных текстов (например, из новостных лент), но и узкоспециализированных источников, характерных для конкретной предметной области. В таких условиях подходы на основе NER становятся необходимым инструментом, позволяющим автоматизировать сложные процессы анализа больших данных, неподвластные ручной обработке.

В рамках настоящего исследования особое внимание уделяется задаче идентификации сущностей категории «Профессия». Извлечение профессий представляет значительный практический интерес для решения ряда актуальных задач:

- автоматическое извлечение информации из текстов вакансий и резюме с целью выявления профессиональных трендов, анализа требований к навыкам и уровням компетенций;
- мониторинг социальных сетей и онлайн-чатов для выявления потребностей рынка труда;
- построение рекомендательных систем для профориентации и карьерного консультирования.

Несмотря на существование множества подходов к NER — от основанных на жёстких лингвистических правилах до нейросетевых моделей — вопрос их сравнительной эффективности остаётся открытым. Это применительно и к распознаванию профессий. Таким образом, целью работы является проведение комплексного сравнительного анализа различных методов распознавания профессий. Данный анализ охватывает как традиционные словарные методы, так и современные глубокие нейронные сети, включая предобученные языковые модели на основе архитектуры трансформеров. Результаты исследования позволяют определить наиболее оптимальный подход, обеспечивающий высокую точность распознавания профессий и устойчивость к разнообразию языковых конструкций в реальных текстах.

2. Данные

Описание корпуса данных. Для проведения сравнительного анализа был целенаправленно создан специализированный корпус текстов, содержащий упоминания профессий. Данный набор текстов включает следующие два ключевых компонента.

1. Синтетические текстовые примеры, искусственно сконструированные для обеспечения широкого охвата связанной с профессиями лексики в различных грамматических формах и синтаксических контекстах.
2. Тексты из открытых источников, репрезентирующие реальную языковую практику. Выборка производилась с онлайн-платформ по поиску работы, из объявлений о вакансиях и резюме.

Для обеспечения высокого качества разметки был применён комбинированный полуавтоматический подход. Первичная обработка осуществлялась с использованием специализированных программных инструментов для предварительного аннотирования. После этого проводилась обязательная ручная верификация и коррекция для устранения неточностей и обеспечения консистентности аннотаций.

Схема разметки данных. Эффективность и точность распознавания сущностей в значительной степени зависит от выбранной схемы разметки данных. Схема разметки данных устанавливает формальные правила идентификации границ сущности и её принадлежности к целевому семантическому классу, что является критически важным для последующего обучения моделей. Различные стратегии аннотации позволяют по-разному обрабатывать многословные сущности и разграничивать их от фоновых слов [3].

В данной работе была применена широко распространенная и зарекомендовавшая себя схема BIO (Begin, Inside, Outside). В рамках данной схемы каждый токен в тексте получает один из трёх тегов:

- B (Begin) — начало именованной сущности;
- I (Inside) — продолжение (внутренняя часть) сущности;
- O (Outside) — токен, не являющийся частью сущности.

Использование BIO-разметки обеспечивает унифицированное и корректное представление как односоставных, так и многосоставных названий профессий, таких как «менеджер по продажам» или «инженер-программист».

Сложности при подготовке корпуса данных. Формирование репрезентативного и качественно размеченного корпуса данных сопряжено с рядом сложностей, среди которых можно выделить следующие.

1. Морфологическая вариативность. Одна и та же профессия может быть представлена в тексте различными грамматическими

формами (номинатив «*программист*», родительный падеж «*ищем программиста*», множественное число «*приглашаем программистов*»). Задача аннотатора — распознать лексему в любой форме и пометить её стандартным тегом (напр., B-Job).

2. Определение границ составных профессий. Многие профессии являются многословными терминами («*менеджер по продажам*», «*инженер-программист*»). Требуется единообразно размечать их границы. Принятое решение (напр., размечать «по продажам» как I-Job) должно быть строго зафиксировано в руководстве по разметке и соблюдаться для всех аналогичных случаев, чтобы избежать противоречий в данных для обучения.

Примеры размеченных данных. Для иллюстрации применённой схемы разметки ниже приведены примеры предложений из подготовленного корпуса в формате BIO-разметки.

О B-Job О B-Job О B-Job О
 Набираем строителей , монолитчиков и разнорабочих на

О О О О
 крупнейшие объекты страны .

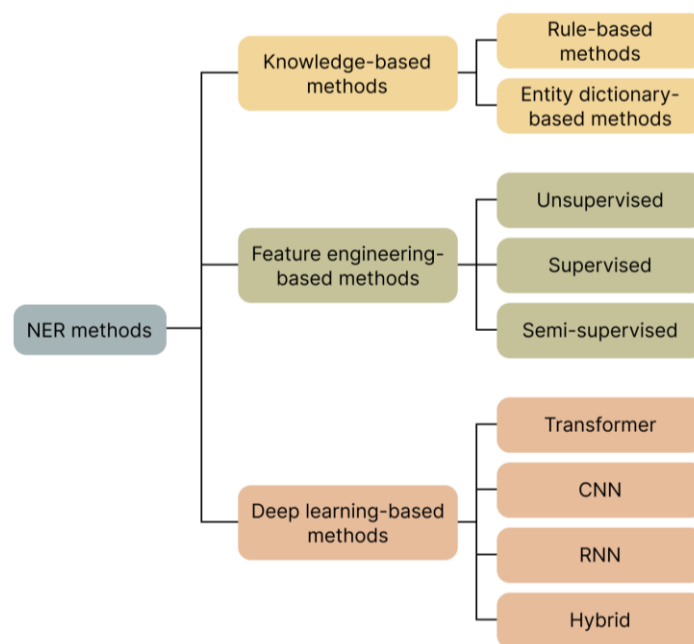
О О B-Job I-Job I-Job О
 Открыта вакансия менеджера по продажам .

Таким образом, сформированный корпус данных обеспечивает репрезентативную основу для обучения и оценки моделей, демонстрируя разнообразие от простых однословных упоминаний профессий до сложных многословных конструкций, характерных для реальной лексики рынка труда.

3. Методы

В области распознавания именованных сущностей можно выделить три основных типа методов: методы, основанные на знаниях; методы, основанные на конструировании признаков; методы, основанные на глубоком обучении (рисунок). Каждый тип имеет свои особенности, преимущества и ограничения [3].

Методы, основанные на знаниях. Методы этой группы стали фундаментом для первых систем NER. Они основываются на использовании заранее сформулированных лингвистических правил, шаблонов и



Основные подходы к NER

словарей [3]. Это позволяет извлекать сущности без необходимости использовать размеченные данные. Ключевым преимуществом данных методов является высокая точность в строго ограниченных предметных областях. Основным недостатком выступает низкая полнота извлечения, обусловленная невозможностью ручного описания всех вариантов упоминаний сущностей. Система не обладает способностью к обобщению и не распознает упоминания, отсутствующие в её словарях.

В рамках данной работы в качестве одного из способов распознавания профессий был применён метод, основанный на использовании специального словаря профессий. Использование словаря позволило эффективно выявлять стандартизированные названия профессий. Преимуществом данного подхода является его простота и высокая интерпретируемость. Недостатки заключаются в слабой способности к обобщению и необходимости постоянного расширения и актуализации словаря для покрытия новых случаев.

Методы, основанные на конструировании признаков. Следующим этапом развития NER стало применение методов машинного обучения, переносящих задачу с ручного описания правил на автоматическое выявление закономерностей. Особенностью этих методов является выделение признаков. Данный процесс заключается в выявлении в тексте информативных характеристик, значимых для классификации. К таким

характеристикам относятся форма слова, часть речи, наличие заглавных букв, контекстные слова и другие.

Одними из наиболее эффективных методов этого направления считаются условные случайные поля (Conditional Random Fields, CRF). Ключевое преимущество CRF перед другими алгоритмами машинного обучения заключается в способности эффективно моделировать зависимости между наблюдениями и метками, учитывая контекстные взаимосвязи [2]. В современных исследованиях CRF часто используется как базовый метод, с которым сравнивают эффективность более сложных нейросетевых моделей.

В рамках данной работы была реализована модель на основе CRF. Её результаты предоставляют важную основу для объективной оценки эффективности других подходов.

Методы, основанные на глубоком обучении. Прорыв в области NER связан с применением методов глубокого обучения, в основе которых лежат нейронные сети. Данные методы демонстрируют высокую эффективность в автоматическом выявлении репрезентативных признаков сущностей из больших наборов данных.

Значительный прогресс достигнут с появлением архитектуры «Трансформер», основанной на механизме внимания. Модели на основе архитектуры трансформеров обладают рядом преимуществ, включая возможность параллельной обработки данных и способность улавливать сложные зависимости в тексте [4]. Использование трансформеров позволило достичь значительно более высокой точности по сравнению с традиционными подходами. Тем не менее, такие модели требуют существенных вычислительных ресурсов и времени на обучение.

В данном исследовании для работы с русскоязычными текстами использовалась предобученная модель RuBERT, которая была дополнительно дообучена на специально размеченном корпусе данных, описанном выше. Её применение позволило достичь наивысшей точности за счёт способности учитывать сложный контекст и распознавать профессии в разнообразных формулировках. Процедура дообучения позволила адаптировать общую языковую модель к специфике задачи распознавания профессий.

4. Результаты

Для оценки качества различных подходов к распознаванию профессий использовались стандартные метрики: Precision (точность), Recall (полнота) и F1-score (сбалансированная метрика, учитывающая и точность, и полноту) [1].

Таблица демонстрирует сравнительные результаты для трёх методов: словарного подхода, CRF и RuBERT.

Сравнение методов распознавания профессий

Метод	Precision	Recall	F1-score
Словарь	0,98	0,55	0,70
CRF	0,84	0,72	0,78
RuBERT	0,87	0,85	0,86

Словарный подход демонстрирует наибольшую точность (0,98) и наименьшую полноту (0,55). Метод обеспечивает минимальное количество ложных срабатываний, но пропускает почти половину релевантных сущностей. Это характерно для данного подхода, который является надёжным для известных паттернов, но не способен к обобщению. Данное свойство делает метод пригодным лишь для ограниченных и строго определённых доменов.

CRF показывает сбалансированное качество модели с улучшением полноты на 17% по сравнению со словарным методом.

RuBERT демонстрирует наилучшие результаты по метрике F1 (0,86). У данной модели оптимальный баланс между точностью и полнотой.

Эволюция качества методов чётко прослеживается в таблице: от высокой точности, но низкой полноты словарей к сбалансированным показателям CRF, а затем к максимальной эффективности нейросетевого подхода.

Таким образом, для большинства реальных приложений рекомендуется использование трансформерных архитектур, как обеспечивающих наиболее полное и точное извлечение сущностей в условиях лингвистического разнообразия русскоязычных текстов.

5. Заключение

В данной работе был проведён комплексный анализ современных подходов к решению задачи распознавания именованных сущностей категории «Профессия» в русскоязычных текстах. Исследование охватило эволюцию методов — от традиционных словарных подходов до современных нейросетевых архитектур на основе трансформеров.

Экспериментальные результаты подтвердили теоретические предположения о преимуществах и ограничениях каждого из подходов. Наилучшие результаты были достигнуты с использованием предобученной языковой модели RuBERT, дообученной на размеченном корпусе текстов. Модель продемонстрировала способность эффективно обрабатывать

сложные случаи: многокомпонентные наименования, морфологическую вариативность и контекстно-зависимые упоминания профессий.

Полученные результаты имеют практическую значимость для разработки систем автоматического анализа текстов в различных прикладных областях: мониторинге рынка труда, анализе вакансий и резюме, социологических исследованиях. Дальнейшие исследования в данном направлении могут быть связаны с расширением размеченного корпуса и созданием энергоэффективных моделей для промышленного применения.

Таким образом, работа вносит вклад в развитие методов обработки естественного языка для русскоязычного сегмента и подтверждает эффективность применения предобученных языковых моделей для решения практических задач информационного поиска и извлечения информации.

Библиографические ссылки

1. Метрики классификации и регрессии. URL: <https://education.yandex.ru/handbook/ml/article/metriki-klassifikacii-i-regressii> (дата обращения: 05.09.2025).
2. *Cariello M. C., Lenci A., Mitkov R.* A Comparison between Named Entity Recognition Models in the Biomedical Domain // *Proceedings of the Translation and Interpreting Technology Online Conference*. 2021. P. 76–84.
3. *Keraghel I., Morbieu S., Nadif M.* Recent Advances in Named Entity Recognition: A Comprehensive Survey and Comparative Study. URL: <https://arxiv.org/abs/2401.10825> (date of access: 05.09.2025).
4. A review of named entity recognition: from learning methods to modelling paradigms and tasks / W. L. Seow [et al.] // *Artificial Intelligence Review*. 2025. Vol. 58, iss. 10. Article no. 315.

ПРОГНОЗИРОВАНИЕ ЦЕН ОПЦИОНОВ НА ОСНОВЕ МОДЕЛИ БЛЭКА-ШОУЛСА

В. И. Лобач¹⁾, В. В. Савицкая²⁾

¹⁾ *Белорусский государственный университет,
Минск, Беларусь, lobach@bsu.by*

²⁾ *Белорусский государственный университет,
Минск, Беларусь, forsmithofficial@gmail.com*

Данная работа посвящена прогнозированию цен опционов на основе модели Блэка-Шоулса. Рассматриваются теоретические основы модели, дифференциальное уравнение и расчетных формул решения уравнения для опционов колл и пут. Особое внимание уделяется модификациям модели для случаев выплаты дискретных и непрерывных дивидендов.

Ключевые слова: опционы; модель Блэка-Шоулса; оценка стоимости.

FORECASTING OPTION PRICES BASED ON THE BLACK-SCHOLES MODEL

V. I. Lobach^{a)}, V. V. Savitskaya^{b)}

^{a)} *Belarusian State University,
Minsk, Belarus, lobach@bsu.by*

^{b)} *Belarusian State University,
Minsk, Belarus, forsmithofficial@bsu.by*

This work is devoted to the forecasting of option prices based on the Black-Scholes model. It covers the theoretical foundations of the model, the differential equation, and the numerical formulas for solving the equation for call and put options. Special attention is given to modifications of the model for cases with discrete and non-continuous dividends.

Keywords: options; Black-Scholes model; valuation.

1. Введение

Опцион – это финансовый инструмент, который предоставляет его владельцу право, но не обязанность, купить или продать базовый актив по заранее установленной цене в определенный момент времени или в течение определенного периода. Имеется два основных типа опционов:

– *колл опционы* дают право владельцу купить лежащий в основе актив в определенную дату по определенной цене;

– *пут опционы* дают право владельцу продать лежащий в основе актив в определенную дату по определенной цене.

По способу исполнения выделяют два типа опционов:

- *американские опционы* исполняются в любое время до даты истечения;
- *европейские опционы* исполняются только в дату истечения.

Покупатели опциона колл (держатели опциона) покупают право купить базовый актив в будущем по определенной цене — цене страйк. Соответственно продавцы опциона колл (иначе называемые подписчики опциона) продают покупателю это право за определенное денежное вознаграждение, называемое премией. Контрактная цена называется ценой исполнения; контрактная дата называется датой истечения, датой исполнения или погашения. Предметом торговли на опционных биржевых площадках является опционная премия. Все остальные параметры опционного контракта — фиксированные [1].

2. Опционные контракты: принципы ценообразования

Опишем, каким образом опцион может предоставлять финансовое плечо с позиции покупателя колл опциона. Приведём некоторые рассуждения в таблице. Посмотрим, как обстоит ситуация на рынке для держателя какого-либо актива и для держателя опциона на такой же актив.

Результаты торгов при держании актива и опциона на такой же актив

	Вложенная сумма, д.е.	Случай роста цены базового актива		Случай падения цены базового актива	
		Прибыль, д.е.	Доходность	Убыток, д.е.	Доходность
Базовый актив	K	dS ($dS > c$)	$\frac{dS}{K} \cdot 100\%$	$-dS$	$-\frac{dS}{K} \cdot 100\%$
Опцион	c ($c < K$)	$dS - c$	$\frac{dS}{dS - c} = 100\% + \frac{c}{dS} \cdot 100\%$	$\begin{cases} -c, dS \geq c; \\ dS - c, dS < c. \end{cases}$	$\begin{cases} \frac{-c}{c} = -100\%; \\ \frac{dS}{c} - 1 = \frac{dS}{c} \cdot 100\% - 100\%. \end{cases}$

K — цена исполнения;

S_T — окончательная цена базового актива;

$dS = |S_T - K|$ — абсолютная величина разности цены исполнения и окончательной стоимости;

c – стоимость опциона (премия). Условие $c < K$ исходит из природы опционов, обратное не несло бы выгоды покупателю.

Доходность с реализации опциона при росте стоимости базового актива составляет более 100%, но потери также велики.

Пришли к тому, что, чтобы определить оптимальную цену опциона, необходимо оценить изменение со временем стоимости актива. Этим вопросом занялся Луи Башелье, а конкретно он взял за предмет изучения стоимость акций. Он исходил из предположения, что изменение цены можно описать математически. Непосредственно цена акций зависит от количества продавцов и покупателей на рынке. Больше продавцов — цена ниже, и наоборот. А на количество этих участников торгов в свою очередь влияет большая совокупность факторов, от погоды, до политической ситуации. Остаётся исходить из того, что акция в любой момент времени может как упасть, так и вырасти. Тогда Башелье пришёл к выводу, что цена акции блуждает случайным образом, и смог связать колебания цен акций с броуновским движением. Теория Роберта Брауна утверждает, что микроскопические частицы двигаются абсолютно случайным образом. В свою очередь ожидаемое положение частицы описывается законом нормального распределения. Тогда плотность распределения изменения цены акции имеет стандартный вид колоколообразной прямой.

Теперь свяжем изменение цен акций с ожидаемой доходностью опциона, вновь рассматривая ситуацию с позиции покупателя опциона колл. Зная закон распределения изменений цены акции, можем получить вероятность того, что покупка опциона принесёт прибыль. Это можно продемонстрировать графически (рис. 1). Закрасим зелёным вероятность остаться в прибыли. Заметим, что увеличение цены опциона означает меньшую вероятность заработка. Не закрашенной осталась вероятность убытка, то есть заработка продавца опциона.

Башелье рассудил, что, умножив прибыль или убыток на их вероятности, можно вычислить ожидаемую доходность опциона (рис. 2). Вновь визуализируем, на этот раз отрицательные и положительные доходности (красные и зелёный цвет на графике соответственно). Иначе это можно назвать доходностью продавцов и покупателей. Башелье справедливо рассудил, что обеим сторонам должны полагаться одинаковые доходности.

Опционы, наравне с другими видами вторичных ценных бумаг, являются ключевым элементом биржевой стратегии хеджирования. *Хеджирование* (от английского слова to hedge — защищать, ограничивать) — это стратегия, при которой участники рынка страхуют свои операции по базовому активу с помощью заключения соответствующих опционных (в общем случае не только таких) контрактов [2].

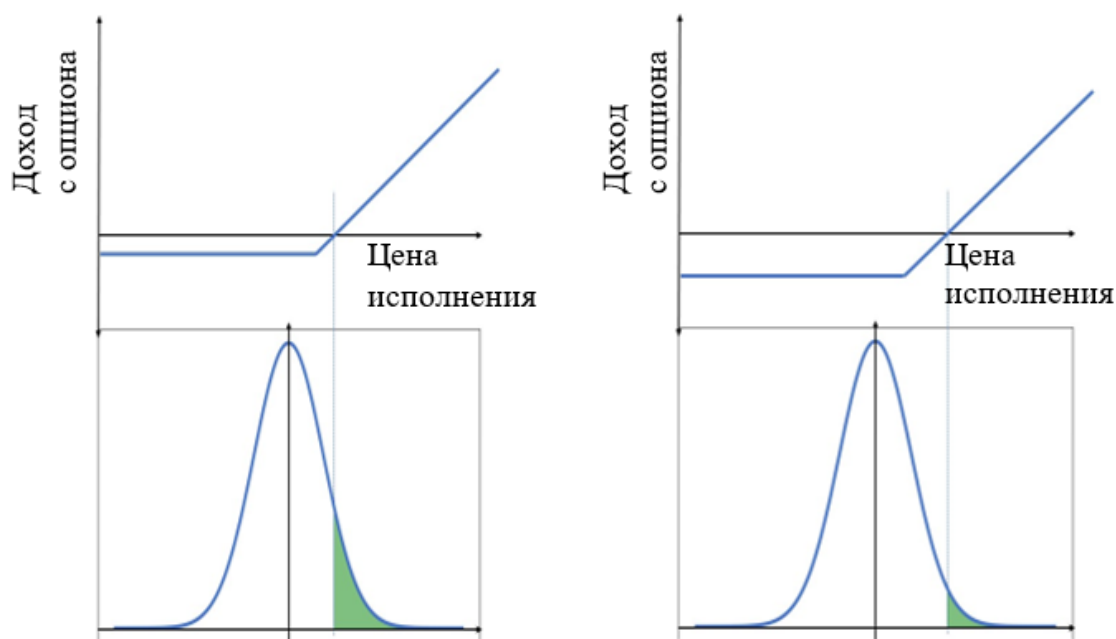


Рис. 1. Связь дохода держателя с покупки опциона колл и изменения цен на актив

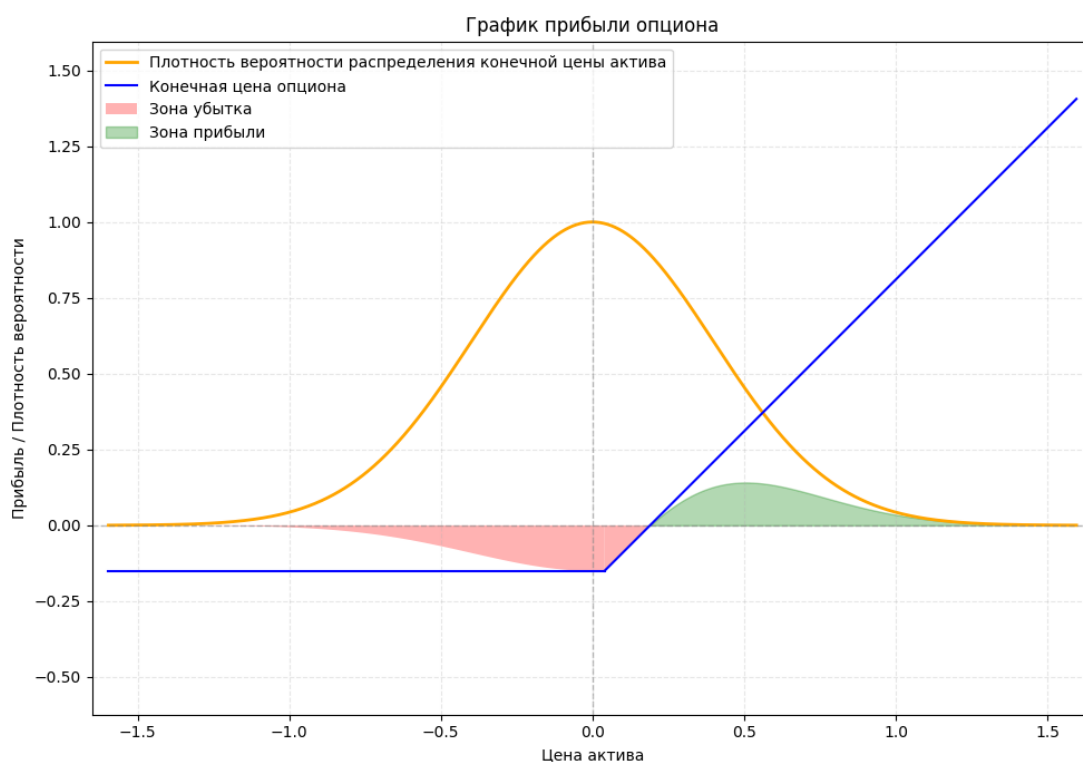


Рис. 2. Доходность с опциона

Заслугой Эдварда Торпа является описание страхования убытков компенсирующими операциями, так называемого динамического хеджирования. Держатель акции продал опцион на её продажу. Если к дате исполнения цена акции вырастет, то держатель при её продаже по опциону

потерпит убыток. Чтобы предотвратить потери, держателю необходимо иметь ещё одну такую же акцию, которая ростом своей цены будет компенсировать убыток по опциону. В случае же, когда цена на акцию падает, держатель продаёт вторую акцию и имеет заработок с того, что цена страйк опциона будет выше текущей цены базового актива.

Торп вывел формулу захеджированного портфеля:

$$\Pi = (-1)V + \Delta S = (-1)V(S(t), t) + \frac{\partial V}{\partial S} S(t), \quad (1)$$

Π – стоимость портфеля;

$V(S(t), t)$ – стоимость финансовой производной, то бишь опциона.

Далее рассуждения Торпа расходятся с тем, что предложили Фишер и Блэк.

3. Модель Блэка-Шоулса

В своей работе о ценах опционов 1973 года Блэк и Шоулс вывели дифференциальное уравнение для цен финансовых производных, зависящих от стоимости акций, не выплачивающих дивиденды. Основная идея авторов заключалась в создании безрискового портфеля, состоящего из двух финансовых инструментов: производной и акции. Доход от этого портфеля приравнивается к доходу от аналогичной инвестиции по безрисковой ставке.

Перечислим некоторые допущения в модели Блэка-Шоулса [3]:

- 1) уровень процентных ставок остается неизменным и известен заранее;
- 2) используются временные сроки исполнения для европейских опционов;
- 3) в течение срока действия опциона дивиденды по базовым акциям не выплачиваются;
- 4) отсутствие взимаемых комиссий;
- 5) рынки являются эффективными.

Фишер Блэк и Майрон Шоулс развивали свою теорию параллельно Торпу. Они брали за основу предположение Башелье о справедливой стоимости опциона и отталкивались от модели безрискового портфеля Торпа. Дальнейшие их рассуждения привели к тому, что на эффективном рынке доходность портфеля не может превышать безрисковую ставку. Были использованы дополненные рассуждения Башелье, которые заключаются в том, что изменение цены на акцию складывается как из случайного изменения цены, так и из общей тенденции рынка. Привёл уравнение к дифференциальному виду же Роберт Мёртон, применив стохастическое исчисление и лемму Ито [4].

$$rV = \frac{\partial V}{\partial t} + rS \frac{\partial V}{\partial S} + \frac{1}{2} \frac{\partial^2 V}{\partial S^2} \sigma^2 S^2. \quad (2)$$

Приведено дифференциальное уравнение Блэка-Шоулса для цены финансовой производной. Приведём финансовый смысл слагаемых.

rV – доходность безрискового актива. Показывает, что ожидаемая доходность производного инструмента должна равняться безрисковой ставке (в условиях нейтрального к риску мира). Т.е. если бы V была инвестирована в безрисковый актив, она росла бы со ставкой r .

$\frac{\partial V}{\partial t}$ – скорость изменения стоимости опциона со временем. Учитывает уменьшение стоимости опциона из-за сокращения времени до экспирации.

$rS \frac{\partial V}{\partial S}$ – влияние дрейфа цены базового актива (ожидаемого роста S со ставкой r) на стоимость опциона. Состоит из безрискового роста цены актива и чувствительности доходности к изменению цены актива. Отражает вклад линейного движения цены актива в стоимость опциона.

$\frac{1}{2} \frac{\partial^2 V}{\partial S^2} \sigma^2 S^2$ – отражает влияние волатильности и нелинейного роста цены актива. Состоит из скорости роста чувствительности и учёта случайных колебаний. Отражает «стоимость хеджирования» против риска резких движений цены (чем выше волатильность, тем дороже опцион).

Чтобы получить из него вариацию для конкретной финансовой производной, необходимо ввести соответствующие краевые условия. Их приводили в предыдущей главе. Например, для европейского опциона колл они имеют вид $V = \max(S_T - K, 0)$. Модель применима и для фьючерсов с краевым условием $V = S_T - K$.

Решением полученного уравнения является формула *Блэка-Шоулса*.

Приведём вид решения дифференциального уравнения – формулы Блэка-Шоулса для вычисления первоначальной стоимости европейского опциона колл [5]:

$$c = S_0 \mathcal{N}(d_+) - Ke^{-rT} \mathcal{N}(d_-), \quad (3)$$

$$d_{\pm} = \frac{\ln \frac{S_0}{K} + \left(r \pm \frac{\sigma^2}{2} \right) T}{\sigma \sqrt{T}},$$

где

K – цена исполнения;

S_0 – стоимость базового актива в момент заключения контракта;

c – стоимость опциона колл;

r – величина безрисковой процентной ставки;

T – время до экспирации;

$\mathcal{N}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(\frac{-x^2}{2}\right) dx$ – кумулятивная функция распределения

нормального распределения;

$\sigma = \sqrt{\frac{1}{N-1} \sum_{t=1}^N (r_t - \mu)^2}$ – волатильность актива, где μ – средняя доход-

ность, N – число наблюдений, $r_t = \ln\left(\frac{S_t}{S_{t-1}}\right)$ – логарифмические доходности.

Считается, что использование 60–200 ежедневных ценовых наблюдений, 40–60 еженедельных ценовых наблюдений или 30–50 ежемесячных ценовых наблюдений является хорошим компромиссом.

Стоимость европейского пут-опциона принимает вид [6]

$$p = Ke^{-rT} N(-d_-) - S_0 N(-d_+), \quad (3)$$

p – стоимость опциона пут.

Вывод формулы будет аналогичен и исходить из начального условия $V(S, 0) = \max(K - S_T, 0)$ или же исходить из правила Пут-колл паритета:

$$c - p = e^{-rT} (S_0 e^{rT} - K).$$

Тогда

$$\begin{aligned} p &= c + Ke^{-rT} - S_0 = S_0 \mathcal{N}(d_+) - Ke^{-rT} \mathcal{N}(d_-) + Ke^{-rT} - S_0 = \\ &= Ke^{-rT} (1 - \mathcal{N}(d_-)) - S_0 (1 - \mathcal{N}(d_+)). \end{aligned}$$

Исходя из симметрии нормального распределения $1 - \mathcal{N}(x) = \mathcal{N}(-x)$. Из чего и выходит формула (4).

Тем не менее, отталкиваясь от вида формулы для колл опциона, вид этой довольно интуитивен. С позиции продавца величина K — это то, что выплатит покупатель, соответственно он заработаем. Функция кумулятивного распределения даёт нам вероятность того, что цена базового актива будет меньше или равна аргументу функции, а взяв аргумент со знаком

минус, описанное неравенство станет «больше или равна», что как раз случай прибыли продавца опциона. Отнимается же упущенная прибыль с базового актива, умноженная на вероятность такового расклада.

4. Учет дивидендов в ценообразовании опционов

В оригинальной статье Блэка и Шоулса не предусматривались выплаты по базовому активу. Дивидендная доходность была добавлена только Мертоном в его работе 1973 года “Theory of Rational Option Pricing”. Тогда будем иметь в виду, что держатель базового актива всё время до экспирации опциона будет получать дивиденды по нему. Соответственно, держатель опциона теряет некоторую прибыль, что будет учитываться в уменьшении стоимости опциона.

1. Дискретные дивиденды [7].

D_i – ожидаемый размер дивиденда в i -ом периоде;

T_i – время в годах до i -ой выплаты дивидендов;

N – ожидаемое количество выплат дивидендов до истечения срока действия опциона.

$$c_D = e^{-rT} \left(S_0 - \sum_{i=1}^N D_i e^{-rT_i} \right) \mathcal{N}(d_+) - e^{-rT} K \mathcal{N}(d_-);$$

$$p_D = e^{-rT} K \mathcal{N}(-d_-) - e^{-rT} \left(S_0 - \sum_{i=1}^N D_i e^{-rT_i} \right) \mathcal{N}(-d_+).$$

Чтобы понять, как учесть дискретные дивидендные выплаты в модель Блэка-Шоулса, важно понять, что уравнение (2) описывает цену опциона при каждой конкретной стоимости базового актива. Все изменения этого актива будут заложены в модель, выбранную для объяснения стохастического процесса изменения цены базового актива. В нашем случае это модель геометрического броуновского движения. Тогда можем утверждать, что дифференциальное уравнение (2) останется неизменным и необходимо описать моменты выплаты дивидендов. В формулах (3) и (4) учтётся снижение цены акции с каждой выплатой дивидендов. При этом размер дивидендов дисконтируется множителем e^{-rT_i} . В финансовом смысле это переводит «будущие деньги» в «настоящие».

2. Непрерывные дивиденды [8].

q – постоянная дивидендная ставка базового актива.

$$c_q = e^{-qT} S_0 \mathcal{N}(d_+^q) - e^{-rT} K \mathcal{N}(d_-^q);$$

$$p_q = e^{-rT} K \mathcal{N}(-d_-^q) - e^{-qT} S_0 \mathcal{N}(-d_+^q).$$

Предположения о непрерывности дивидендных выплат и постоянной ставке дивидендной доходности должны быть учтены при расчете параметров:

$$d_{\pm}^q = \frac{\ln \frac{S_0}{K} + \left(r - q \pm \frac{\sigma^2}{2} \right) T}{\sigma \sqrt{T}}.$$

Исходя из экономической логики, безрисковая ставка r отражает альтернативную доходность, которую инвестор мог бы получить, дивиденды q — это часть доходности акции, которая не остаётся в цене актива, а выплачивается акционеру. Тогда чистая доходность для держателя акции в процентах составит $r - q$. Не трудно понять, что отличающийся от оригинального лишь заменой множителя r на $r - d$. Тогда аналогично и изменится решение формула для опциона-колл будет отличаться лишь заменой множителя r на $r - q$:

$$\begin{aligned} c_q &= e^{(r-q)T} \left(S_T \mathcal{N}(d_+^q) - K \mathcal{N}(d_-^q) \right) = e^{(r-q)T} \left(S_0 e^{rT} \mathcal{N}(d_+^q) - K \mathcal{N}(d_-^q) \right) = \\ &= e^{-qT} S_0 \mathcal{N}(d_+^q) - e^{-rT} K \mathcal{N}(d_-^q), \\ d_{\pm}^q &= \frac{\ln \frac{S_0}{K} + \left((r - q) \pm \frac{\sigma^2}{2} \right) T}{\sigma \sqrt{T}}. \end{aligned}$$

5. Заключение

В данной работе исследованы возможности прогнозирования цен опционов на основе модели Блэка-Шоулса. Рассмотрены теоретические основы опционных контрактов, их экономическая сущность и применение в хеджировании рисков. Проанализирован исторический контекст формирования модели — от исследований Башелье о случайном характере движения цен до разработки Торпом концепции динамического хеджирования. Изучена модель Блэка-Шоулса-Мертон: введено базовое дифференциальное уравнение, раскрыт финансовый смысл его компонентов и получено аналитическое решение для европейских опционов. Исследованы модификации модели для учёта дискретных и непрерывных дивидендных выплат.

Проведенный анализ подтверждает, что модель Блэка-Шоулса, несмотря на строгие предпосылки, остается фундаментальной основой количественных финансов. Её значение заключается не только в предоставлении расчётных формул, но и в создании теоретического базиса для

оценки производных финансовых инструментов через принципы арбитражного ценообразования и построения безрисковых портфелей.

Библиографические ссылки

1. *Медведев Г. А.* Математические модели финансовых рисков : учебное пособие : в 2 ч. Минск : БГУ, 1999. Ч. 1. Риски из-за неопределенности процентных ставок.
2. *Балабушкин А. Н.* Опционы и фьючерсы : методическое пособие. Фондовая биржа РТС, 2004.
3. Формула Блэка-Шоулса. URL: https://www.synset.com/wiki/Формула_Блэка-Шоулса (дата обращения: 09.12.2024).
4. *Поттосина С. А., Алёхина А. Э.* Математика рынка ценных бумаг (с элементами технического анализа) : учебно-метод. пособие. Минск : БГУИР, 2012. URL: <https://libeldoc.bsuir.by/handle/123456789/536> (дата обращения: 16.09.2025).
5. Математика опционов или модель Блэка-Шоулса. URL: <https://habr.com/ru/articles/552194/> (дата обращения: 11.12.2024).
6. *Буренин А. Н.* Фьючерсные, форвардные и опционные рынки : учебное пособие. М. : Тривола, 1994.
7. *Company R., Gonzalez A. L., Jodar L.* Numerical solution of modified Black–Scholes equation pricing stock options with discrete dividend // Mathematical and Computer Modelling. Vol. 44, iss. 11–12. P. 1058–1068.
8. *Wu L., Yang X.* An Efficient Difference Algorithm for Black-Scholes Equation with Payment of Dividend // Proceedings of the 2012 2nd International Conference on Computer and Information Application (ICCIA 2012) / Paris : Atlantis Press, 2012. P. 470–473.

ПРОГНОЗИРОВАНИЕ ДОХОДНОСТИ ФИНАНСОВЫХ АКТИВОВ НА ОСНОВЕ МОДЕЛИ ВАСИЧЕКА

С. В. Лобач¹⁾, Е. В. Остапук²⁾

¹⁾ Белорусский государственный университет,
Минск, Беларусь, fpm.lobach@gmail.com

²⁾ Белорусский государственный университет,
Минск, Беларусь, fpm.ostapuk@bsu.by

В работе исследуется применение модели Васичека для прогнозирования доходности финансовых активов. Модель Васичека представляет собой однофакторную модель процентных ставок, основанную на процессе Орнштейна-Уленбека. Проведен анализ различных методов оценки параметров модели и их практическое применение на данных казначейских векселей США. Показана эффективность модели для краткосрочного прогнозирования процентных ставок.

Ключевые слова: модель Васичека; процентные ставки; прогнозирование; финансовые активы; оценка параметров.

FORECASTING THE PROFITABILITY OF FINANCIAL ASSETS BASED ON THE VASICEK MODEL

S. V. Lobach^{a)}, A. V. Astapuk^{b)}

^{a)} Belarusian State University,
Minsk, Belarus, fpm.lobach@gmail.com

^{b)} Belarusian State University,
Minsk, Belarus, fpm.ostapuk@bsu.by

The paper examines the application of the Vasicek model to predict the availability of financial assets. The analysis of various methods for estimating model parameters and their practical application on the data of US Treasury bills is carried out. The effectiveness of the model for short-term forecasting of interest rates is shown.

Keywords: Vasicek model; interest rates; forecasting; financial assets; parameter estimation.

1. Введение

В настоящее время прогнозирование доходности является неизменным и необходимым фактором для анализа финансовых активов. Прогнозирование является важным при инвестировании и управлении рисками.

Когда рынок понятен и стабилен, то в силу уже сложившихся правил и законов финансового рынка установлено, что участник рынка, вложивший в некоторый момент времени определенную сумму в некоторый финансовый проект заданной продолжительности, получит после его выполнения сумму, рассчитываемую через интенсивность получения дохода с течением времени [1].

Однако, когда рыночная ситуация неопределенна, а процентные ставки стохастичны, найти стоимость финансовых активов становится сложной задачей.

Чтобы прогнозировать поведение ставок на нестабильном рынке, была разработана модель Олдриха Васичека в 1977 году, это мощный инструмент для оценки процентных ставок, что позволяет также прогнозировать стоимость акций и облигаций.

2. Модель Васичека

Модель была разработана в 1977 году Олдрихом Васичеком [2]. В данной модели предполагается, что значение процентной ставки колеблется около среднего значения, полученного за некоторый промежуток времени. Модель является однофакторной, так как в ней присутствует только один фактор неопределённости.

Процессы изменения процентных ставок были определены с помощью стохастических дифференциальных уравнений.

Модель Васичека классифицируется как однофакторная модель, а процентная ставка называется мгновенной процентной ставкой. Однофакторные модели предполагают, что существует только один источник риска. Эти модели могут использоваться перед применением многофакторных моделей как предварительный анализ, однако однофакторные модели дают хорошее представление о процентных ставках. Более того, они могут быть решены аналитически, а наличие аналитических решений позволяет быстро оценить облигации, опционы, цены и параметры хеджирования [3].

В модели Васичека спот ставка $r(t)$ определяется так называемым процессом Орнштейна-Уленбека, и уравнение имеет следующий вид:

$$dr(t) = \alpha(\beta - r(t))dt + \sigma dW(t), \quad r(0) = r_0, \quad (1)$$

где W_t – процесс Винера, моделирующей случайный рыночный фактор риска.

Наиболее важной особенностью, которую демонстрирует эта модель и которую объяснил Васичек, является возврат к среднему значению, что

означает, что если процентная ставка больше долгосрочного среднего значения ($r > \beta$), то коэффициент α , $\alpha > 0$, приводит к тому, что дрейф становится отрицательным (дрейф представляет собой компонент, который определяет, как быстро и в каком направлении переменная возвращается к своему среднему значению), так что ставка будет снижаться в направлении β . Аналогично, если процентная ставка меньше, чем долгосрочное среднее ($r < \beta$), то коэффициент α , $\alpha > 0$, делает дрейф положительным, так что ставка будет повышаться в направлении β . Таким образом, коэффициент α – скорость приведения процентной ставки к ее долгосрочному уровню [4].

Одной из наиболее привлекательных особенностей модели Васичека является то, что она имеет решение в закрытой форме (аналитическое решение) для r_t .

Начинаем решение стохастического дифференциального уравнения (1) с вычисления производной от $e^{\alpha t} r_t$:

$$d(e^{\alpha t} r_t) = r_t \alpha e^{\alpha t} dt + e^{\alpha t} dr_t, \quad (2)$$

$$d(e^{\alpha t} r_t) = r_t \alpha e^{\alpha t} dt + e^{\alpha t} [\alpha(\beta - r_t)dt + \sigma dW_t], \quad (3)$$

$$\int_s^t d(e^{\alpha u} r_u) = \int_s^t e^{\alpha u} [\alpha\beta du + \sigma dW_u], \quad (4)$$

$$e^{\alpha t} r_t = e^{\alpha s} r_s + \beta(e^{\alpha t} - e^{\alpha s}) + \sigma \int_s^t e^{\alpha(t-u)} dW_u. \quad (5)$$

Решением дифференциального уравнения (1) будет

$$r_t = e^{-\alpha(t-s)} r_s + \beta(1 - e^{-\alpha(t-s)}) + \sigma \int_s^t e^{-\alpha(t-u)} dW_u. \quad (6)$$

Процесс, приведенный в уравнении (6), соответствует нормальному Гауссовскому распределению с математическим ожиданием [5]

$$E(r_t) = \beta + (r_s - \beta)e^{-\alpha(t-s)} \quad (7)$$

и дисперсией

$$D(r_t) = \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha(t-s)}). \quad (8)$$

Долгосрочное распределение процесса Орнштейна-Уленбека является стационарным и является нормальным со средним значением β и дисперсией $\frac{\sigma^2}{2\alpha}$.

3. Метод максимального правдоподобия для оценки параметров модели

Метод максимального правдоподобия (ММП) находит параметры, которые максимизируют вероятность наблюдения имеющихся данных.

Модель Васичека описывается стохастическим дифференциальным уравнением:

$$dr(t) = \alpha(\beta - r(t))dt + \sigma dW(t), \quad (9)$$

где W_t — винеровский процесс (броуновское движение).

Поскольку процесс Орнштейна-Уленбека, на котором основана модель Васичека, имеет нормальное распределение, условное распределение r_{t+1} при известном r_t будет нормальным:

$$r_{t+1} \sim N\left(\beta + (r_t - \beta)e^{-\alpha\Delta t}, \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha\Delta t})\right), \quad (10)$$

где Δt — шаг дискретизации (например, 1 день).

Функция правдоподобия $L(\alpha, \beta, \sigma)$ строится как произведение плотностей вероятности для всех пар (r_t, r_{t+1}) . Логарифм правдоподобия:

$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(Var) - \sum_{t=1}^n \frac{(r_{t+1} - \mu_t)^2}{2Var}, \quad (11)$$

где $\mu_t = \beta + (r_t - \beta)e^{-\alpha\Delta t}$, а $Var = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha\Delta t})$.

Параметры α, β, σ находятся численно (например, с помощью градиентного спуска), чтобы максимизировать $\ln L$ в формуле (11).

У данного метода есть несомненные преимущества, он учитывает всю информацию из данных, а также асимптотически эффективен, то есть даёт минимальную дисперсию оценок при больших выборках. Однако

присутствует и ряд недостатков, такие как чувствительность к выбросам и требование численной оптимизации, что может быть вычислительно затратным.

4. Оценка параметров модели Васичека и построение прогнозов на основе модели

Применим модель Васичека к реальным данным. В данном случае используются ежедневные данные с начала 2025 года до 9 мая 2025 года о ставке по казначейским векселям на вторичном рынке США [6].

Для оценки воспользуемся дискретной формой модели Васичека:

$$\Delta r_t = \alpha(\beta - r_{t-1})dt + \varepsilon_t(0, \sigma^2). \quad (12)$$

Преобразуем:

$$r_t - r_{t-1} = \alpha\beta - \alpha r_{t-1} + \varepsilon_t(0, \sigma^2), \quad (13)$$

где единицей Δt в данном случае, для ежедневных данных, будет являться день, $\varepsilon_t(0, \sigma^2)$ — независимо и одинаково распределённая случайная величина. Константа в этой модели — оценка произведения $\alpha\beta$, а α в этой регрессии — оценка отрицательного значения α , отрицательного значения скорости возврата к среднему (дрейфа).

Сначала вычисляются первые значения дельт процентной ставки, для этого из текущего значения процентной ставки отнимается предыдущее, что можно видеть в табл. 1.

Таблица 1

Вычисление первых дельт

observation_date	DTB3		delta
2025-01-02	4,22		
2025-01-03	4,20		−0,02
2025-01-06	4,21		0,01
2025-01-07	4,21		0,00
2025-01-08	4,21		0,00

Отдельно добавляется колонка, отвечающая за лаг, то есть за значения, смещённые на один день, что отображено в табл. 2.

Таблица 2

Вычисление лаговых значений

observation_date	DTB3		delta	lag
2025-01-02	4,22			
2025-01-03	4,20		–0,02	4,22
2025-01-06	4,21		0,01	4,20
2025-01-07	4,21		0,00	4,21
2025-01-08	4,21		0,00	4,21

Данные предварительные вычисления помогают использовать функцию **ЛИНЕЙН(D3:D85;E3:E85;1;1)**, которая рассчитывает статистику для ряда с применением метода наименьших квадратов, чтобы вычислить прямую линию, которая наилучшим образом аппроксимирует имеющиеся данные и затем возвращает массив, который описывает полученную прямую. Результат представлен в табл. 3.

Таблица 3

Результат работы линейной регрессии

	$-\alpha$	$\alpha \cdot \beta$	
coef	–0,336212168	1,41488612	
std err	0,083102325	0,349722492	
	0,168106084	0,010433845	sigma
	16,36818412	81	
	0,001781924	0,008818076	

Проверка статистической значимости коэффициентов, которые были рассчитаны, необходима, чтобы убедиться, что модель действительно в значительной степени объясняет динамику процентных ставок по векселям. Чтобы рассчитать t-статистику, коэффициенты делятся на соответствующие стандартные ошибки: $-\frac{0,33621}{0,08310} = -4,04576$, $\frac{1,41488}{0,34972} = 4,04573$.

Далее используется функция **СТЮДЕНТ.РАСП.2Х(ABS(H9);I6)**, которая возвращает двустороннее t-распределение Стьюдента, где H9 – числовое значение, для которого необходимо вычислить распределение, I6 – число степеней свободы. Результат работы данной функции говорит о статистической значимости коэффициентов регрессии, так как значение $p\text{-val} < 0,05$.

Перед началом прогнозирования необходимо вычислить параметры α и β . Из табл. 1 видно, что $-\alpha = -0,336212168$, следовательно,

$\alpha = 0,336212168$. Из этой же таблицы заметим, что $\alpha \cdot \beta = 1,41488612$, значит, $\beta = 4,208313253$.

Благодаря данной модели возможен прогноз процентных ставок, а также их волатильности в любой момент времени в будущем, в том числе в долгосрочной перспективе. Для прогнозирования будем отталкиваться от наблюдения за 2 мая 2025, чтобы в дальнейшем сравнить действительные и прогнозируемые значения. Значение ставки составляет 4,22; выберем некоторый временной горизонт в днях для прогноза. Прогноз будет строиться на следующие 5 дней.

Чтобы составить прогноз, необходимо рассмотреть, каков курс сейчас и как быстро он вернется к равновесному значению в течение каждого из 5 дней. Для это используется следующая формула:

$$E(r_{t+T}) = r_0 e^{-\alpha T} + \beta(1 - e^{-\alpha T}). \quad (14)$$

Из формулы (14) можно заметить, что при бесконечном значении времени t коэффициент β и будет являться прогнозом процентной ставки.

Также можно использовать модель Васичека для вычисления дисперсии значения в каждый из прогнозируемых 5 дней.

$$Var(r_{t+T}) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}). \quad (15)$$

По формуле (15) дисперсия для ставки по казначейским векселям на вторичном рынке США составила 0,000156. Волатильность же составит значение квадратного корня дисперсии: 0,0125. Эти результаты означают, что прогнозное значение будет отклоняться на примерно 0,0125

Сводные результаты по всем дням прогнозирования можно увидеть в табл. 4.

Таблица 4

Результаты по дисперсии и волатильности для прогнозных значений

Горизонт прогноза (дни)	5	4	3	2	1
Прогноз ставки	4,210489	4,211358	4,212575	4,214278	4,216663
Дисперсия ставки	0,000156	0,000150	0,000140	0,000119	7,925E-05
Волатильность ставки	0,012501	0,012284	0,011847	0,010941	0,008902

Из формулы (15) можно заметить, что при бесконечном значении времени t коэффициент $\frac{\sigma^2}{2\alpha}$ и будет являться дисперсией процентной ставки.

Можно заметить, что чем более неустойчиво само равновесие, тем более стохастические факторы влияют на процентную ставку в краткосрочной перспективе, тем более волатильными будут проценты.

Очевидно, что чем меньше параметр сходимости параметр α , тем дольше процентная ставка может колебаться и совершенно непредсказуемо.

Благодаря тому что в оценке модели не были задействованы 5 последних наблюдений, есть возможность провести сравнительный анализ полученных прогнозных значений и реальных данных, что отображено в табл. 5.

Таблица 5

Сравнительная таблица прогнозных и реальных значений процентной ставки

Дата	Прогноз	Реальные данные
2025-05-09	4,210489049	4,22
2025-05-08	4,211358575	4,23
2025-05-07	4,212575596	4,22
2025-05-06	4,214278981	4,22
2025-05-05	4,216663101	4,22

В табл. 5 видно, что прогнозные значения не сильно отклоняются от реальных значений, из чего можно сделать вывод, что метод применим для прогнозирования процентных ставок.

Аналогично только что проделанному, можно смоделировать более длительный прогноз. Для этого используются месячные данные за последние 5 лет по казначейским векселям на вторичном рынке США.

В результате получили следующие оценки параметров модели:

Скорость возврата к среднему (α): 0,1000;

Долгосрочное среднее (β): 0,0500;

Волатильность (σ): 0,0200.

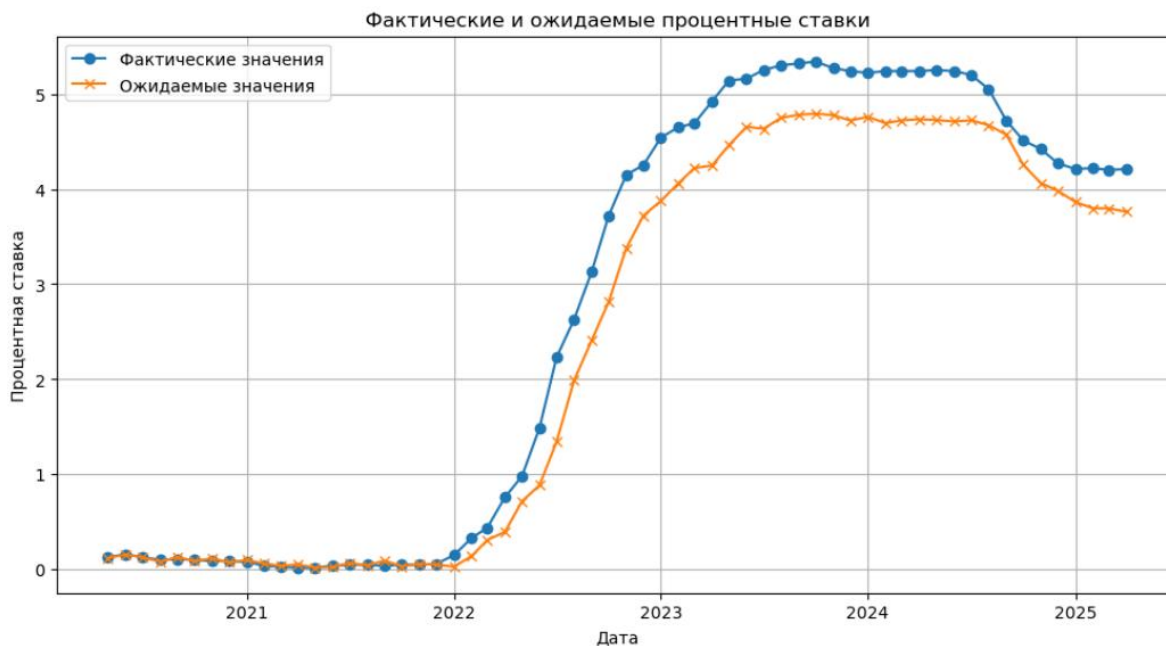
Также графический результат прогноза представлен на рисунке.

5. Заключение

В данной работе получены следующие результаты:

1) построены алгоритмы оценивания параметров на основе модели Васичека;

2) дано графическое представление результатов моделирования стоимости казначейских векселей на вторичном рынке США на основе модели Васичека;



Фактические и ожидаемые процентные ставки

3) проведены компьютерные эксперименты по прогнозированию процентной ставки на основе модели Васичека в отношении казначейских векселей на вторичном рынке США.

Библиографические ссылки

1. Медведев Г. А. Математические модели финансовых рисков : учебное пособие : в 2 ч. Минск : БГУ, 1999. Ч. 1. Риски из-за неопределенности процентных ставок.
2. Vasicek O. An Equilibrium Characterization of the Term Structure // J. Financial Econom. 1977. Vol. 5., no. 2. P.177–188.
3. Ayranci G., Özgürel B. Monte Carlo Simulation for Vasicek Interest Rate Model Parameters // Digital Proceeding Of The ISDS'2014, Side, Turkey, May 10–14, 2014. P. 87–94. URL: https://www.researchgate.net/publication/275823482_Monte_Carlo_Simulation_for_Vasicek_Interest_Rate_Model (date of access: 16.09.2025).
4. Önalın Ö. Vasicek ve CIR Modelleri Kullanılarak Oynaklık ve Faiz Oranlarının Modellenmesi // Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi. 2015. Vol. 27, iss. 2. P. 329–344.
5. Гамбаров Г., Шевчук И., Балабушкин А. Оценка срочной структуры процентных ставок. URL: <https://fs.moex.com/files/12008/> (дата обращения: 10.04.2025).
6. 3-Month Treasury Bill Secondary Market Rate, Discount Basis (DTB3) // Federal Reserve Economic Data | FRED | St. Louis Fed. URL: <https://fred.stlouisfed.org/series/DTB3/> (date of access: 09.05.2025).

АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ПО ДАННЫМ РЕАЛЬНОГО ВРЕМЕНИ С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ СМЕШАННОЙ ЧАСТОТЫ

В. И. Малюгин

*Белорусский государственный университет,
Минск, Беларусь, Malugin@bsu.by*

В статье представлены эконометрические модели по данным смешанной частоты для текущего прогнозирования (наукастинга) месячных индексов цен производителей и потребителей в белорусской экономике на основе ежедневных обменных курсов белорусского рубля. Модели имеют высокую точность прогнозов (наукастов) и позволяет установить опережающее влияние российского индекса цен производителей промышленной продукции на белорусские индексы.

Ключевые слова: модели смешанной частоты; данные реального времени индексы цен; обменные курсы валют; причинный анализ; наукастинг.

ANALYSIS AND FORECASTING BASED ON REAL-TIME DATA USING MIXED-FREQUENCY MODELS

V. I. Malugin

*Belarusian State University,
Minsk, Belarus, Malugin@bsu.by*

The paper presents econometric models based on mixed-frequency data for current forecasting (nowcasting) of monthly producer and consumer price indices in the Belarusian economy based on daily exchange rates of the Belarusian ruble. The models have high forecast (nowcasting) accuracy and allow establishing the leading influence of the Russian producer price index on Belarusian indices.

Keywords: mixed-frequency models; real-time data; price indices; exchange rates; causal analysis; nowcasting.

1. Введение

Модели по данным смешанной частоты, появившиеся и получившие интенсивное развитие в последние два десятилетия [1], предоставляют новую методологию совместного анализа и прогнозирования временных рядов с различной частотой наблюдения, актуальную в условиях роста объема и разнообразия доступной в реальном времени информации. Процедуры

построения моделей по смешанным данным используют методы машинного обучения. Широкое распространение указанные модели получили в задачах эконометрического моделирования, анализа и прогнозирования. Так, использование в моделях для месячных и квартальных макроэкономических показателей высокочастотных данных (например, ежедневных) сделало возможным создание автоматизированных платформ для мониторинга макроэкономических условий в режиме реального времени [2, 3].

Эконометрические модели, основанные на данных смешанной частоты, актуальны в связи с необходимостью текущего прогнозирования (наукастинга) ключевых макроэкономических показателей по поступающим в реальном времени данным, до появления их официальных значений с существенной задержкой. В [4] предлагаются модели по смешанным данным MIDAS для прогнозирования индекса потребительских цен в белорусской экономике, в частности, квартальная модель индекса по его прошлым месячным значениям и данным реального времени в виде ежедневных изменений обменных курсов основных валют. В данном докладе представляется более широкий состав одномерных и многомерных моделей по смешанным данным MIDAS, MIDAS-GETS и MF-VAR, которые помимо текущего прогнозирования, позволяют эффективно проводить анализ причинно-следственных связей между используемыми переменными с разной частотой наблюдения.

2. Описание моделей

В работе построены следующие модели по смешанным данным:

- модель MIDAS/PDL (*Mixed Data Sampling*) с полиномиально распределенными лагами Алмона (*Polynomial Distributed Lags – PDL*) [1];
- модель MIDAS-GETS/Auto/IIS. на основе подхода «от общего к частному» (*General-To-Specific – GETS*) [5] с использованием машинного обучения для поиска аномальных наблюдений и включения в модель фиктивных переменных (*Impulse Indicators of Saturation – IIS*) с последующей оптимизацией лаговой структуры для высокочастотных переменных;
- векторная авторегрессионная модель по данным смешанной частоты MF-VAR (*Mixed-Frequency Vector Autoregression*) для совместного прогнозирования нескольких низкочастотных переменных с использованием высокочастотных данных [6]. Модель допускает применение теста «причинной зависимости по Грейнджеру» [7] для анализа взаимосвязи между эндогенными переменными, а также анализ откликов на импульсные воздействия (*Impulse Response Analysis*) [8].

3. Используемые данные и задачи исследования

Для построения моделей применяются временные ряды индексов цен с июня 2017 г. по декабрь 2024 г., предоставляемые Национальным статистическим комитетом Республики Беларусь, Федеральной службой государственной статистики Российской Федерации, а также официальные обменные курсы белорусского рубля Национального банка Республики Беларусь. В представляемых моделях используются следующие обозначения:

- PPI_RB_t – сезонно-скорректированный временной ряд индекса цен производителей продукции в промышленности РБ («месяц к месяцу» в %);
- CPI_RB_t – сезонно-скорректированный временной ряд индекса потребительских цен РБ («месяц к месяцу» в %);
- PPI_RU_t – временной ряд индекса цен производителей промышленной продукции РФ («месяц к месяцу» в %) без сезонной корректировки;
- $RUR\_BYN\_D_t$ и $USD\_BYN\_D_t$ – временные ряды ежедневных темпов роста («день ко дню», в %) официальных обменных курсов белорусского рубля по отношению к российскому рублю и доллару США.

По причине высокой степени интеграции белорусской и российской экономик [9], а также значительной доли объемов промышленного производства в структуре ВВП наибольший интерес представляет индекс цен производителей продукции в промышленности. В связи с этим решаются следующие задачи: 1) анализ взаимосвязи между индексами цен производителей продукции в промышленности Беларуси и России; 2) краткосрочное прогнозирование и наукастинг индекса PPI_RB на основе альтернативных типов моделей.

4. Анализ прогностической способности моделей и причинно-следственной связи между индексами

Для оценки прогностической способности построенных моделей используются ретроспективные и вневыборочные прогнозы. Вневыборочные ежемесячные прогнозы строятся на весь 2024 год с помощью алгоритма *скользящего расширяющегося окна* (*sliding expanding window*), который последовательно исключает прогнозный месяц из периода оценивания модели. Прогнозы на основе моделей MIDAS, MIDAS-GETS и MF-VAR по смешанным данным получаются в режиме наукастинга, то есть с учетом внутримесячной динамики ежедневных изменений обменных курсов. Традиционные модели авторегрессии ARX и VARX [10] строят краткосрочные прогнозы по среднемесячным изменениям обменных курсов. В табл. 1 для целевого показателя PPI_RB приводятся характеристики точности ретроспективных и вневыборочных прогнозов RMSE, MAE и MAPE [10] на основе одномерных моделей.

Таблица 1

Точность прогнозов для целевого показателя PPI_RB

<i>Ретроспективные прогнозы</i>			
Показатель	MIDAS	MIDAS-GETS	ARX
RMSE	0,2973	0,1753	0,2883
MAE	0,2411	0,1284	0,2352
MAPE	0,2396	0,1276	0,2338
<i>Вневыборочные прогнозы</i>			
Показатель	MIDAS	MIDAS-GETS	ARX
RMSE	0,2438	0,2382	0,2812
MAE	0,2166	0,2031	0,2350
MAPE	0,2156	0,2020	0,2336

Анализ причинно-следственной связи между индексами PPI_RB, CPI_RB и PPI_RU проводится на основе моделей векторной авторегрессии MF-VAR и VARX с использованием теста Грейнджера [7]. В табл. 2 представлены результаты, полученные на основе модели MF-VARX.

Таблица 2

Тестирование причинной зависимости на основе модели MF-VAR

Зависимая переменная	Независимая переменная (P-значение)			
CPI_RB	PPI_RB (0,5805)	PPI_RU (0,0367)	USD_BYN (0,0018)	RUR_BYN (0,0755)
PPI_RB	CPI_RB (0,0089)	PPI_RU (0,0424)	USD_BYN (0,0513)	RUR_BYN (0,0001)
PPI_RU	CPI_RU (0,0391)	PPI_RB (0,5534)	USD_BYN (0,2761)	RUR_BYN (0,0597)

Формально тест Грейнджера для рассматриваемого периода времени приводит к следующим выводам: изменения потребительских цен являются причиной изменения цен производителей продукции в промышленности; общим «третьим» фактором для обоих индексов в белорусской экономике являются цены производителей в промышленности России; изменения обменных курсов основных валют имеют опережающий характер для обоих индексов цен в белорусской экономике.

5. Анализ откликов на импульсные шоки

Результаты анализа откликов на импульсные шоки [8] для модели MF-VAR (рис. 1) приводят к следующим выводам: реакция индекса PPI_RB на шок со стороны индекса PPI_RU достигает максимальной силы

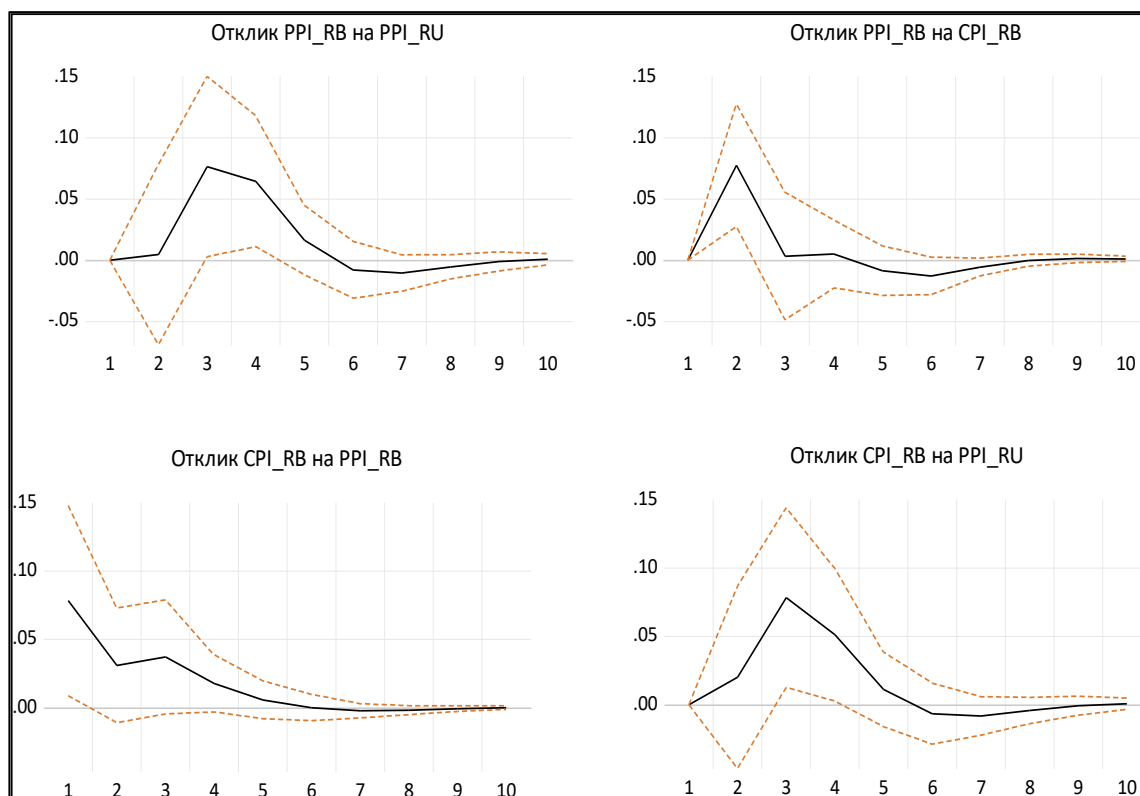


Рис. 1. Анализ откликов на импульсные шоковые воздействия:
синяя линия – отклик переменной; красные линии – доверительный интервал

в течение второго месяца с последующим периодом затухания до шести месяцев; рост индекса PPI_RU максимально проявляется в росте индекса CPI_RB в течение двух месяцев; рост индекса PPI_RB в ответ на шок со стороны индекса CPI_RB достигает максимального уровня в течение первого месяца, после чего эффект шока быстро затухает.

6. Заключение

Модели по данным смешанной частоты демонстрируют более высокую точность ретроспективных и вневыборочных прогнозов белорусских индексов цен по сравнению с моделями, использующими агрегированные данные. Анализ причинно-следственных связей позволяет сделать вывод о том, что в рассматриваемый период времени (до конца 2024 года) индекс цен производителей промышленной продукции России являлся высоко значимым опережающим фактором для обоих индексов цен в экономике Беларуси. Изменения обменных курсов российского рубля и доллара США при этом также имеют опережающий характер для обоих индексов цен в белорусской экономике. Более полное описание моделей и полученных результатов дается в [11].

Библиографические ссылки

1. *Andreou E., Ghysels E., Kourtellos A.* Regression models with mixed sampling frequencies // *Journal of Econometrics*. 2010. № 58. P. 246–261.
2. *Giannone D., Reichlin L., Small D.* Nowcasting: The real-time informational content of macroeconomic data // *Journal of Monetary Economics*. 2008. Vol. 55, no. 4. P. 665–676.
3. *Macroeconomic Nowcasting and Forecasting with Big Data / B. Bok [et al.]* // *Annual Review of Economics*. 2018. Vol 10, iss. 1. P. 615–643.
4. *Малюгин В. И.* Краткосрочное прогнозирование и наукастинг темпов роста инфляции на основе моделей по смешанным данным // *Банковский вестник*. 2024. № 1/726. С. 23–36.
5. *Campos J. Ericsson N., Hendry D.* General-to-specific Modeling: An Overview and Selected Bibliography // *International Finance Discussion Paper*. 2005. Article no. 838.
6. *Foroni C., Marcellino M.* A survey of econometric methods for mixed frequency data // *Working Paper 2013/06*. Norges Bank, 2013.
7. *Granger C. W. J.* Investigating Causal Relations by Econometric Models and Cross-spectral Methods // *Econometrica*. 1969. № 37(3). P. 424–438.
8. *Sims C. A.* Policy Analysis with Econometric Models // *Brookings Papers on Economic Activity*. 1985. Vol. 1. P. 107–164.
9. *Малюгин В. И., Новопольцев А. Ю.* Связь темпов роста экономик Беларуси и России в условиях шоков: эконометрический анализ и прогнозирование // *Экономика. Моделирование. Прогнозирование*. 2022. Вып. 16. С. 236–250.
10. *Харин Ю. С., Малюгин В. И., Харин А. Ю.* Эконометрическое моделирование. Минск : БГУ. 2003.
11. *Малюгин В. И.* Модели по данным смешанной частоты и их применение в причинно-следственном анализе и наукастинге индексов цен производителей и потребителей // *Банковский вестник*. 2025. № 8/745. С. 3–17.

МЕТОДИКА ТОЧНОГО ИЗМЕРЕНИЯ ТЕМПЕРАТУРЫ ТОЧКИ РОСЫ С ПОМОЩЬЮ СОВОКУПНОСТИ ЛИНЕЙНЫХ УРАВНЕНИЙ

С. Г. Ольков, О. А. Олькова

*Казанский инновационный университет им. В. Г. Тимирязова,
Казань, Россия, olkovsg@mail.ru*

В статье представлена новейшая методика измерения температуры точки росы, необходимая для составления точных государственных стандартов, сводов правил, технических регламентов, инструкций в различных отраслях народного хозяйства.

Ключевые слова: точка росы; физика; экономика; метеорология; авиация; строительство; сельское хозяйство; судебно-строительная экспертиза; безопасность; строительные работы; эксперимент; матрица; методика; расследование преступлений.

METHODOLOGY FOR ACCURATE MEASUREMENT OF DEW POINT TEMPERATURE USING A SET OF LINEAR EQUATIONS

S. G. Olkov, O. A. Olkova

*Kazan Innovative University named after V. G. Timiryasov,
Kazan, Russia, olkovsg@mail.ru*

The article presents the newest methodology of dew point temperature measurement, which is necessary for drawing up accurate state standards, codes of practice, technical regulations, instructions in different sectors of the economy.

Keywords: dew point; physics; economics; meteorology; aviation; construction; agriculture; forensic construction expertise; safety; construction works; experiment, matrix; methodology; crime investigation.

1. Введение

Определение точки росы имеет важное значение при решении широкого спектра задач во многих областях жизнедеятельности человека.

В настоящее время существуют специальные таблицы, с помощью которых определяют точку росы [1]. Указанные таблицы были получены эмпирическим путем с помощью проведения физических экспериментов, и содержат дискретные значения температуры точки росы. При этом как

температура окружающей среды, так и относительная влажность воздуха также взяты дискретно с большим или меньшим шагом. Например, для относительной влажности воздуха шаг обычно составляет 5%.

Учеными из различных отраслей знания предпринимались попытки разработать более точную методику измерения точки росы. Подробный обзор таких попыток приведен в работе Марка Лоуренса [2]. Успехом эти попытки увенчаны не были ввиду применения неадекватного математического аппарата.

Авторами данной статьи предлагается получить более точные результаты при определении точки росы и работать с непрерывными данными. Этот результат достигается путем снятия линии уровня с поверхности температуры точки росы и получения 49 линейных уравнений с высочайшей степенью точности, позволяющих измерять температуру точку росы в непрерывном режиме. Было получено 14 линейных уравнений, функционально описывающих температуру точки росы при непрерывной температуре воздуха и фиксированных значениях относительной влажности воздуха и 35 линейных уравнений, функционально описывающих температуру точки росы, при непрерывной относительной влажности воздуха и фиксированных значениях температуры воздуха.

Линейные уравнения, описывающие значение температуры точки росы, являются практически функциональными, и получить более точные результаты другим способом невозможно, поскольку в полученных уравнениях устраняются (гасятся) все возможные ошибки, которые могли возникнуть при проведении конкретных экспериментов.

2. Теоретическая часть

В табл. 1 представлена температура точки росы в зависимости от относительной влажности воздуха и температуры воздуха в табличном (дискретном виде). Это таблица размером 35 строк на 14 столбцов, содержащая 490 возможных значений температуры точки росы. Могут быть и другие подобные таблицы с большим количеством строк и столбцов. Их использование не повлияет на качество проведенного нами исследования и точность полученных нами научных результатов, ибо для корректного применения метода наименьших квадратов относительно пространственных табличных данных достаточно 7 наблюдений на независимую переменную, а у нас всюду применяется одна независимая переменная – это либо температура воздуха, либо относительная его влажность. То есть фактических наблюдений у нас более чем достаточно. При установлении функциональной связи между температурой точки росы и относительной

влажностью воздуха мы использовали по 35 наблюдений, а при установлении функциональной зависимости между температурой точки росы и температурой воздуха – по 14 наблюдений.

Таблица 1

Таблица определения температуры точки росы в зависимости от температуры воздуха и относительной влажности воздуха

Температура точки росы °C															
Относительная влажность воздуха		30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%
Температура воздуха		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	-10°C	-23,2	-21,8	-20,4	-19	-17,8	-16,7	-15,8	-14,9	-14,1	-13,3	-12,6	-11,9	-10,6	-10
2	-5°C	-18,9	-17,2	-15,8	-14,5	-13,3	-11,9	-10,9	-10,2	-9,3	-8,8	-8,1	-7,7	-6,5	-5,8
3	0°C	-14,5	-12,8	-11,3	-9,9	-8,7	-7,5	-6,2	-5,3	-4,4	-3,5	-2,8	-2	-1,3	-0,7
4	+2°C	-12,8	-11	-9,5	-8,1	-6,8	-5,8	-4,7	-3,6	-2,6	-1,7	-1	-0,2	-0,6	1,3
5	+4°C	-11,3	-9,5	-7,9	-6,5	-4,9	-4	-3	-1,9	-1	0	0,8	1,6	2,4	3,2
6	+5°C	-10,5	-8,7	-7,3	-5,7	-4,3	-3,3	-2,2	-1,1	-0,1	0,7	1,6	2,5	3,3	4,1
7	+6°C	-9,5	-7,7	-6	-4,5	-3,3	-2,3	-1,1	-0,1	0,8	1,8	2,7	3,6	4,5	5,3
8	+7°C	-9	-7,2	-5,5	-4	-2,8	-1,5	-0,5	0,7	1,6	2,5	3,4	4,3	5,2	6,1
9	+8°C	-8,2	-6,3	-4,7	-3,3	-2,1	-0,9	0,3	1,3	2,3	3,4	4,5	5,4	6,2	7,1
10	+9°C	-7,5	-5,5	-3,9	-2,5	-1,2	0	1,2	2,4	3,4	4,5	5,5	6,4	7,3	8,2
11	+10°C	-6,7	-5,2	-3,2	-1,7	-0,3	0,8	2,2	3,2	4,4	5,5	6,4	7,3	8,2	9,1
12	+11°C	-6	-4	-2,4	-0,9	0,5	1,8	3	4,2	5,3	6,3	7,4	8,3	9,2	10,1
13	+12°C	-4,9	-3,3	-1,6	-0,1	1,6	2,8	4,1	5,2	6,3	7,5	8,6	9,5	10,4	11,7
14	+13°C	-4,3	-2,5	-0,7	0,7	2,2	3,6	5,2	6,4	7,5	8,4	9,5	10,5	11,5	12,3
15	+14°C	-3,7	-1,7	0	1,5	3	4,5	5,8	7	8,2	9,3	10,3	11,2	12,1	13,1
16	+15°C	-2,9	-1	0,8	2,4	4	5,5	6,7	8	9,2	10,2	11,2	12,2	13,1	14,1
17	+16°C	-2,1	-0,1	1,5	3,2	5	6,3	7,6	9	10,2	11,3	12,2	13,2	14,2	15,1
18	+17°C	-1,3	0,6	2,5	4,3	5,9	7,2	8,8	10	11,2	12,2	13,5	14,3	15,2	16,6
19	+18°C	-0,5	1,5	3,2	5,3	6,8	8,2	9,6	11	12,2	13,2	14,2	15,3	16,2	17,1
20	+19°C	0,3	2,2	4,2	6	7,7	9,2	10,5	11,7	13	14,2	15,2	16,3	17,2	18,1
21	+20°C	1	3,1	5,2	7	8,7	10,2	11,5	12,8	14	15,2	16,2	17,2	18,1	19,1
22	+21°C	1,8	4	6	7,9	9,5	11,1	12,4	13,5	15	16,2	17,2	18,1	19,1	20
23	+22°C	2,5	5	6,9	8,8	10,5	11,9	13,5	14,8	16	17	18	19	20	21
24	+23°C	3,5	5,7	7,8	9,8	11,5	12,9	14,3	15,7	16,9	18,1	19,1	20	21	22
25	+24°C	4,3	6,7	8,8	10,8	12,3	13,8	15,3	16,5	17,8	19	20,1	21,1	22	23
26	+25°C	5,2	7,5	9,7	11,5	13,1	14,7	16,2	17,5	18,8	20	21,1	22,1	23	24
27	+26°C	6	8,5	10,6	12,4	14,2	15,8	17,2	18,5	19,8	21	22,2	23,1	24,1	25,1
28	+27°C	6,9	9,5	11,4	13,3	15,2	16,5	18,1	19,5	20,7	21,9	23,1	24,1	25	26,1
29	+28°C	7,7	10,2	12,2	14,2	16	17,5	19	20,5	21,7	22,8	24	25,1	26,1	27
30	+29°C	8,7	11,1	13,1	15,1	16,8	18,5	19,9	21,3	22,5	22,8	25	26	27	28
31	+30°C	9,5	11,8	13,9	16	17,7	19,7	21,3	22,5	23,8	25	26,1	27,1	28,1	29
32	+32°C	11,2	13,8	16	17,9	19,7	21,4	22,8	24,3	25,6	26,7	28	29,2	30,2	31,1
33	+34°C	12,5	15,2	17,2	19,2	21,4	22,8	24,2	25,7	27	28,3	29,4	31,1	31,9	33
34	+36°C	14,6	17,1	19,4	21,5	23,2	25	26,3	28	29,3	30,7	31,8	32,8	34	35,1
35	+38°C	16,3	18,8	21,3	23,4	25,1	26,7	28,3	29,9	31,2	32,3	33,5	34,6	35,7	36,9

На наш взгляд, идеальным аналитическим решением для определения зависимости температуры точки росы от температуры воздуха и относительной влажности воздуха будет снятие линии уровня с поверхности температуры точки росы (рис. 1). То есть мы рассматриваем температуру точки росы в зависимости от температуры воздуха при фиксированных значениях относительной влажности воздуха и зависимость температуры

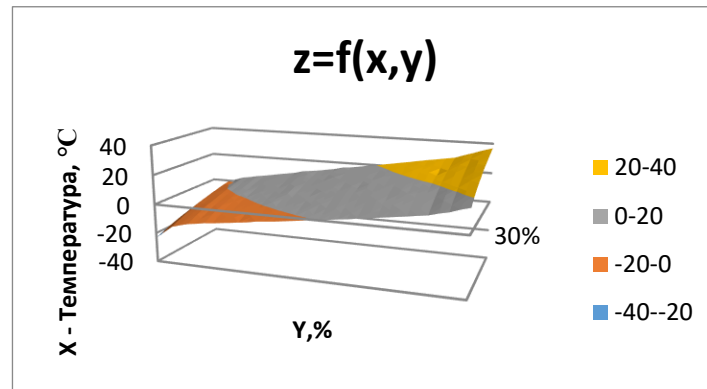


Рис. 1. Поверхность температуры точки росы в зависимости от температуры воздуха и относительной влажности воздуха

точки росы от относительной влажности воздуха при фиксированных значениях температуры воздуха.

По эмпирическим (опытным) данным получена практически идеальная зависимость температуры точки росы от температуры воздуха (рис. 2), имеющая ясную физическую интерпретацию – если при относительной влажности воздуха равной 30% температура воздуха меняется на 1 градус Цельсия, то температура точки росы меняется в среднем на 0,806 градусов Цельсия.

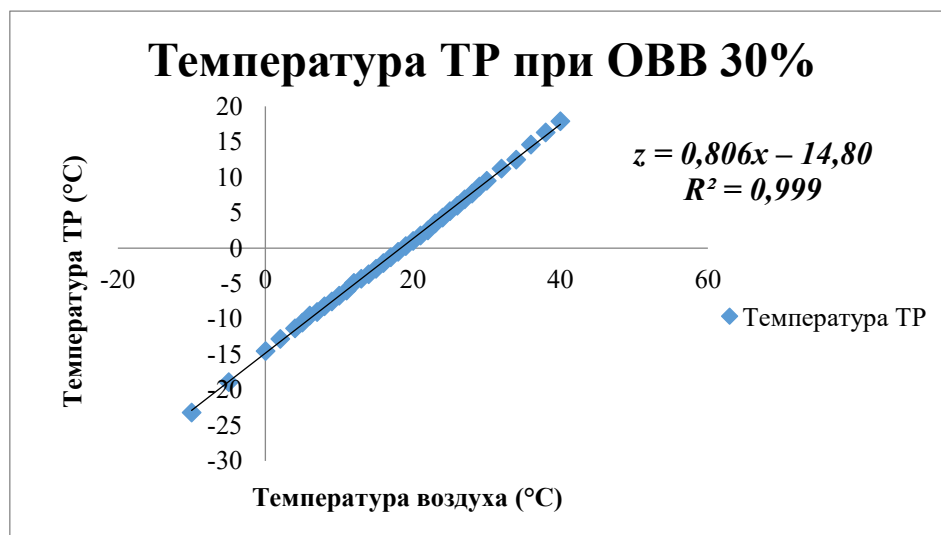


Рис. 2. Зависимость температуры точки росы от температуры воздуха при фиксированной относительной влажности воздуха равной 30%

Зависимость является положительной и практически функциональной (коэффициент корреляции между переменными модели практически равен 1).

Коэффициент детерминации практически равен единице. Еще ближе к единице ($R^2 = 0,999$) линейный коэффициент корреляции Пирсона ($r = \sqrt{0,999} = 0,9995$), что говорит о сильной положительной корреляционной связи между исследуемыми переменными близкой к идеальной функциональной связи.

Зависимости при фиксированных значениях относительной влажности воздуха приведены в табл. 2. Проверка качества на примере одной из полученных моделей представлена в табл. 3.

Таблица 2

Аналогичные зависимости для фиксированных значений относительной влажности воздуха и непрерывных значениях температуры воздуха

1) $z_{30} = 0,806x - 14,8.$	6) $z_{55} = 0,904x - 7,872.$	11) $z_{80} = 0,966x - 3,119.$
2) $z_{35} = 0,86x - 13,26.$	7) $z_{60} = 0,916x - 6,752.$	12) $z_{85} = 0,978x - 2,356.$
3) $z_{40} = 0,852x - 11,65.$	8) $z_{65} = 0,932x - 5,781.$	13) $z_{90} = 0,983x - 1,534.$
4) $z_{45} = 0,872x - 10,27.$	9) $z_{70} = 0,944x - 4,828.$	14) $z_{95} = 0,988x - 0,653.$
5) $z_{50} = 0,888x - 8,99.$	10) $z_{75} = 0,952x - 3,934.$	

Таблица 3

Проверка качества полученного оценочного уравнения регрессии при фиксированной относительной влажности воздуха 45% в эконометрическом статистическом пакете *Gretl*. $z_{45} = 0,872x - 10,27$

	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	-10,2735	0,0814126	-126,2	<0,0001	***
X (температура)	0,872137	0,00400121	218,0	<0,0001	***
Среднее зав. перемен	4,328571		Ст. откл. зав. перемен		10,23430
Сумма кв. остатков	2,471841		Ст. ошибка модели		0,273686
R-квадрат	0,999306		Испр. R-квадрат		0,999285
F(1, 33)	47510,23		P-значение (F)		1,05e-53
Лог. правдоподобие	-3,281115		Крит. Акаике		10,56223
Крит. Шварца	13,67293		Крит. Хеннана-Куинна		11,63604

Примечание. Модель 2: МНК, использованы наблюдения 1–35. Зависимая переменная: **Z (45%)** Из таблицы видно, что параметры уравнения (коэффициенты) значимы на 1% уровне значимости, т.е. вероятность ошибки не превышает 1%. Коэффициент корреляции практически равен единице, о чем можно судить, взяв корень квадратный из коэффициента детерминации (**R-квадрат**), что говорит в пользу, практически, функциональной связи между переменными. Судя по коэффициенту регрессии, связь является положительной. Согласно статистике Фишера, уравнение в целом статистически значимо.

Таким образом, прогноз по данному оценочному уравнению регрессии является идеальным.

На рис. 3 и в табл. 4 приведена точная зависимость температуры точки росы от относительной влажности воздуха при фиксированных значениях температуры воздуха в виде совокупности линейных уравнений по эмпирическим табличным данным.

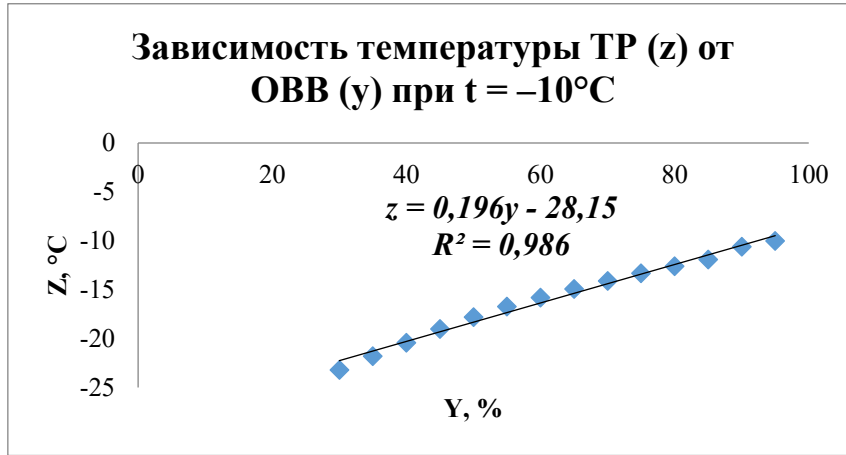


Рис. 3. Зависимость температуры точки росы от относительной влажности воздуха при фиксированной температуре воздуха равной минус 10 градусов Цельсия

Таблица 4

Уравнения зависимости температуры точки росы от относительной влажности воздуха при фиксированных значениях температуры воздуха

1) $z_{(-10)} = 0,196y - 28,15$ ($R^2 = 0,986$).	19) $z_{(18)} = 0,266y - 7,157$ ($R^2 = 0,983$).
2) $z_{(-5)} = 0,192y - 23,84$ ($R^2 = 0,972$).	20) $z_{(19)} = 0,270y - 6,475$ ($R^2 = 0,984$).
3) $z_{(0)} = 0,209y - 19,55$ ($R^2 = 0,973$).	21) $z_{(20)} = 0,271y - 5,596$ ($R^2 = 0,981$).
4) $z_{(2)} = 0,205y - 17,63$ ($R^2 = 0,972$).	22) $z_{(21)} = 0,273y - 4,844$ ($R^2 = 0,981$).
5) $z_{(4)} = 0,215y - 16,47$ ($R^2 = 0,981$).	23) $z_{(22)} = 0,274y - 3,960$ ($R^2 = 0,978$).
6) $z_{(5)} = 0,218y - 15,86$ ($R^2 = 0,984$).	24) $z_{(23)} = 0,277y - 3,146$ ($R^2 = 0,979$).
7) $z_{(6)} = 0,219y - 14,84$ ($R^2 = 0,985$).	25) $z_{(24)} = 0,278y - 2,307$ ($R^2 = 0,98$).
8) $z_{(7)} = 0,224y - 14,48$ ($R^2 = 0,985$).	26) $z_{(25)} = 0,282y - 1,603$ ($R^2 = 0,981$).
9) $z_{(8)} = 0,228y - 13,92$ ($R^2 = 0,989$).	27) $z_{(26)} = 0,285y - 0,774$ ($R^2 = 0,98$).
10) $z_{(9)} = 0,234y - 13,34$ ($R^2 = 0,989$).	28) $z_{(27)} = 0,285y - 0,079$ ($R^2 = 0,981$).
11) $z_{(10)} = 0,239y - 12,82$ ($R^2 = 0,986$).	29) $z_{(28)} = 0,289y - 0,763$ ($R^2 = 0,98$).
12) $z_{(11)} = 0,241y - 12,04$ ($R^2 = 0,988$).	30) $z_{(29)} = 0,287y - 1,716$ ($R^2 = 0,981$).
13) $z_{(12)} = 0,249y - 11,48$ ($R^2 = 0,99$).	31) $z_{(30)} = 0,295y + 2,351$ ($R^2 = 0,977$).
14) $z_{(13)} = 0,253y - 10,8$ ($R^2 = 0,987$).	32) $z_{(32)} = 0,297y + 4,089$ ($R^2 = 0,981$).
15) $z_{(14)} = 0,253y - 10,07$ ($R^2 = 0,986$).	33) $z_{(34)} = 0,305y + 5,105$ ($R^2 = 0,983$).
16) $z_{(15)} = 0,256y - 9,357$ ($R^2 = 0,985$).	34) $z_{(36)} = 0,306y + 7,189$ ($R^2 = 0,982$).
17) $z_{(16)} = 0,261y - 8,698$ ($R^2 = 0,985$).	35) $z_{(38)} = 0,305y + 9,044$ ($R^2 = 0,979$).
18) $z_{(17)} = 0,267y - 8,076$ ($R^2 = 0,986$).	

Пусть нас интересует значение точки росы при 12 градусах Цельсия и относительной влажности воздуха равной 53%. Берем уравнение: $z_{(12)} = 0,249y - 11,48$ и вместо y подставляем значение 53, то есть

записываем $z_{(12)} = 0,249 \cdot 53 - 11,48 = 1,717$, что близко к табличному значению 1,6 при относительной влажности воздуха 50%.

3. Прикладная часть

Методика точного аналитического измерения температуры точки росы

1. Пусть нас интересует значение температуры точки росы при температуре воздуха равной 10°C и относительной влажности воздуха 45%. Берем уравнение $z_{45} = 0,872x - 10,27$ и подставляем в него значение тем-

пературы воздуха: $z_{45} = 0,872 \cdot 10 - 10,27 = -\frac{31}{20} \approx -1,55^\circ\text{C}$. В данном слу-

чае отклонение от табличного значения, равного $-1,7$, составляет $e = 1,7 - (-1,55) = 0,15$, или 8,8%. Учитывая большую точность корреляционно-регрессионного уравнения, полученного по длинному пространственному ряду, можно списать разницу на погрешность экспериментальных измерений и считать, что расчет по уравнению является более точным. Или

можно взять среднее значение: $\frac{1,7 + 1,55}{2} = 1,625$. Здесь уже отличие от

табличного значения составляет 4,4% $\left(\frac{1,7 - 1,625}{1,7} \cdot 100\% = \frac{75}{17} \approx 4,4\% \right)$.

2. Пусть нас интересует значение точки росы при температуре воздуха $25,5^\circ\text{C}$ и относительной влажности воздуха 73%. Берем два уравнения: 1) $z_{70} = 0,944x - 4,828$ и 2) $z_{75} = 0,952x - 3,934$, и находим среднее

значение его коэффициентов, получив новое уравнение: $z_{73} = \frac{0,944 + 0,952}{2}x - \frac{4,828 + 3,934}{2}$, то есть $z_{73} = \frac{237}{250}x - \frac{4381}{1000}$, или в десятич-

ных дробях: $z_{73} = 0,948x - 4,381$, откуда $z_{73} = 0,948 \cdot 25,5 - 4,381 =$

$= 19,793^\circ\text{C}$. Если посмотреть в табличные значения, то для относительной влажности воздуха 75% и температуры воздуха 25°C температура точки росы равна 20°C , а для 70% относительной влажности воздуха при темпе-

ратуре воздуха 25°C она составит 18,8. То есть $\frac{20 + 18,8}{2} = 19,4^\circ\text{C}$.

Линейные уравнения температуры точки росы мы можем получить, как при постоянной относительной влажности воздуха (при изменении температуры воздуха), так и при постоянной температуре воздуха (при изменении относительной влажности воздуха).

Спрашивается, а как определить температуру точки росы при относительной влажности воздуха, например, 33%, 88% и так далее? –

В этом случае можно усреднить коэффициенты: 1) свободный член; 2) коэффициент пропорциональности двух ближайших точек, в пределах которых находится относительная влажность воздуха. Например, для относительной влажности воздуха 88% нам нужно усреднить коэффициенты для относительной влажности воздуха 85 и 90%, и использовать эти коэффициенты для определения точки росы. Можно также комбинировать 14 и 35 уравнений, при отыскании конкретного значения температуры точки росы на поверхности температуры точки росы в ее конкретной координате, сужая окрестность вокруг оцениваемой координаты.

Преимущество предлагаемой методики помимо прочего состоит в том, что осуществлен переход от дискретных табличных значений к непрерывным функциям, описывающим бесконечное множество теоретически возможных комбинаций температуры точки росы по относительной влажности воздуха и температуре воздуха. Так, таблица эмпирических данных размером 35X14 дает всего 490 возможных значений точки росы при дискретных значениях относительной влажности воздуха и температуры воздуха. А предлагаемые уравнения дают бесконечное множество возможных вариантов с высочайшей точностью оценки температуры точки росы при непрерывных значениях относительной влажности воздуха и температуры воздуха и не ограничивают область определения функции по температуре воздуха и относительной влажности воздуха.

Библиографические ссылки

1. Свод правил по проектированию и строительству проектирование тепловой защиты зданий СП 23-101-2004 / Группа Ж24, ОКС 91.120.01, Дата введения 01.06.2004.
2. *Lawrence M. G.* The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications // The Bulletin of the American Meteorological Society (BAMS). 2005. Vol. 86, iss. 2. P. 225–233.
3. *Alduchov O. A., Eskridge R. E.* Improved Magnus form approximation of saturation vapor pressure / J. Appl. Meteor. 1996. Iss. 35. P. 601–609.
4. *Magnus G.* Versuche über die Spannkkräfte des Wasserdampfs. Ann. // Phys. Chem. 1844. Iss. 61. P. 225–247.

ОБЪЕМ ИНФОРМАЦИИ В МАТРИЦАХ ПРЯМЫХ ЗАТРАТ РАЗНОЙ РАЗМЕРНОСТИ В МОДЕЛИ ЛЕОНТЬЕВА «ЗАТРАТЫ – ВЫПУСК»

В. А. Пархименко¹⁾, А. А. Быков²⁾

¹⁾ Белорусский государственный университет информатики и радиоэлектроники,
Минск, Беларусь, parkhimenko@bsuir.by

²⁾ Белорусский государственный экономический университет,
Минск, Беларусь, bikov_a@bseu.by

Рассматривается проблема размерности матрицы прямых затрат в модели Леонтьева с точки зрения содержащегося в матрице объема информации о реальных экономических явлениях и процессах. Предлагается концептуальный подход к получению оценок объема информации в матрицах прямых затрат разного уровня детализации / агрегирования относительно «истинного» объема информации в экономической деятельности субъектов хозяйствования.

Ключевые слова: информация; межотраслевой баланс; модель Леонтьева; таблицы «затраты – выпуск»; экономическое моделирование.

THE INFORMATION CONTENT OF DIRECT REQUIREMENTS MATRICES OF VARYING DIMENSIONS IN THE LEONTIEF INPUT-OUTPUT MODELS

U. A. Parkhimenka^{a)}, A. A. Bykau^{b)}

^{a)} Belarusian State University of Informatics and Radioelectronics,
Minsk, Belarus, parkhimenko@bsuir.by

^{b)} Belarusian State Economic University,
Minsk, Belarus, bikov_a@bseu.by

This paper addresses the issue of the dimensionality of the direct requirements matrix in the Leontief model from the perspective of the amount of information it contains about real economic phenomena and processes. An approach is proposed for estimating the information content in direct requirements matrices of different levels of detail/aggregation relative to the ‘true’ amount of information in the economic activities of companies.

Keywords: information; input-output table; Leontief model; input-output analysis; economic modeling.

1. Введение

Матрица прямых в модели Леонтьева «затраты – выпуск» [1] представляет собой квадратную матрицу, в ij -й ячейке которой указывается величина прямых затрат продукции (товаров и услуг) i -й отрасли, потребляемой в j -й отрасли на единицу выпуска (1 рубль) продукции j -й отрасли.

Обычная практика для прикладных экономических расчетов использовать подобного вида матрицу размерностью от нескольких десятков до нескольких сотен отраслей. Например, в Республике Беларусь в настоящее время размерность таблиц «затраты – выпуск» (на основе которых рассчитывается матрица прямых затрат и другие компоненты модели Леонтьева) составляет 83×83 , в Российской Федерации – 99×99 , в США – 398×398 .

Очевидно, что более детализированные таблицы содержат значительно больше информации об экономических явлениях и процессах, наблюдаемых в национальной экономике за годовой период. Вместе с тем создание детализированных таблиц требует значительных расходов ресурсов и времени.

В этом контексте научный интерес вызывают оценки объема информации в матрицах прямых затрат разной размерности в модели Леонтьева «затраты – выпуск».

2. Оценка «истинного» объема информации

Предложим концептуальный подход к осмыслению и поиску решения для этой задачи.

Попробуем представить «истинный» объем информации, который «содержится» непосредственно в самих реально протекающих экономических процессах. Для этого от уровня отрасли (совокупности предприятий) перейдем к наименьшему ее компоненту – одному предприятию.

Пусть в данный момент в стране есть N организаций (компаний, предприятий, учреждений), которые, помимо выпуска конечной продукции и услуг для потребления, выпускают так называемые промежуточные товары и оказывают услуги, которые затем потребляются как ресурсы в других производственных процессах.

Если представить, что каждая организация осуществляет поставки своей продукции каждой другой организации, то этот процесс можно представить таблицей транзакций (потоков товаров и услуг) в стоимостном выражении размерностью $N \times N$. Число потенциальных транзакций, таким образом, будет равно N^2 .

Из этой таблицы путем деления (по каждому j -му столбцу) абсолютных значений на соответствующие выпуски продукции j -й отрасли

($j = \overline{1, N}$) можно перейти к матрице прямых затрат максимально возможной размерности (степени детализации), назовем ее условно «истинной».

Пусть в каждой ячейке матрицы мы можем записать число от 0 до 1 с точностью в шесть знаков после запятой. Тогда объем информации, который потенциально будет содержаться в одной ячейке, $-\log_2 1000001 \approx \approx 19,9 \approx 20$ бит. А во всей матрице – $20 \cdot N^2$ бит.

По данным Белстата, в 2024 году в Беларуси были зарегистрировано 130 463 коммерческих организаций, в том числе 11 384 малых организаций, 2 093 средних и 1 598 крупных, т.е. без учета так называемых микро-организаций – 15 075.

Отталкиваясь от общего числа коммерческих организаций и нереалистично допуская наличие транзакций между всеми компаниями, получим оценку информации, содержащейся в «истинной» матрице: $20 \cdot N^2 \approx \approx 20 \text{ бит} \cdot 17,02 \text{ млрд ячеек} \approx 39,49 \text{ Гбайт}$.

При переходе к таблицам меньшей размерности, например, 83×83 , как это принято для составления таблиц «затраты – выпуск» в Республике Беларусь по числу видов экономической деятельности (ВЭД), получаем объем информации, равный $20 \cdot 83^2 = 20 \text{ бит} \cdot 6889 \text{ ячеек} \approx 16,82 \text{ Кбайт}$, т.е. всего лишь $4 \cdot 10^{-5} \%$ от информации, содержащейся в истинной матрице.

Очевидно, однако, что в реальной действительности наша «истинная» матрица будет содержать огромное количество нулевых ячеек, так как ни одна из наугад выбранных организаций не будет взаимодействовать с каждой другой организацией без исключения, а лишь с некоторым их числом. Прежде всего потому, что технологический процесс отдельного производства требует только определенного набора входных ресурсов (например, для производства автотракторной техники нужен металл и пластмассы, но не продукции животноводства). Кроме того, в одно и то же время на рынке находится множество аналогичных компаний-конкурентов, которые, как правило, не работают с одними и теми же поставщиками или клиентами.

Вряд ли в настоящее время возможны точные оценки на этом «истинном» детализированном уровне матрицы прямых затрат, так как такие данные не собираются и не анализируются статистическими органами. (Возможно, при построении цифровой экономики и полному отражению всех экономических потоков и транзакций в централизованной информационной системе такие данные будут доступны государству как архитектору и регулятору цифровой экономики.) Неизвестны и статистические характеристики и распределения этих данных на этом уровне.

Предложим, однако, некоторый эвристический подход, который, с нашей точки зрения, позволят сделать хотя бы грубые и приблизительные оценки.

Будем исходить из того допущения, что можно охарактеризовать всех экономических субъектов (в данном случае коммерческие организации) средним числом поставщиков и клиентов – μ . Отталкиваясь от данных Белстата за 2024 год и предполагаемого числа поставщиков / клиентов у разного типа организаций, рассчитаем средневзвешенное значение параметра μ (табл. 1).

Таблица 1

Оценка параметра μ

Тип организаций	Количество субъектов	Предполагаемое среднее число поставщиков / клиентов (μ)
Микроорганизации (до 15 человек)	115 388	10
Малые организации (от 16 до 100 человек)	11 384	100
Средние организации (до 250 человек)	2 093	1 000
Крупные организации (свыше 250 человек)	1 598	10 000
Коммерческие организации, ВСЕГО	130 463	≈ 156

Если отталкиваться от числа 156 из табл. 1, означающего, что в среднем на одну организацию приходится 156 поставщиков/клиентов, то количество отличных от нуля ячеек в «истинной» матрице прямых затрат будет равно не N^2 , а $\mu \cdot N$ (в каждом столбце – μ , а во всей матрице – $\mu \cdot N$) и количество информации, соответственно, – $20 \cdot \mu \cdot N$ бит, что для случая белорусской экономики дает нам оценку в 49 688,1 Кбайт.

3. Оценка объема информации в матрицах прямых затрат разной размерности

Далее предположим, что матрицы, чья размерность во много раз меньше размерности «истинной» матрицы (в случае белорусской экономики мы полагаем, что речь может идти о размерности в несколько сотен отраслей) и примерно соответствует сложившейся статистической практики разработки таблиц «затраты – выпуск», – для таких матриц ввиду их существенной агрегированности объем информации может быть оценен уже по первоначальной формуле, т.е. как $\approx 20 \cdot N^2$, так как количество нулевых ячеек будет крайне незначительным.

В табл. 2 сведены расчетные оценки объема информации в матрицах прямых затрат в модели Леонтьева при разной их размерности и при указанном выше допущении.

Таблица 2

Оценка объема информации в матрицах разной размерности

Размерность матрицы (N)	Потенциальный объем информации, Кбайт	Объем информации от «истинного» объема
130 463	49 688,1	100,0%
500	610,4	1,23%
250	152,6	0,31%
100	24,4	0,05%
83	16,8	0,03%

Полученное значение 0,03% информации, содержащейся в отчетственных таблицах «затраты – выпуск» (точнее говоря, в таком элементе, как матрица прямых затрат), на первый взгляд, вряд ли может рассматриваться как достаточное. И даже матрицы размерностью около 500 строк/столбцов не позволяют говорить о «репрезентации» большого объема информации по отношению к «истинной» матрице прямых затрат.

С другой стороны, экономическая наука часто оперирует такими агрегированными показателями, как ВВП, индекс потребительских цен, уровень безработицы, которые представляют собой просто скалярные значения. Использование системы взаимоувязанных количественных показателей, например, в форме модели Леонтьева «затраты – выпуск», в любом случае представляет собой шаг в сторону более полного описания реальных экономических явлений и процессов.

4. Заключение

Полученные оценки (см. табл. 2), конечно, являются очень грубыми и предварительными.

Остаются не более чем экспертными предположениями такие «логические звенья» рассмотренного выше подхода, как оценка параметра $\mu = 156$, так и утверждение, что для матриц размерностью меньшей 1000 количество нулевых ячеек пренебрежительно мало.

Видны и направления для развития и уточнения предложенного подхода.

Агрегирование таблиц меньшей размерности из таблиц большей размерности получается путем суммирования соответствующих ячеек. Это может приводить к необходимости отведения большего количества бит на хранение информации в каждой ячейке. Таким образом, эта тенденция

будет противодействовать (как представляется, незначительно) общему уменьшению объема информации в таблицах меньшей размерности.

Следует учесть информацию об импортных товарах и услугах.

Кроме того, таблицы «затраты – выпуск» содержат и другие компоненты: конечный спрос по категориям институциональных единиц (домохозяйства, государственные учреждения и т.д.).

Крайне важный вопрос (не затрагиваемый в данном исследовании) критерия построения самих таблиц: по отраслевому принципу или продуктовому.

Наконец, важно учесть статистическое распределение значений в матрице прямых затрат и оценить информационную энтропию по Шеннону [2]. Даже если в качестве эвристического предположения можно предположить, что те 156 значений в каждом столбце (или строке) матрицы прямых затрат имеют распределение Парето (это довольно традиционное предположение для многих процессов в экономике), то вполне можно утверждать, что объем информации, измеряемый как информационная энтропия по Шеннону, будет для «истинного» случая значительно меньше, чем указано в табл. 2.

Библиографические ссылки

1. *Miller R. E., Blair P. D.* Input-output analysis: foundations and extensions. Cambridge university press, 2022.
2. *Shannon C. E.* A Mathematical Theory of Communication // Bell System Technical Journal. 1948. Vol. 27, iss. 3. P. 379–423.

МЕТОДОЛОГИЯ РАСЧЕТА ВЫСОКОЧАСТОТНЫХ СУБ-ИНДЕКСОВ ИПЦ С ПРИМЕНЕНИЕМ ВЕБ-СКРЕЙПИНГА И КЛАССИФИКАЦИИ ДАННЫХ

Т. Т. Сафиуллин

*Белорусский государственный университет,
Минск, Беларусь, SafiullinTT@bsu.by*

В статье представлена методика расчета высокочастотных суб-индексов потребительских цен на основе данных интернет-ритейла Беларуси. Описаны разработка и применение ETL-пайплайна для веб-скрейпинга товаров с онлайн-платформ, их очистки и предварительной обработки. Основное внимание уделено задаче автоматической классификации товаров по категориям потребительской корзины НСК РБ с использованием методов машинного обучения. Результаты работы формируют основу для создания оперативной системы мониторинга инфляции.

Ключевые слова: веб-скрейпинг; индекс потребительских цен; машинное обучение; классификация товаров.

METHODOLOGY FOR CALCULATING HIGH-FREQUENCY CPI SUB-INDEXES USING WEB SCRAPING AND DATA CLASSIFICATION

T. T. Safiullin

*Belarusian State University,
Minsk, Belarus, SafiullinTT@bsu.by*

The abstract should be informative. The article presents a methodology for calculating high-frequency consumer price sub-indexes based on data from online retailers in Belarus. It describes the development and application of an ETL pipeline for web scraping product data from online platforms, including its cleaning and preprocessing. The primary focus is on the task of automatic product classification according to the categories of the consumer basket defined by the National Statistical Committee of Belarus. The results form the basis for creating a real-time inflation monitoring system.

Keywords: web scraping; consumer price index; machine learning; product classification.

1. Введение

Современные экономические исследования все чаще опираются на анализ веб-данных, позволяющий получать актуальную и детализированную информацию о рыночных процессах. Одним из ключевых инструментов сбора таких данных является веб-скрейпинг — автоматизированное

извлечение информации с веб-сайтов. Как отмечают исследователи, этот метод становится все более популярным среди национальных статистических служб для расчета официальных ценовых показателей [1]. Особую ценность представляет возможность ежедневного мониторинга изменений, что позволяет получать высокочастотные данные для оценки воздействия экономической политики и поддержки принятия решений [2].

Актуальность веб-скрейпинга в экономических исследованиях подтверждается его успешным применением в международной практике. Исследования демонстрируют устойчивую связь между онлайн- и оффлайн-ценами [3–5], что делает веб-данные ценным источником для построения суб-индексов потребительских цен (ИПЦ) [6], прогнозирования статистических показателей [7] и выявления изменений инфляционных тенденций [8]. Достоинством суб-индексов является возможность получения оценки инфляционных процессов до появления официальных значений индексов цен, рассчитываемых национальными статистическими органами. Например, в статье [2] показано, как веб-скрейпинг данных крупных интернет-магазинов и торговых сетей помогает выявлять расхождения между официальной и реальной инфляцией в Аргентине. В работе [9] авторы используют скрейпинг для подсчета фактического ИПЦ и анализа ценовой реакции на экономические санкции в России. Эти примеры демонстрируют, что автоматизированный сбор данных с торговых онлайн-платформ обеспечивает более оперативный и точный мониторинг экономических показателей по сравнению с традиционными методами.

В Беларуси, как показано в работе В. И. Малюгина [10], применение веб-скрейпинга и методов наукастинга позволяет осуществлять мониторинг потребительской корзины на основе веб-данных о ценах в реальном времени. Автор демонстрирует, что использование веб-данных в сочетании со статистическими методами и машинным обучением дает возможность не только отслеживать текущую стоимость продовольственной корзины и бюджета прожиточного минимума, но и прогнозировать их динамику для различных категорий населения. Однако ручной сбор данных затруднен из-за большого объема информации и ее постоянного обновления. Основная проблема при использовании веб-скрейпинга заключается в получении неструктурированных данных, требующих сложной обработки перед анализом.

Цель данной работы состоит в представлении результатов выполнения первых этапов вычисления суб-индексов на основе веб-данных, включая: автоматизированный сбор и обработка данных о товарах, представленных в торговых сетях, классификация товаров по категориям потребительской корзины в Беларуси.

В рамках исследования решаются следующие задачи:

- разработка парсера для сбора веб-данных с белорусских торговых онлайн-платформ;
- очистка и предварительная обработка неструктурированных данных;
- разметка данных для обучения моделей классификации;
- классификация товаров по категориям Национального статистического комитета Республики Беларусь (НСК РБ) с использованием методов машинного обучения.

2. Методология сбора данных

Начальным этапом исследования является сбор структурированных данных с популярных белорусских онлайн-платформ e-dostavka.by и green-dostavka.by, выбранных благодаря их рыночной значимости и репрезентативному товарному ассортименту. Для реализации процесса сбора данных разработана специализированная ETL-система на языке Python.

При работе со статическим контентом используется стек технологий, включающий библиотеки Requests для HTTP-запросов и BeautifulSoup для парсинга DOM-структуры. BeautifulSoup обеспечивает гибкий поиск элементов через CSS-селекторы, позволяя точно извлекать ключевые атрибуты: наименование товара, цену, категорию и уникальный артикул, служивший первичным ключом для интеграции данных.

Анализ сайта e-dostavka.by показывает необходимость обработки динамически генерируемого контента. Для решения этой задачи применяется библиотека Selenium WebDriver в headless-режиме Chrome, что обеспечило корректное исполнение клиентских скриптов и доступ к динамическим элементам DOM.

В ходе реализации пайплайна решаются несколько критических проблем: битые ссылки и нестабильности DOM-структуры. Для обработки битых ссылок внедряется система валидации URL с экспоненциальным backoff, где невалидные ссылки маркировались специальным флагом. Проблема нестабильности DOM-структуры решается через алгоритм избыточного поиска элементов с использованием комбинации CSS-селекторов. Для товаров со скидками реализуется двухэтапная верификация ценовых данных с исключением некорректных записей из финального набора.

Особое внимание уделяется обеспечению воспроизводимости процесса сбора данных. Реализованная система логирования и контрольных точек позволяет отслеживать прогресс сбора, возобновлять процесс после прерываний и гарантировать целостность данных при сетевых сбоях. В результате получается устойчивый пайплайн, обеспечивающий стабильное поступление актуальных рыночных данных для последующего анализа. Собранный корпус данных включает более 23 000 товарных позиций

с еженедельным обновлением и создает надежную основу для моделирования потребительской корзины.

3. Предварительная обработка и категоризация данных

На этапе первичной обработки данных последовательно выполняются ключевые процедуры очистки и нормализации. Первоначально из выборки исключаются записи с отсутствующими ключевыми признаками (название товара, цена), поскольку их включение может привести к искажению аналитических результатов. Далее устраняются дубликаты, возникающие из-за особенностей категориальной структуры исходных платформ, где один товар может относиться к нескольким пересекающимся категориям. Эта процедура позволяет сформировать уникальный и непротиворечивый набор данных.

Важным этапом становится фильтрация нерелевантных категорий, не соответствующих целям исследования. В частности, исключаются сезонные товары (новогодние украшения, праздничная атрибутика и т.д.), поскольку их включение нарушает репрезентативность данных для анализа устойчивых потребительских тенденций.

Процесс категоризации включает несколько последовательных этапов:

1. автоматическое сопоставление категорий магазинов с официальной классификацией НСК РБ на основе векторных представлений и расчета косинусного сходства;
2. создание словаря соответствий для однозначно идентифицируемых категорий;
3. разработка правила для обработки неоднозначных случаев на основе анализа ключевых слов в названиях товаров.

Результатом предварительной обработки становятся два независимых набора данных:

- green-dostavka.by: 13 096 записей;
- e-dostavka.by: 9 994 записей.

Каждый набор содержит единообразную структуру:

- название товара;
- цена товара;
- исходная категория товара в магазине;
- соответствующая категория НСК РБ.

Анализ распределения категорий (рис. 1) выявляет существенный дисбаланс классов, что требует применения специальных метрик оценки качества на этапе классификации. В исследовании используются F1-мера (гармоническое среднее precision и recall) и ROC-AUC (площадь под характеристической кривой), которые демонстрируют устойчивость к

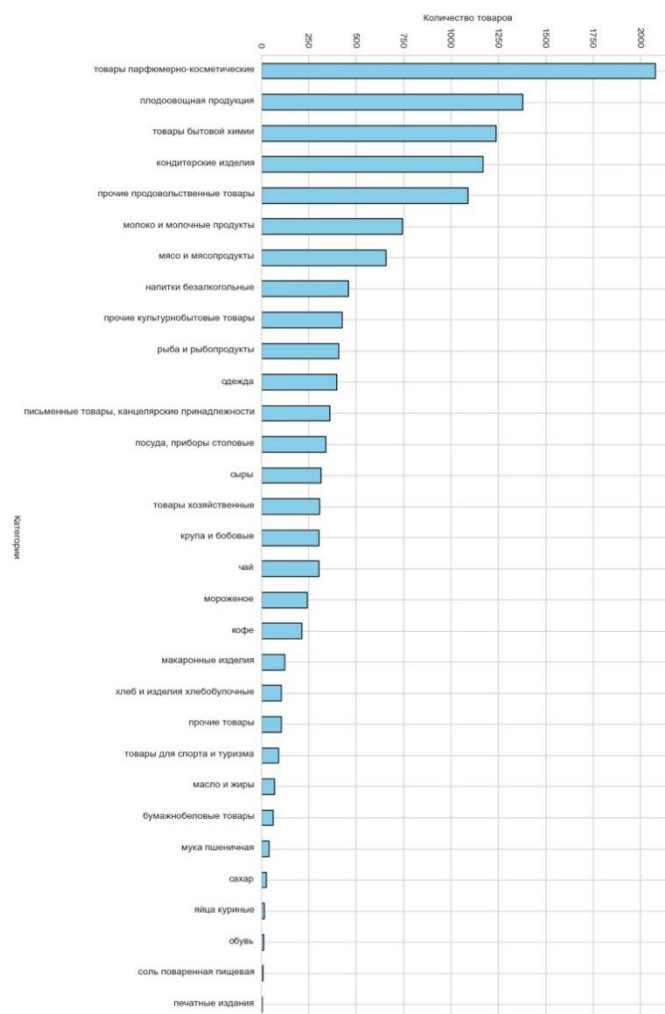


Рис. 1. Распределение категорий НСК РБ

дисбалансу классов в отличие от стандартной метрики ассигасу. Эти метрики позволяют получить объективную оценку эффективности модели, минимизируя влияние доминирующих категорий на итоговые показатели.

4. Экспериментальное исследование эффективности алгоритмов классификации товаров

В рамках исследования эффективности различных подходов к классификации товаров был проведен комплексный анализ на данных белорусских онлайн-ритейлеров. Первоначальный эксперимент использовал данные платформы green-dostavka.by, где в качестве входных признаков применялось комбинированное текстовое представление, объединяющее название товара и его магазинную категорию. Этот признак подвергался TF-IDF векторизации перед разделением на стандартные обучающую и тестовую выборки в соотношении 80 на 20.

Тщательная оптимизация гиперпараметров методом решетчатого поиска проводилась для пяти различных алгоритмов машинного обучения: логистическая регрессия, случайный лес, градиентный бустинг, SVM с RBF-ядром и ансамблевая модель, включающая SVM и XGBoost. Сравнительный анализ показал, что все модели продемонстрировали исключительно высокое качество классификации по метрике ROC-AUC (рис. 2), превышающее 0,99, что свидетельствует об отличной разделяющей способности всех рассмотренных алгоритмов для данной задачи.

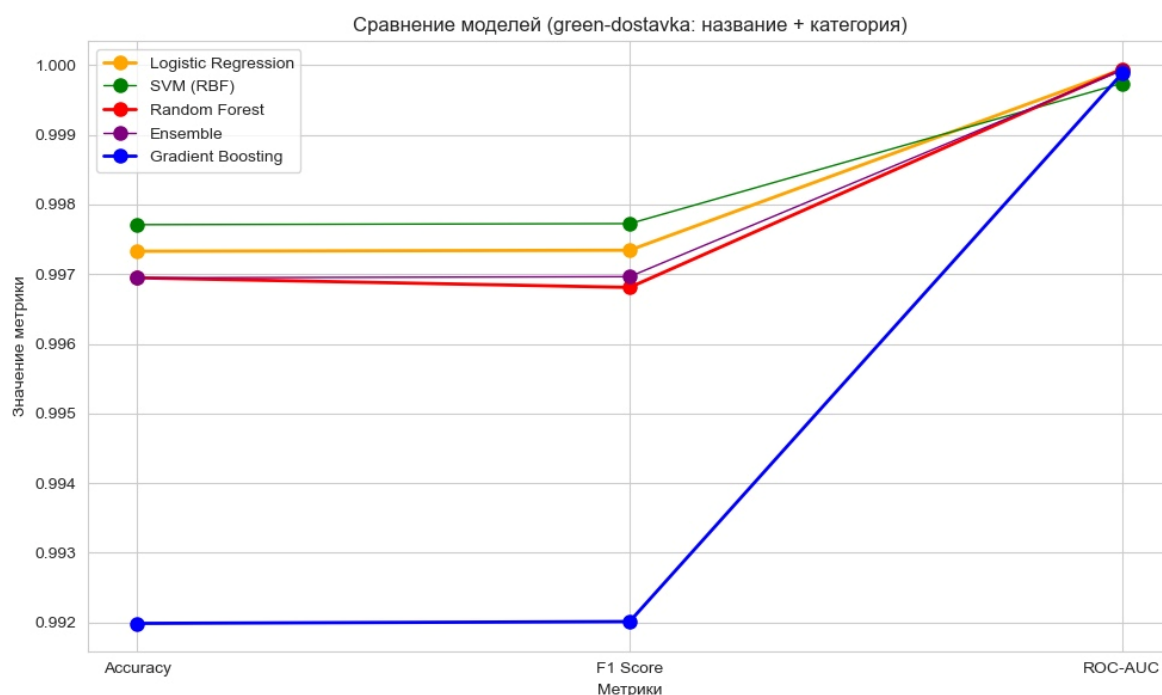


Рис. 2. График сравнения моделей, обученных на названиях и категориях

Наиболее сбалансированные результаты по комплексной метрике F1 и ассигасу продемонстрировали SVM, логистическая регрессия и ансамблевая модель. Примечательно, что градиентный бустинг, несмотря на высокий ROC-AUC, показал относительно более низкие значения F1-меры, что может объясняться его повышенной чувствительностью к дисбалансу классов в обучающей выборке. Это наблюдение подчеркивает важность выбора соответствующих метрик оценки в зависимости от специфики задачи.

Дальнейшее исследование было сосредоточено на разработке универсального классификатора, использующего только названия товаров без привязки к конкретной категориальной структуре магазина. Такой подход показал несколько сниженные, но все равно весьма высокие показатели качества ($F1 > 0,85$), при этом обеспечивая существенно более широкую применимость решения. В этом сценарии SVM и ансамблевая модель

вновь подтвердили свою эффективность, тогда как случайный лес показал более скромные результаты, особенно по F1-мере (рис. 3).

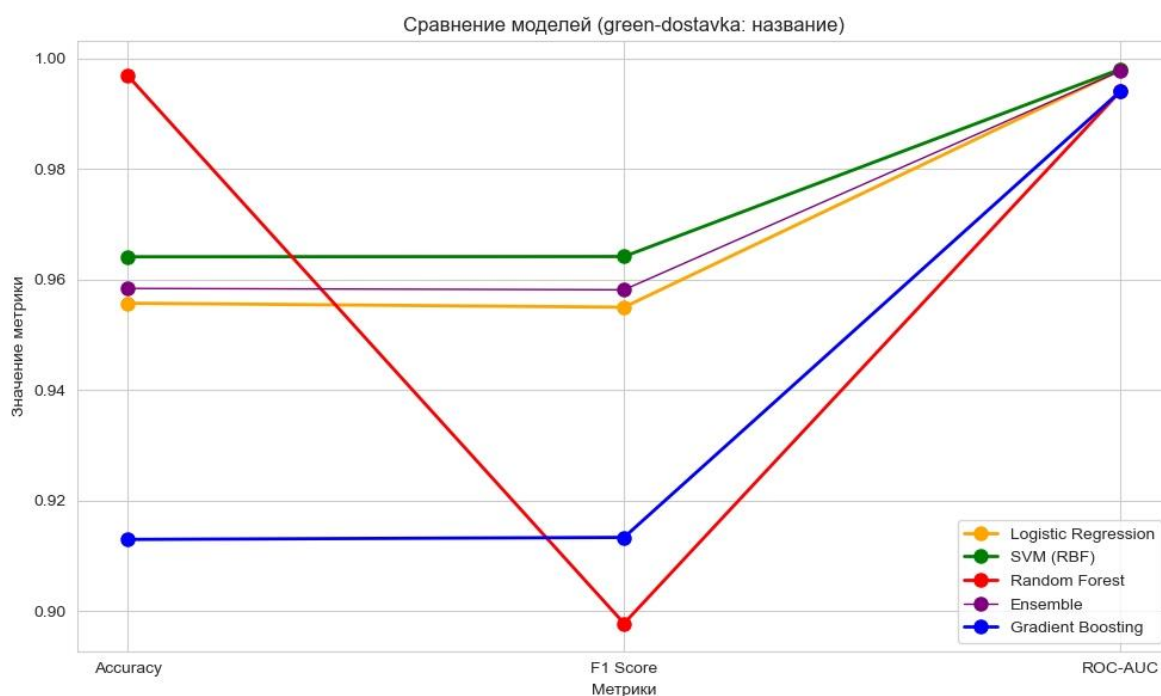


Рис. 3. График сравнения моделей обученных на названиях

Критически важным этапом стала валидация лучших моделей на независимых данных с платформы e-dostavka.by. Устойчивость показателей качества ($\text{ROC-AUC} > 0,99$) для всех протестированных моделей свидетельствует о хорошей обобщающей способности разработанных решений и их применимости для различных онлайн-ритейлеров (рис. 4). Особенно впечатляющим оказалось сохранение высокого качества при переходе к универсальной классификации только по названиям товаров, что открывает перспективы создания кросс-платформенного решения.

5. Заключение

Проведенное исследование позволяет сделать следующие ключевые выводы:

- веб-скрейпинг в сочетании с методами машинного обучения доказал свою эффективность для автоматической классификации товаров потребительской корзины в Беларуси;
- наилучшие результаты показали SVM и ансамблевые модели, продемонстрировавшие высокую точность и устойчивость к дисбалансу классов;

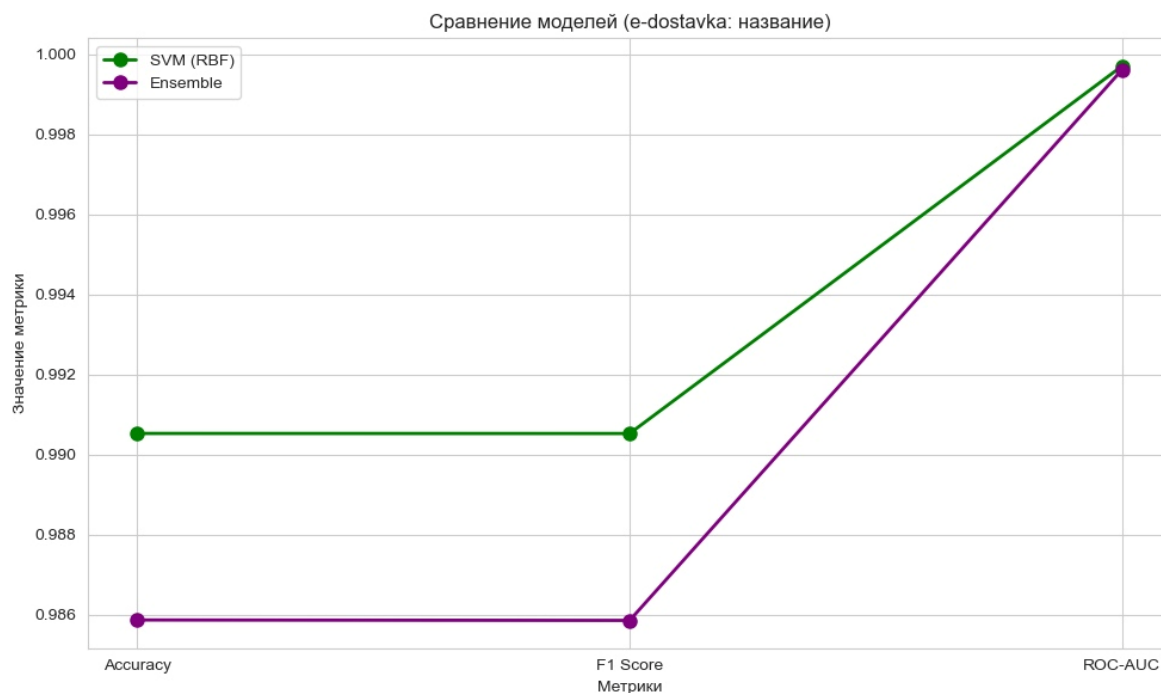


Рис. 4. Проверка лучших моделей на новом наборе данных

- универсальные классификаторы, работающие только с названиями товаров, обеспечивают кросс-платформенную применимость при сохранении высокого качества классификации.

На основе полученных результатов можно сформулировать следующие практические рекомендации:

- для максимально точной классификации в рамках одной платформы рекомендуется использовать SVM с комбинированными признаками (название + категория);
- для кросс-платформенных решений оптимальным выбором являются ансамблевые модели, работающие только с названиями товаров;
- особое внимание при внедрении следует уделить обработке дисбаланса классов и регулярному обновлению обучающих данных.

Важным следующим этапом данного исследования станет расчет индекса потребительских цен на основе собранных и классифицированных веб-данных. Это позволит:

- создать оперативную систему мониторинга инфляционных процессов;
- обеспечить альтернативный источник для верификации официальной статистики;
- реализовать методологию наукастинга для краткосрочного прогнозирования ценовой динамики.

Полученные результаты имеют важное значение для совершенствования систем мониторинга потребительской корзины и могут быть использованы государственными органами статистики, аналитическими центрами и ритейл-компаниями для повышения качества экономического анализа и прогнозирования.

Библиографические ссылки

1. *Eurostat*. Practical Guidelines on Web Scraping for the HICP. 2020. URL: <https://ec.europa.eu/eurostat/documents/272892/12032198/Guidelines-web-scraping-HICP-11-2020.pdf/> (date of access: 09.09.2025).
2. *Cavallo A.* Online and official price indexes: Measuring Argentina's inflation // *Journal of Monetary Economics*. 2013. Vol. 60, iss. 2. P. 152–165.
3. *Cavallo A.* Are online and offline prices similar? Evidence from large multi-channel retailers // *American Economic Review*. 2017. Vol. 107, iss. 1. P. 283–303.
4. *Cavallo A.* Scraped data and sticky prices // *Review of Economics and Statistics*. 2018. Vol. 100, iss. 1. P. 105–119.
5. E-commerce and price setting: evidence from Europe / G. Strasser [et al.] // *Occasional Paper Series*. No. 320. European Central Bank, 2023. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4514760 (date of access: 09.09.2025).
6. *Harchaoui T. M., Janssen R. V.* How can big data enhance the timeliness of official statistics? // *International Journal of Forecasting*. 2018. Vol. 34, iss. 2. P. 225–234.
7. *Aparicio D., Bertolotto M. I.* Forecasting inflation with online prices // *International Journal of Forecasting*. 2020. Vol. 36, iss. 2. P. 232–247.
8. *Cavallo A., García Zavaleta G.* Detecting structural breaks in inflation trends: A high-frequency approach / HBS Working Paper, Harvard Business School, 2023.
9. *Benchimol J., Palumbo L.*, Sanctions and Russian Online Prices // *Journal of Economic Behavior & Organization*. 2024. Vol. 225. P. 483–521.
10. *Малюгин В. И., Якубович А. В.* Анализ и прогнозирование стоимости потребительской корзины в режиме реального времени // *Экономика. Моделирование. Прогнозирование*. 2020. Вып. 14. С. 235–241.

ТЕОРЕТИЧЕСКОЕ ИССЛЕДОВАНИЕ И ЭКСПЕРИМЕНТАЛЬНАЯ ПРОВЕРКА АЛГОРИТМОВ РАСПОЗНАВАНИЯ, ОСНОВАННЫХ НА МЕРЕ ПРЕЦЕДЕНТНОСТИ

М. А. Филиппов

*Белорусский государственный университет,
Минск, Беларусь, filippov.maxim52@gmail.com*

Основная цель данного исследования — реализовать алгоритмы распознавания, основанные на мере прецедентности, и сравнить их с другими алгоритмами классификации. В результате работы были решены три задачи классификации, на основе которых были проведены сравнения точности и временной эффективности применённых алгоритмов. Для автоматизации процесса тестирования предложен проект соответствующей системы.

Ключевые слова: мера прецедентности; функция сравнения; решающее правило; классификация; автоматизированное тестирование.

THEORETICAL STUDY AND EXPERIMENTAL VERIFICATION OF RECOGNITION ALGORITHMS BASED ON THE MEASURE OF PRECEDENCE

M. A. Filippov

*Belarusian State University,
Minsk, Belarus, filippov.maxim52@gmail.com*

This study investigates metric algorithms based on the measure of precedence. The main purpose is to implement these algorithms and to compare their performance to other classification algorithms. As a result of this work, three classification problems were solved using different algorithms. Accuracy and time efficiency of the applied algorithms were compared. To automate the process of testing, a project of the corresponding system is proposed.

Keywords: measure of precedence; comparison function; decision rule; classification; automated testing.

1. Введение

В современном мире вопросы эффективного распознавания и классификации данных являются актуальными. Имеет место ряд алгоритмов, основанных на мерах сходства, прецедентности и подобия, которые могут

использоваться для её решения. В данном исследовании подробно рассматриваются алгоритмы, основанные на мере прецедентности, которая предполагает анализ предыдущих примеров (прецедентов) для принятия решений в новых ситуациях. Кроме того, будут рассмотрены теоретические аспекты данной меры, а также проведены экспериментальные исследования для оценки эффективности алгоритмов, основанных на ней.

Для использования алгоритмов распознавания в практических задачах необходимо экспериментально подтвердить их работоспособность и определить область их применения. Поэтому задача разработки соответствующей системы, способной проводить комплексные эксперименты, представляет несомненный интерес как с теоретической, так и с практической точки зрения. Концепция такой системы будет предложена в рамках данной работы.

2. Общий алгоритм

Задача классификации формулируется следующим образом:

Пусть задано множество объектов X , разбитое на подмножества (классы) $X_1, \dots, X_l$ ($l \in \mathbb{N}$). Классы не пересекаются, т.е. $X_i \cap X_j = \emptyset \quad \forall i \neq j$ ($i, j \in \{1, \dots, l\}$). Из заданного множества выбирается подмножество объектов $X^0 \subset X$, удовлетворяющее условию $X^0 \cap X_i \neq \emptyset \quad (\forall i \in \{1, \dots, l\})$. Дополнительно для каждого объекта $x \in X^0$ известна информация о его принадлежности к одному из классов $X_1, \dots, X_l$. Эта информация задается в виде информационного вектора $P(x) = (P_1(x), \dots, P_l(x))$, компоненты которого задаются следующим образом:

$$P_i(x) = \begin{cases} 1, & \text{если } x \in X_i, \\ 0, & \text{иначе.} \end{cases}$$

Общий алгоритм решения задачи, основанного на мере прецедентности включает следующие шаги [5, с. 286].

Шаг 0. Разбиение выборки X^0 на две части. Этот шаг может повторяться неоднократно. Каждый раз множество X^0 разбивается на две части $X_{\text{обуч}}^0$ и $X_{\text{контр}}^0$, называемые соответственно обучающей и контрольной выборками.

Шаг 1. Определение функции попарного сравнения объектов из X (функции прецедентности) вида (1).

$$s : X \times X \rightarrow \mathbb{R}. \quad (1)$$

Каждый объект $x \in X$ сравнивается с объектами обучающей выборки $X_{\text{обуч}}^0$. В результате, объекту x может быть поставлен в соответствие вектор $(s(x, x_1), \dots, s(x, x_m))$, где первые m_1 компонент являются результатами сравнения объекта x с объектами из обучающей выборки $X_{1(\text{обуч})}^0$, следующие m_2 компонент являются результатами сравнения с объектами из $X_{2(\text{контр})}^0$ и т.д.

Шаг 2. Определение функции сравнения объектов из X и объектов из обучающего множества $X_{i(\text{обуч})}^0$ вида (2).

$$f_i : \mathbb{R}^m \rightarrow \mathbb{R}, \quad i \in \{1, \dots, l\}. \quad (2)$$

С помощью функции (2) каждому вектору $(s(x, x_1), \dots, s(x, x_m)) \in \mathbb{R}^m$ можно сопоставить вектор $(f_1(x), \dots, f_l(x)) \in \mathbb{R}^l$.

Шаг 3. Определение решающего правила P^A вида (3).

$$P^A : \mathbb{R}^l \rightarrow \mathbb{B}_2^l, \quad \mathbb{B}_2 = \{0, 1\}. \quad (3)$$

Вектор $P^A(x) = (P_1^A(x), \dots, P_l^A(x))$ обычно называют классификационным, т.к. его значения интерпретируются следующим образом: объект $x \in X$ заносится (или нет) алгоритмом A в класс X_i , если $P_i^A(x) = 1$ ($= 0$).

Таким образом, для построения конкретного алгоритма необходимо выбрать функции (1) и (2), а также решающее правило (3).

3. Выбор гиперпараметров

Чтобы определить необходимые для алгоритма функции, нужно вычислить некоторые параметры [5, с. 304].

Для всех классов $i \in \{1, \dots, l\}$ и для всех признаков $j \in \{1, \dots, n\}$ вычисляем

$$\begin{aligned} b_{ij} &= (m_i)^{-1} \sum_{x_t \in X_{i(\text{обуч})}^0} x_{tj}, \\ b_j &= l^{-1} \sum_{i=1}^l b_{ij}, \\ a_{ij} &= |b_{ij} - b_j|, \end{aligned} \quad (4)$$

где x_{tj} — значение признака с номером j в векторе $x_t \in X_{i(\text{обуч})}^0$ ($t \in \{1, \dots, m_i\}$).

Используя параметры (4), функции (1) – (3) могут быть определены следующими способами:

Алгоритм 1 (A₁)

Функция (1) зависит от параметров (4). В формуле ниже предполагается, что x_1 — произвольный объект из X , $x_2 \in X_{i(\text{обуч})}^0$, и для всех таких x_2 вычисляем:

$$s_1(x_1, x_2) = \left(\sum_{j=1}^n a_{ij} \right)^{-1} \times \left(\sum_{j=1}^n (-1)^u \times a_{ij} \right),$$

где

$$u = \begin{cases} 1, & \text{если } x_{1j} \neq x_{2j}, \\ 2, & \text{иначе.} \end{cases}$$

Для этого алгоритма функция (2) для каждого $i \in \{1, \dots, l\}$ определяется как

$$f_i(x) = \max_{x_t \in X_{i(\text{обуч})}^0} \{s_1(x, x_t)\},$$

а решающее правило (3) вычисляется на основе вычисления максимума расстояния до класса:

$$P_i^A(x) = \begin{cases} 1, & \text{если } f_i(x) = \max_{i \in \{1, \dots, l\}} \{f_1(x), \dots, f_l(x)\}, \\ 0, & \text{иначе.} \end{cases}$$

Алгоритм 2 (A₂)

Другой способ определить функции (1) – (3) — инвертировать соответствующие функции, описанные в алгоритме A₁. Из этой идеи следуют следующие формулы:

$$\begin{aligned} s_2(x_1, x_2) &= 1 - s_1(x_1, x_2), \\ f_i(x) &= \min_{x_j \in X_{i(\text{обуч})}^0} \{s_2(x, x_j)\}, \\ P_i^A(x) &= \begin{cases} 1, & \text{если } f_i(x) = \min_{i \in \{1, \dots, l\}} \{f_1(x), \dots, f_l(x)\}, \\ 0, & \text{иначе.} \end{cases} \end{aligned}$$

Алгоритм 3 (A₃)

В данном алгоритме функция (1) определяется через диаметр множества. Для каждого подмножества $X_i^0 = X^0 \cap X_i$ ($|X_i^0| = m_i$), используя

функцию $s_2(x_1, x_2)$, построенную в рамках алгоритма A_2 , диаметр множества X_i^0 вычисляется следующим образом:

$$d_i = \max_{(x_u, x_v) \in X_i^0 \times X_i^0} s_2(x_u, x_v).$$

Для полученного множества $(d_1, d_2, \dots, d_l)$ вычисляем

$$d = \max(d_1, d_2, \dots, d_l).$$

Используя полученный диаметр, функцию прецедентности определяем следующим образом:

$$s_3(x_1, x_2) = \max \left\{ 0, 1 - \frac{s_2(x_1, x_2)}{d} \right\}.$$

Функции (2) и (3) определяются аналогично алгоритму A_1 :

$$f_i(x) = \max_{x_j \in X_{i(\text{обуч})}^0} s(x, x_j),$$

$$P_i^A(x) = \begin{cases} 1, & \text{если } f_i(x) = \max_{i \in \{1, \dots, l\}} \{f_1(x), \dots, f_l(x)\}, \\ 0, & \text{иначе.} \end{cases}$$

4. Кодирование переменных

Вышеописанные алгоритмы могут работать только с данными, представленными в двоичной форме (все параметры объектов должны принимать значения либо 0, либо 1). Однако, используя различные методы кодирования, почти любую переменную можно представить в виде нескольких переменных со значениями 0 или 1.

Применение того или иного метода кодирования зависит от типа переменной. Существует два основных типа переменных: категориальные и количественные. В случае категориальной переменной может быть использован метод кодирования с одним активным состоянием (one-hot encoding). С количественными переменными всё немного сложнее, но можно использовать метод бинарного кодирования с некоторыми предварительными шагами в случае дробного типа.

Метод кодирования с одним активным состоянием [4] представляет собой способ преобразования категориальных переменных в двоичный

формат. Он создаёт новые столбцы для каждого возможного значения параметра и заполняет их нулями и единицами, где 1 означает наличие данной характеристики у объекта, а 0 — её отсутствие. Основной целью метода является обеспечение эффективного использования категориальных данных в моделях машинного обучения.

Существует две основные причины использования этого метода.

Во-первых, он устраняет порядковость. Многие категориальные переменные не имеют внутреннего порядка (например, переменная пол, принимающая значения «мужчина» или «женщина»). Если бы мы присваивали числовые значения (например, мужчина = 0, женщина = 1), модель могла бы ошибочно интерпретировать это как ранжирование, что могло бы привести к предвзятым прогнозам. Метод кодирования с одним активным состоянием устраняет этот риск, рассматривая каждую категорию независимо.

Во-вторых, он улучшает производительность модели, предоставляя более детализированное представление о категориальных переменных. Это позволяет моделям фиксировать сложные зависимости внутри данных, которые могли бы быть упущены, если бы категориальные переменные рассматривались как единые сущности.

Для кодирования количественной переменной можно использовать метод бинарного кодирования. Чтобы закодировать целочисленную переменную, её значения нужно представить в двоичной форме. Для каждого бита этой формы создаётся новая переменная. Бинарное кодирование осуществляется через присваивание каждой созданной переменной значения соответствующего бита.

Для кодирования дробной переменной её нужно умножить на 10^n в степени числа десятичных знаков. Это преобразует число с плавающей запятой в целое число, сохраняя его масштаб. Далее новая целочисленная переменная кодируется, как описано выше.

В отличие от предыдущего метода кодирования, бинарное кодирование сохраняет порядковость и масштаб. Например, значение 186 по-прежнему будет рассматриваться как большее, чем 19, что важно для количественных переменных.

5. Эксперименты и результаты

Для выполнения экспериментальной проверки описанных алгоритмов были выбраны несколько наборов данных с различными типами параметров.

Набор данных о сотрудниках [1] содержит информацию о работниках неизвестной компании. Цель этого набора данных — предсказать, был ли сотрудник уволен или нет, т.е. это задача бинарной классификации. Все переменные в наборе данных являются категориальными. Набор данных

также несбалансирован, что означает, что количество уволенных сотрудников значительно отличается от количества сотрудников, которые не были уволены.

Набор данных для предсказания сердечных заболеваний [2] содержит медицинскую информацию о некоторых пациентах. Цель состоит в том, чтобы определить, есть ли у них какие-либо сердечные заболевания или нет, т.е. это также задача бинарной классификации. Переменные в основном количественные. Набор данных слегка несбалансирован: пациентов с сердечными заболеваниями немного больше, чем здоровых.

Набор данных цен мобильных телефонов [3] содержит технические характеристики мобильных телефонов. Цель состоит в том, чтобы предсказать категорию стоимости телефона. Существует 4 возможные категории: низкая стоимость, средняя стоимость, высокая стоимость и очень высокая стоимость. Это означает, что данная задача относится к многоклассовой классификации. В наборе данных представлены как количественные, так и категориальные переменные. Набор данных идеально сбалансирован: количество мобильных телефонов в каждой категории стоимости одинаково.

Все вычисления проводились на процессоре Intel(R) Core(TM) i9-14900HX 2.20 GHz. Результаты экспериментов отражены в табл. 1–3. Лучшие результаты по каждому набору данных выделены полужирным шрифтом. Гиперпараметры для алгоритмов были подобраны методом поиска по сетке. Они отражены в табл. 4.

Таблица 1

Результаты применения различных алгоритмов машинного обучения к набору данных о сотрудниках

Алгоритм	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Время обучения (с)	Время тестирования (с)
A_1	69,62	57,49	59,80	58,62	0,0002	5,3051
A_2	69,62	57,49	59,80	58,62	0,0002	5,5125
A_3	69,98	58,13	59,30	58,71	5,1424	5,4723
Логистическая регрессия	70,52	69,71	67,29	67,68	0,0052	0,0008
Метод опорных векторов	77,40	78,58	74,00	74,85	0,5556	0,1573
Дерево решений	79,75	81,42	76,48	77,49	0,0038	0,0010
Случайный лес	77,58	79,84	73,70	74,61	0,0192	0,0021

Таблица 2

Результаты применения различных алгоритмов машинного обучения к набору данных о пациентах

Алгоритм	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Время обучения (с)	Время тестирования (с)
A_1	69,62	57,49	59,80	58,62	0,0002	5,3051
A_2	69,62	57,49	59,80	58,62	0,0002	5,5125
A_3	69,98	58,13	59,30	58,71	5,1424	5,4723
Логистическая регрессия	70,52	69,71	67,29	67,68	0,0052	0,0008
Метод опорных векторов	77,40	78,58	74,00	74,85	0,5556	0,1573
Дерево решений	79,75	81,42	76,48	77,49	0,0038	0,0010
Случайный лес	77,58	79,84	73,70	74,61	0,0192	0,0021

Таблица 3

Результаты применения различных алгоритмов машинного обучения к набору данных о ценах мобильных телефонов

Алгоритм	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Время обучения (с)	Время тестирования (с)
A_1	69,62	57,49	59,80	58,62	0,0002	5,3051
A_2	69,62	57,49	59,80	58,62	0,0002	5,5125
A_3	69,98	58,13	59,30	58,71	5,1424	5,4723
Логистическая регрессия	70,52	69,71	67,29	67,68	0,0052	0,0008
Метод опорных векторов	77,40	78,58	74,00	74,85	0,5556	0,1573
Дерево решений	79,75	81,42	76,48	77,49	0,0038	0,0010
Случайный лес	77,58	79,84	73,70	74,61	0,0192	0,0021

Таблица 4

Гиперпараметры, подобранные методом поиска по решётке

Алгоритм	Набор данных о сотрудниках	Набор данных о пациентах	Набор данных о ценах мобильных телефонов
Метод опорных векторов	<i>Kernel=RBF</i>	<i>Kernel=Linear</i>	<i>Kernel=Linear</i>

Алгоритм	Набор данных о сотрудниках	Набор данных о пациентах	Набор данных о ценах мобильных телефонов
Дерево решений	<i>Max tree depth=8</i>	<i>Max tree depth=3</i>	<i>Max tree depth=5</i>
Случайный лес	<i>Criterion=LogLoss</i> <i>Number of trees=13</i> <i>Max tree depth=8</i>	<i>Criterion=LogLoss</i> <i>Number of trees=18</i> <i>Max tree depth=5</i>	<i>Criterion=Entropy</i> <i>Number of trees=19</i> <i>Max tree depth=5</i>

Легко заметить, что дерево решений показало наилучший результат для двух из трёх задач. Этот алгоритм является не только самым точным, но и наиболее эффективным по времени. Однако задачу по определению стоимости мобильных телефонов лучше решил метод опорных векторов, что указывает на высокую линейную разделимость этих данных.

Что касается метрических алгоритмов, можно легко заметить, что A_1 и A_2 показали одинаковую точность. Это было предсказуемо, так как они не имеют значительных различий. Что касается A_3 , его качество классификации немного отличается при работе с набором данных о сотрудниках. Однако это повышение точности незначительно по сравнению со временем, которое требуется для обучения.

6. Заключение

В заключение следует отметить, что на текущем этапе исследования метрические алгоритмы малоэффективны по сравнению с другими алгоритмами машинного обучения. Необходимо провести дополнительные эксперименты с различными комбинациями функций сравнения и правил принятия решений. Это может повысить точность алгоритмов для определённого набора задач.

Для проведения таких экспериментов целесообразно реализовать автоматизированную систему тестирования. Эта система должна соответствовать нескольким требованиям. Во-первых, она должна работать с любыми табличными данными. Следовательно, необходимо, чтобы она автоматически адаптировалась к различным размерам данных и выполняла разведочный анализ данных, хотя бы его простейшие шаги: удаление дубликатов, обработку пропущенных значений, выявление возможных выбросов и т. д. Во-вторых, система должна автоматически кодировать все небинарные переменные, иначе метрические алгоритмы не смогут работать с такими данными. В-третьих, программное обеспечение должно позволять пользователю выбирать набор алгоритмов для тестирования. В идеале оно должно поддерживать метрические алгоритмы, классические алгоритмы машинного обучения и нейросетевые модели, разработанные с

использованием библиотек Keras или PyTorch, так как эти библиотеки являются наиболее популярными.

Наконец, система должна формировать табличный отчёт, отображающий производительность алгоритмов, выделять и сохранять лучшие модели. Из этого отчёта пользователь сможет понять, какой алгоритм лучше всего решает конкретную задачу.

Описанная система значительно упростит процесс тестирования алгоритмов. Она может существенно снизить требования к специалистам по машинному обучению, поскольку им не придётся изучать языки программирования для решения задач распознавания объектов. В будущем планируется реализация такой системы. Конечно, аналогичное программное обеспечение может быть разработано и для решения других задач машинного обучения, таких как регрессия, кластеризация и т. д.

Библиографические ссылки

1. Employee Dataset. URL: <https://www.kaggle.com/datasets/tawfikelmetwally/employee-dataset> (date of access: 29.03.2025).
2. Heart Prediction Dataset (Quantum). URL: <https://www.kaggle.com/datasets/shantanugarg274/heart-prediction-dataset-quantum> (date of access: 29.03.2025).
3. Mobile Price Classification Dataset. URL: <https://www.kaggle.com/datasets/iabhishekoofficial/mobile-price-classification> (date of access: 29.03.2025).
4. One Hot Encoding in Machine Learning. URL: <https://www.geeksforgeeks.org/ml-one-hot-encoding/> (date of access: 11.03.2025).
5. Выбранные науковыя працы Беларускага дзяржаўнага ўніверсітэта : У 7 т. / Адк. рэд. М. І. Юрчук. Мн.: БДУ, 2001. Т. 6. Матэматыка.

ПРОГНОЗИРОВАНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ РОССИЙСКОЙ ФЕДЕРАЦИИ НА ОСНОВЕ ГИБРИДНОГО ПОДХОДА

С. Т. Цаплин

*Московский государственный технический университет имени Н. Э. Баумана,
Москва, Россия, tsaplin@bmstu.ru*

Рассматривается задача кратко- и среднесрочного прогнозирования социально-экономических временных рядов Российской Федерации в условиях ограниченной длины наблюдений и структурной нестабильности. Предложена гибридная архитектура, сочетающая SARIMAX, XGBoost и Time-LLM (residual-коррекция), а также динамическое ансамблирование с инерционным сглаживанием весов.

Ключевые слова: прогнозирование временных рядов; гибридные модели; SARIMAX; XGBoost; большие языковые модели.

FORECASTING SOCIO-ECONOMIC INDICATORS OF THE RUSSIAN FEDERATION BASED ON A HYBRID APPROACH

S. T. Tsaplin

*Bauman Moscow State University,
Moscow, Russia, tsaplin@bmstu.ru*

The paper presents a hybrid system for forecasting Russian socio-economic time series under data scarcity and structural instability. The system integrates SARIMAX, XGBoost, and residual Time-LLM correction with dynamic ensemble weighting and inertia smoothing.

Keywords: time series forecasting; hybrid models; SARIMAX; XGBoost; large language models.

1. Введение

В современных условиях цифровизации социально-экономических процессов и их высокой волатильности задача надежного прогнозирования ключевых показателей приобретает особую актуальность. Многие из этих показателей, такие как данные о вакансиях по регионам или числе поступивших в учебные заведения, часто представлены короткими временными рядами (менее 40 наблюдений) и обладают сложной динамикой, включающей сезонность, тренды и структурные сдвиги. Традиционные

статистические методы, например, модели класса ARIMA (Autoregressive Integrated Moving Average) и SARIMA (Seasonal ARIMA), хотя и обладают хорошей интерпретируемостью, демонстрируют ограниченную точность и высокую чувствительность к малым объемам данных и резким изменениям в структуре ряда [1, 2, 3]. Это делает их применение затруднительным в условиях, когда доступные данные ограничены, а требования к точности прогнозов высоки.

С другой стороны, современные методы машинного обучения, такие как градиентный бустинг (например, XGBoost), способны выявлять сложные нелинейные зависимости и паттерны в данных. Однако их эффективность часто зависит от наличия объемных обучающих выборок и тщательной фичеризации, а интерпретируемость таких моделей может быть низкой [4, 5, 6]. Кроме того, они подвержены риску переобучения при работе с короткими временными рядами, что приводит к нестабильным и ненадежным прогнозам. Появление больших языковых моделей (LLM) открывает новые возможности для интеграции экспертных знаний и контекстуальной информации в процесс прогнозирования, предлагая механизмы для коррекции прогнозов, основанных на статистических и машинных моделях.

Таким образом, возникает острая необходимость в разработке гибридных моделей, которые могли бы объединить преимущества различных подходов, нивелируя их недостатки [7, 8, 9, 10]. Целью данной работы является создание и формализация гибридной системы прогнозирования, способной обеспечить надежные и точные прогнозы для коротких и волатильных социально-экономических временных рядов. Предлагаемая система интегрирует статистическую модель SARIMAX для учета линейных и сезонных зависимостей, алгоритм XGBoost для выявления нелинейных паттернов, а также компоненты LLM для остаточной и экспертной коррекции прогнозов. Важным элементом системы является модуль динамического ансамблирования с экспоненциально-взвешенными весами, который адаптирует вклад каждой базовой модели в итоговый прогноз в зависимости от ее текущей эффективности.

2. Данные и постановка задачи

Задача прогнозирования социально-экономических временных рядов в условиях ограниченных данных и высокой волатильности требует особого подхода [11, 12]. В данной работе рассматривается задача построения гибридной системы прогнозирования для одномерных временных рядов целевой величины Y_t . Эти ряды характеризуются небольшой длиной (порядка 20–40 наблюдений) и подвержены сезонным колебаниям и структурным сдвигам. В качестве примеров таких рядов выступают месячные ряды

вакансий по регионам и укрупнённым профессиональным группам (с горизонтами прогноза $h = 1, 3, 6$ месяцев) и годовые ряды числа поступивших на направления подготовки в вузах (с горизонтами $h = 1, 2, 3$ года).

Пусть Y_t — стационаризованный одномерный ряд целевой величины, а X_t — вектор внутренних экзогенных регрессоров, включающий календарные индикаторы, Фурье-термы и лаги ряда. Требуется построить прогноз $\hat{Y}(t+h)$ на горизонт $h \in \{1, \dots, H\}$ с минимизацией ошибки прогнозирования и обеспечением устойчивости к смене режимов при длинах рядов порядка 20–40 наблюдений. Важным условием является исключение использования новостных источников в базовых моделях для повышения воспроизводимости и снижения зависимости от внешних шумовых факторов.

3. Типизация временных рядов и предобработка

Для корректной идентификации структуры временных рядов и подготовки данных к моделированию применяется диагностический контур, включающий тесты стационарности и детекцию сезонности. Этот этап критически важен для коротких и волатильных рядов, поскольку позволяет адекватно трансформировать исходные данные, устраняя тренды и сезонные компоненты.

3.1. Детекция сезонности и тесты стационарности

Детекция сезонности: оценка основного периода сезонности s выполняется с использованием периодограммы и анализа автокорреляционной (ACF) и частной автокорреляционной (PACF) функций. Периодограмма позволяет выявить доминирующие частоты в ряду, а ACF/PACF помогают определить порядок сезонной составляющей. Выбранный период s используется для сезонного дифференцирования и формирования сезонных Фурье-компонент.

Тесты стационарности: для выявления наличия единичного корня и стабильности ряда вокруг тренда/константы применяются тесты Дики-Фуллера (ADF) [13] и Квятковского-Филлипса-Шмидта-Шина (KPSS) [14, 15].

- ADF-тест используется для проверки нулевой гипотезы о наличии единичного корня (нестационарности). Если p-value теста ADF значимо, это указывает на стационарность ряда.
- KPSS-тест проверяет нулевую гипотезу о стационарности ряда вокруг детерминированного тренда или константы. Если p-value теста KPSS не значимо, это подтверждает стационарность.

3.2. Выбор порядков дифференцирования

На основе результатов тестов стационарности определяются порядки несезонного d^* и сезонного D^* дифференцирования. Для коротких временных рядов ($T < 40$ наблюдений) накладывается ограничение на суммарный порядок дифференцирования: $d^* + D^* \leq 2$. Это позволяет избежать чрезмерного дифференцирования, которое может привести к потере информации и увеличению дисперсии ошибок прогноза.

Трансформированный ряд Z_t формируется следующим образом:

$$Z_t = \nabla^{d^*} \nabla_s^{D^*} Y_t, \quad (1)$$

где $\nabla = (1 - B)$ — оператор несезонной разности, B — оператор обратного сдвига, и $\nabla_s = (1 - B^s)$ — оператор сезонной разности. Такая трансформация устраняет тренд и сезонность, приводя ряд к стационарному виду, что необходимо для применения моделей класса ARMA.

3.3. Формирование экзогенных признаков

Для повышения точности прогнозирования используются внутренние экзогенные признаки, которые формируются на основе самого временного ряда и календарных данных.

Формируются следующие экзогенные внутренние признаки X_t для SARIMAX и матрицы признаков F_t для XGBoost.

- *Календарные индикаторы*: включают в себя такие переменные, как день недели, месяц, квартал, а также индикаторы праздников (детерминированные dummy-переменные). Эти признаки позволяют учесть влияние календарных эффектов на динамику ряда.
- *Гармоники Фурье для сезонности*: для более гибкого моделирования сезонных паттернов используются синусоидальные и косинусоидальные компоненты: $\cos\left(\frac{2\pi kt}{s}\right)$ и $\sin\left(\frac{2\pi kt}{s}\right)$, где k — порядок гармоник, а s — период сезонности. Эти компоненты позволяют аппроксимировать сложные сезонные формы.
- *Лаги ряда и разности* включают лаги исходного ряда $Y_{\{t-1\}}, Y_{\{t-2\}}, \dots$ и разности $\nabla Y_t = Y_t - Y_{\{t-1\}}$. Эти признаки отражают автокорреляционную структуру ряда.
- *Скользящие статистики*: рассчитываются такие статистики, как скользящее среднее (mean), стандартное отклонение (std),

минимум (min), максимум (max) и медиана на различных окнах. Эти признаки позволяют уловить локальные изменения в дисперсии и уровне ряда.

4. Архитектура системы

Предложенная гибридная система прогнозирования реализована как модульная архитектура, обеспечивающая гибкость и масштабируемость. Общая блок-схема архитектуры представлена на рисунке.

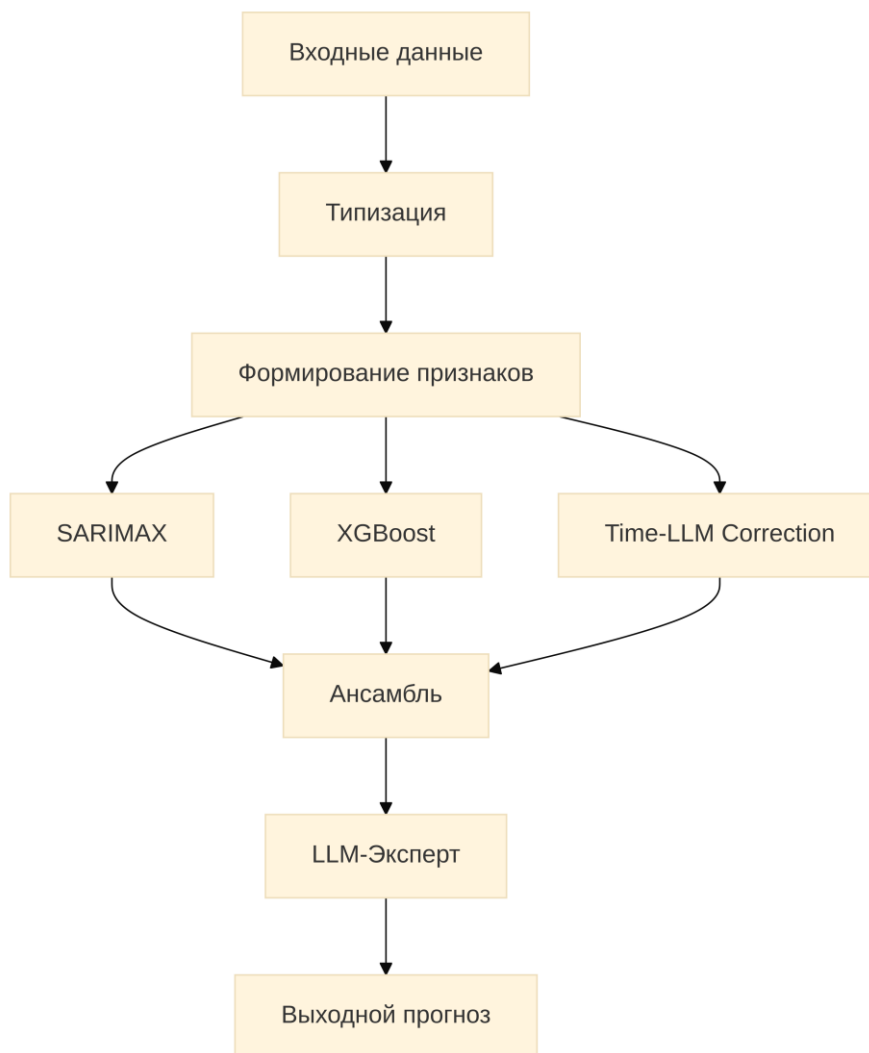


Схема архитектуры гибридной системы прогнозирования

Примечание. Входные данные поступают в модуль типизации временного ряда, затем параллельно обрабатываются компонентами SARIMAX и XGBoost, результаты корректируются модулем Time-LLM, агрегируются динамическим ансамблем с EWA-весами, проходят финальную экспертную LLM-коррекцию и формируют выходной прогноз.

4.1. Поток обработки данных

Поток обработки данных в системе организован следующим образом.

1. Входные данные: система принимает одномерные временные ряды целевой величины Y_t (например, ряды вакансий или числа поступивших).
2. Модуль типизации: входные данные поступают в модуль типизации временного ряда, где выполняются тесты стационарности (ADF, KPSS) и детекция сезонности. На этом этапе определяются оптимальные порядки дифференцирования (d^*, D^*) и период сезонности (s) , а также формируются экзогенные признаки X_t и матрица признаков F_t .
3. Параллельная обработка базовыми моделями: трансформированные данные и сформированные признаки параллельно обрабатываются тремя основными компонентами:
 - SARIMAX: модель SARIMAX строит прогноз, учитывая линейные и сезонные зависимости, а также влияние экзогенных регрессоров X_t .
 - XGBoost: алгоритм XGBoost использует матрицу признаков F_t для выявления нелинейных зависимостей и формирования прогноза.
 - Time-LLM (residual-коррекция): модуль Time-LLM анализирует остатки базовых моделей и внутренние признаки, генерируя корректирующую добавку $\Delta LLM(h)$.
4. Агрегатор ансамбля: прогнозы от SARIMAX, XGBoost и скорректированные прогнозы от Time-LLM поступают в модуль динамического ансамблирования. Здесь рассчитываются инерционно-сглаженные экспоненциально-взвешенные веса (EWA-веса), которые определяют вклад каждой модели в итоговый ансамблевый прогноз $\hat{Y}(t+h)$.
5. LLM-эксперт: опционально, при высоких разногласиях между прогнозами базовых моделей или при обнаружении аномалий, активируется модуль LLM-эксперта. Он формирует микрокоррекцию $\delta_{expert}(h)$, которая применяется к ансамблевому прогнозу для получения финального прогноза $\hat{Y}_{final}(t+h)$.
6. Выходной прогноз: система выдает финальный прогноз, который может включать точечные оценки и интервальные прогнозы.

5. Обсуждение ограничений и рисков

Несмотря на значительные преимущества, гибридная система имеет определенные ограничения и риски, которые необходимо учитывать.

- Короткие ряды: работа с короткими временными рядами всегда сопряжена с риском переобучения и нестабильности идентификации параметров, особенно для SARIMAX. Хотя AICс и ограничения на порядки дифференцирования помогают снизить этот риск, он полностью не устраняется.
- Контроль LLM-коррекции: эффективность LLM-коррекции сильно зависит от качества контекста и механизмов клиппинга. Чрезмерное или неконтролируемое влияние LLM может привести к нестабильности прогнозов [16, 17, 18]. Поэтому строгое ограничение величины поправок и использование процедур self-consistency являются критически важными.
- Стабилизация весов ансамбля: динамические веса ансамбля, хотя и адаптивны, могут быть чувствительны к шуму в оценках ошибок. Использование инерционного сглаживания и регуляризации помогает стабилизировать вклад моделей, но требует тщательной настройки.
- Устойчивость к структурным сдвигам: хотя гибридный подход более устойчив к структурным сдвигам, чем отдельные модели, рекомендуется проводить регулярную проверку устойчивости и анализ остатков на автокорреляции для своевременного выявления проблем.

В целом, предложенная гибридная система представляет собой мощный инструмент для прогнозирования коротких социально-экономических временных рядов, обеспечивая баланс между интерпретируемостью, гибкостью и точностью. Дальнейшие исследования могут быть направлены на более глубокий анализ устойчивости к различным типам структурных сдвигов и автоматизацию процесса настройки гиперпараметров для всех компонентов системы.

6. Заключение

В данной работе была разработана и формализована гибридная система прогнозирования, предназначенная для работы с короткими и волатильными социально-экономическими временными рядами. Система объединяет интерпретируемость статистической модели SARIMAX, гибкость алгоритма машинного обучения XGBoost [19], эвристическую силу LLM-коррекции и устойчивость динамического ансамбля с

экспоненциально-взвешенными весами. Предложенный подход позволяет эффективно решать задачу прогнозирования в условиях ограниченных данных, обеспечивая повышение точности и устойчивости прогнозов на горизонте до пяти шагов.

Ключевые особенности разработанной системы включают:

- универсальный модуль типизации рядов, который адаптирует предобработку данных (дифференцирование, детекция сезонности) к специфике коротких временных рядов, предотвращая переобучение и потерю информации;
- интеграцию SARIMAX и XGBoost, что позволяет учитывать как линейные, так и нелинейные зависимости в данных, используя преимущества каждого метода;
- LLM-компоненты (residual и экспертная коррекция), которые обеспечивают контролируемые и обоснованные поправки к прогнозам, особенно в ситуациях высокой неопределенности, при этом исключая прямое использование новостных источников в базовых моделях для повышения воспроизводимости;
- динамическое ансамблирование с инерционными EWA-весами, которое адаптивно регулирует вклад каждой базовой модели в итоговый прогноз, повышая робастность системы к структурным сдвигам и изменениям в динамике ряда.

Перспективы дальнейших исследований включают разработку многомерных моделей для совместного прогнозирования по регионам или отраслям, создание вероятностных прогнозов (квантильных/прогностических интервалов) на основе исторических ошибок и бутстрепа [20], исследование возможностей федеративного обучения для работы с распределенными данными, а также автоматизированную диагностику дрейфа моделей в реальном времени для поддержания высокой точности прогнозов в долгосрочной перспективе.

Библиографические ссылки

1. Time Series Analysis: Forecasting and Control / G. E. P. Box [et al.]. 5th ed. Wiley, 2015.
2. Скобцов В. Ю., Соколов Б. В., Чжан В. А. Гибридные нейросетевые модели мониторинга данных временных рядов сложных объектов // Известия высших учебных заведений. Приборостроение. 2024. Т. 67, № 1. С. 5–12.
3. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne : OTexts, 2021. URL: <https://otexts.com/fpp3/> (date of access: 10.09.2025).
4. Альчаков В. В. Оценка методов машинного обучения для прогнозирования сезонных временных рядов // Вестник Казанского государственного энергетического университета. 2023. № 3 (59). С. 100–107.

5. Рапаков Г. Г. Исследование LSTM-нейросетевого подхода при моделировании временных рядов // Вестник Московского государственного областного университета. Серия: Физика-Математика. 2023. № 2. С. 78–85.
6. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. P. 785–794.
7. Аботалёв М. С. А. Прогнозирование социально-экономических показателей Российской Федерации на основе гибридного подхода : дис. ... канд. техн. наук. Москва, 2024. 150 с.
8. Аверкин А. Н., Ярушев С. А. Гибридные методы прогнозирования временных рядов на финансовых рынках // Международная конференция по мягким вычислениям и измерениям. 2015. Т. 1. С. 15–18.
9. Казакова Д. С. Гибридные модели предиктивной аналитики в креативной индустрии // International Journal of Open Information Technologies. 2024. Т. 12, № 2. С. 110–115.
10. Толочко И. А. Система стратегического планирования предприятия оборонно-промышленного комплекса на основе логико-лингвистического моделирования : дис. ... канд. экон. наук : 08.00.05 / Санкт-Петербургский государственный экономический университет. Санкт-Петербург, 2021. 200 с.
11. Ярушев С. А., Аверкин А. Н. Когнитивные гибридные системы поддержки принятия решений и прогнозирования // Программные продукты и системы. 2017. № 4. С. 637–643.
12. Чжан Ю. Объяснимый искусственный интеллект в прогнозировании временных рядов : дис. ... канд. физ.-мат. наук : 1.2.2 / Санкт-Петербургский государственный университет. Санкт-Петербург, 2024. 180 с.
13. Dickey D. A., Fuller W. A. Distribution of the Estimators for Autoregressive Time Series With a Unit Root // Journal of the American Statistical Association. 1979. Vol. 74, iss. 366. P. 427–431.
14. Testing the null hypothesis of stationarity against the alternative of a unit root / D. Kwiatkowski // Journal of Econometrics. 1992. Vol. 54, iss. 1–3. P. 159–178.
15. Hurvich C. M., Tsai C.-L. Regression and Time Series Model Selection in Small Samples // Biometrika. 1989. Vol. 76, iss. 2. P. 297–307.
16. Наумов В. Н., Буйневич М. В., Синещук М. Ю. Анализ и прогнозирование временных рядов кибератак на информационную систему ведомственного вуза: возможности и ограничения методов // Труды учебных заведений связи. 2025. № 1. С. 12–20.
17. Чучуева И. А. Модель прогнозирования временных рядов по выборке максимального подобия : автореф. дис. ... канд. техн. наук : 05.13.18 / Московский государственный технический университет имени Н. Э. Баумана. М., 2012. 154 с.
18. Герасименко Д. В. Извлечение тренировочных данных: Риски и решения в контексте безопасности LLM // Известия высших учебных заведений. Приборостроение. 2024. № 1. С. 80–87.
19. Tashman L. J. Out-of-sample tests of forecasting accuracy: an analysis and review // International Journal of Forecasting. 2000. Vol. 16, iss. 4. P. 437–450.
20. Bergmeir C., Benítez J. M. A note on the validity of cross-validation for evaluating autoregressive time series prediction // Computational Statistics & Data Analysis. 2012. Vol. 56, iss. 6. P. 2301–2312.

О ПРОГНОЗИРОВАНИИ МНОГОМЕРНЫХ ДВОИЧНЫХ ВРЕМЕННЫХ РЯДОВ

С. А. Шибалко, Ю. С. Харин

*НИИ прикладных проблем математики и информатики,
Белорусский государственный университет,
Минск, Беларусь, shibalko2003@bk.ru, kharin@bsu.by*

В данной статье рассматривается задача статистического анализа и прогнозирования многомерных двоичных временных рядов. Для решения этих проблем предложены две малопараметрические модели эргодической цепи Маркова порядка s . Построены состоятельные статистические оценки параметров моделей. Приведены алгоритм оценивания параметров и алгоритм прогнозирования.

Ключевые слова: многомерный двоичный временной ряд; цепь Маркова порядка s ; нейросетевая модель; статистическое оценивание параметров; статистическое прогнозирование.

FORECASTING OF MULTIVARIATE BINARY TIME SERIES

S. A. Shibalko, Yu. S. Kharin

*Research Institute for Applied Problems of Mathematics and Informatics,
Belarusian State University,
Minsk, Belarus, shibalko2003@bk.ru, kharin@bsu.by*

This article is devoted to statistical analysis and forecasting of multivariate binary time series. For solving these problems two parsimonious models of ergodic Markov chain of order s are determined. Consistent statistical estimators for model parameters and algorithms for parameters estimation and forecasting are developed.

Keywords: multivariate binary time series, Markov chain of order s , neural network-based model, statistical estimation, statistical forecasting.

1. Введение

Временной ряд представляет собой последовательность данных, полученных с определённым временным интервалом. Такие ряды играют ключевую роль в анализе реальных процессов. Сегодня они применяются в самых разных сферах — от экономики и финансов до информатики,

эпидемиологии, производства и метеорологии. По мере роста вычислительных возможностей появляются всё новые методы их анализа. Теория статистического анализа временных рядов [1] глубоко развита для «непрерывных» моделей. Однако процесс цифровизации экономики и других сфер жизни приводит к увеличению объёма данных, представленных в дискретном пространстве состояний с дискретным временем. Для описания таких данных используются дискретные временные ряды, включая бинарные. В связи с этим задача статистического анализа многомерных дискретных временных рядов становится актуальной в рамках как математической, так и прикладной статистики.

2. Математическая модель многомерных двоичных временных рядов

Примем обозначения: Z – множество целых чисел, $\mathbb{R}^k$ – k -мерное евклидово пространство, $V = \{0,1\}$ – двоичный алфавит, $\mathbb{N}$ – множество натуральных чисел, штрих у матрицы – символ транспонирования.

Определим на вероятностном пространстве $(\Omega, \mathcal{F}, P)$ N -мерный ($N \in \mathbb{N}$) двоичный временной ряд $X_t = (x_{t1}, \dots, x_{tN})' \in V^N$, порожденный семейством условных распределений вероятностей:

$$P\{X_t = J_t \mid \mathcal{F}_{t-1}\} = P\{X_t = J_t \mid X_{t-1} = J_{t-1}, \dots, X_{t-s} = J_{t-s}\}, t \in Z, \quad (1)$$

где $x_{tl} \in V$ – двоичная (бинарная) случайная величина, задающая компоненту номер l временного ряда в момент времени $t \in Z$, $J_t = (j_{tl}) \in V^N$ – значение двоичного случайного вектора X_t в момент времени $t \in Z$, $\mathcal{F}_{t-1} = \sigma\{X_\tau: \tau \leq t-1\}$ – σ -алгебра случайных событий, порожденных указанными в скобках случайными векторами, s – глубина предыстории (памяти процесса), $s \in \mathbb{N}$.

Рассмотрим ситуацию, когда при фиксированной s -предыстории $X_{t-1} = J_{t-1}, \dots, X_{t-s} = J_{t-s}$ случайные величины $x_{t1}, \dots, x_{tN}$ условно независимы:

$$\begin{aligned} P\{X_t = J_t \mid X_{t-1} = J_{t-1}, \dots, X_{t-s} = J_{t-s}\} = \\ = \prod_{l=1}^N P\{x_{tl} = j_{tl} \mid X_{t-1} = J_{t-1}, \dots, X_{t-s} = J_{t-s}\}, J_t = (j_{tl}) \in V^N, \end{aligned} \quad (2)$$

где условное распределение l -го бита x_{tl} при условии, что фиксирована s -предыстория, представимо в виде

$$P\{x_{il} = j_{il} \mid X_{t-1} = J_{t-1}, \dots, X_{t-s} = J_{t-s}\} = \begin{cases} p_l(J_{t-1:t-s}), & j_{il} = 1, \\ 1 - p_l(J_{t-1:t-s}), & j_{il} = 0; \end{cases} \quad (3)$$

здесь $J_{t-1:t-s} = (J'_{t-1}, J'_{t-2}, \dots, J'_{t-s})' \in V^{Ns}$ – составной двоичный вектор-столбец s -предыстории. Заметим, что условная независимость (2) в общем случае не эквивалентна безусловной независимости, так что компоненты $x_{t1}, \dots, x_{tN}$ могут быть зависимы между собой через общую для них предысторию.

3. Малопараметрические модели многомерных двоичных временных рядов

Модель (1) – (3) полностью задается N функциями от $N \cdot s$ переменных каждая, таким образом для задания этой вероятностной модели требуется $N \cdot 2^{Ns}$ независимых параметров. Из-за «проклятия размерности» для уменьшения потребления памяти компьютера и времени работы алгоритмов оценивания параметров моделей необходимо использовать малопараметрические модели [2].

3.1. Малопараметрическая модель на основе базисных функций

Обозначим: $J_{s:1} = (J'_s, J'_{s-1}, J'_1)' \in V^{Ns}$ – составной вектор-столбец, задающий s -предысторию; $\{\psi_k(J_{s:1})\} = \psi_1(J_{s:1}), \psi_2(J_{s:1}), \dots, \psi_m(J_{s:1})$ – семейство линейно независимых на V^{Ns} базисных функций.

Для l -ой компоненты определим малопараметрическую модель на основе базисных функций [3]:

$$p = p(J_{s:1}) = F\left(\sum_{k=1}^m b_k \psi_k(J_{s:1})\right), \quad J_{s:1} \in V^{Ns}, \quad (4)$$

где $F(\cdot)$ – некоторая дважды непрерывно дифференцируемая функция распределения $0 < F(x) < 1$, $B = (b_k) \in \mathbb{R}^m$ – вектор-столбец неизвестных параметров модели; здесь индекс l компоненты для упрощения обозначений опущен. Модель (4) имеет $m < m^+ = 2^{Ns}$ неизвестных параметров.

3.2. Малопараметрическая нейросетевая модель

Главная проблема малопараметрической модели (4) заключается в выборе базисных функций $\{\psi_k\}$. Одним из вариантов решения данной проблемы является использование искусственных нейронных слоев для

аппроксимации базисных функций $\{\psi_k\}$. Рассмотрим малопараметрическую нейросетевую модель.

Для l -ой ($l = 1, \dots, N$) компоненты определим малопараметрическую нейросетевую модель [4] (индекс компоненты l опущен для упрощения обозначений):

$$p = p(J_{s:1}) = F\left(\sum_{k=1}^m b_k F_k\left(\sum_{l=1}^{Ns} a_{kl} j_l\right)\right), \quad (5)$$

где $F(\cdot)$ – некоторая дважды непрерывно дифференцируемая функция распределения $0 < F(x) < 1$, $F_1, \dots, F_m$ – некоторые заданные абсолютно непрерывные функции распределения, $B = (b^k) \in \mathbb{R}^m$ и $A_k = (a_{k,1}, \dots, a_{k,Ns})' \in \mathbb{R}^{Ns}$, $k = 1, \dots, m$ – неизвестные вектор-столбцы коэффициентов (параметров) модели. Таким образом, модель для l -ой компоненты описывается двух-слойной искусственной нейронной сетью с Ns входами, одним выходом, m нейронами на первом слое и одним нейроном на втором слое; при этом функции $F_1, \dots, F_m$ называют функциями активации. Модель (5) имеет $m(Ns + 1)$ неизвестных параметров.

Представим модель (5) в матричном виде:

$$p = p(J_{s:1}) = F(B'F(A'J_{s:1})), J_{s:1} \in V^{Ns}, \quad (6)$$

где $F = (F_1, \dots, F_m)' : \mathbb{R}^m \rightarrow \mathbb{R}^m$ – векторная функция, осуществляющая функциональное преобразование покомпонентно, элементами которой являются функции $\{F_k\}$, $A = (A_1, \dots, A_m) \in \mathbb{R}^{Ns \times m}$ – составная матрица коэффициентов (параметров) модели.

Л е м м а 1. Для любого набора параметров $\theta^{(1)} = (B, A_1, \dots, A_m) \in \mathbb{R}^{m(Ns+1)}$ модели (5) существует $m!$ эквивалентных ему наборов параметров, отличающихся перестановкой пар $(b_1, A_1), \dots, (b_m, A_m)$.

4. Статистическое оценивание параметров и прогнозирование

Пусть наблюдается реализация длины T двоичного N -мерного двоичного временного ряда $X_{1:T} = (X_1, \dots, X_T) \in V^{TN}$. Для оценивания вектор-столбцов параметров моделей (4), (5) применим FBE-метод [5], основанный на многомерных частотах.

4.1. Малопараметрическая модель на основе базисных функций

Примем следующие обозначения: $\Psi(J_{s:1}) = (\psi_k(J_{s:1})) \in \mathbb{R}^{m \times 1}$ – вектор-столбец, состоящий из базисных функций $\psi_k(J_{s:1})$, $(1; J_{s:1}) = (1, x_{s+1,1}, \dots, x_{s+1,l-1}, J'_s, J'_{s-1}, \dots, J'_1)' \in V^{Ns+l}$ – «расширенный» вектор-столбец,

$$v_s(J_{s:1}) = \sum_{t=s}^T \mathbf{1}\{X_t = J_s, \dots, X_{t-s+1} = J_1\},$$

$$v_{s+1}(1; J_{s:1}) = \sum_{t=s}^{T-1} \mathbf{1}\{x_{t+1,l} = 1, X_t = J_s, \dots, X_{t-s+1} = J_1\}$$

– частоты s -грамм.

Построим частотную оценку условной вероятности (1):

$$\hat{p}(J_{s:1}) = \begin{cases} \frac{T-s}{T-s+1} \cdot \frac{v_{s+1}(1; J_{s:1})}{v_s(J_{s:1})}, & J_{s:1} \in J^{(s)}, \\ \frac{1}{2}, & J_{s:1} \notin J^{(s)}; \end{cases} \quad (7)$$

где $J^{(s)} = \{J_{s:1} \in V^{Ns}, t = s, s+1, \dots, T : v_s(J_{s:1}) > 0\} \subseteq V^{Ns}$ – подмножество s -предысторий, имеющих ненулевые частоты в $X_{1:T}$. И следующие статистики:

$$\hat{u}(J_{s:1}) = F^{-1}(\hat{p}(J_{s:1})) \in \mathbb{R}, E = (e_k) = \sum_{J_{s:1} \in J^{(s)}} \hat{u}(J_{s:1}) \Psi(J_{s:1}) \in \mathbb{R}^{m \times 1},$$

$$D = (d_{ij}) = \sum_{J_{s:1} \in J^{(s)}} \Psi(J_{s:1}) \Psi'(J_{s:1}) \in \mathbb{R}^{m \times m}, D^{-1} = (\bar{d}_{ij}) \in \mathbb{R}^{m \times m}.$$

Т е о р е м а 1. Если определитель $|D| \neq 0$, то FBE-оценка

$$\hat{B} = D^{-1}E \quad (8)$$

при $T \rightarrow +\infty$ состоятельна, т.е. сходится по вероятности к истинному значению B^0 .

Т е о р е м а 2. FBE-оценка $\hat{B}$ является асимптотически несмещенной при $T \rightarrow +\infty$.

Перепишем (8) в следующем виде: $\hat{B} = D^{-1}(\bar{\Psi}U)$, где $U = (\hat{u}(J_{s:1}) | J_{s:1} \in J^{(s)})' \in \mathbb{R}^{|J^{(s)}| \times 1}$ – вектор-столбец, состоящий из $\hat{u}(J_{s:1})$ для всех s -предысторий $J_{s:1}$ из $J^{(s)}$, $\bar{\Psi} \in \mathbb{R}^{m \times |J^{(s)}|}$ – матрица, в k -ой строке которой стоит базисная функция $\psi_k(J_{s:1})$ для всех s -предысторий $J_{s:1}$ из $J^{(s)}$.

Т е о р е м а 3. В условиях Теоремы 1 FBE-оценка $\hat{B}$ имеет асимптотически нормальное распределение: $\sqrt{T}(\hat{B} - B^0) \xrightarrow{T \rightarrow \infty} N_m(0_m, \Sigma)$ с нулевым математическим ожиданием и ковариационной матрицей Σ , равной $\Sigma = D^{-1}\bar{\Psi}\bar{F}(\Sigma_p)^{-1}\bar{F}'\bar{\Psi}'(D^{-1})'$, где $\Sigma_p = \text{diag}\left(\frac{\pi(J_{s:1})}{p^0(J_{s:1})(1-p^0(J_{s:1}))}\right) \in \mathbb{R}^{|J^{(s)}| \times |J^{(s)}|}$, $\bar{F} = \text{diag}\left(\frac{dF^{-1}(z)}{dz}\right) \in \mathbb{R}^{|J^{(s)}| \times |J^{(s)}|}$, $\pi(J_{s:1})$ – стационарное распределение вероятностей.

4.2 Малопараметрическая нейросетевая модель

Определим квадратичную функцию потерь:

$$W(B, A) = \sum_{J_{s:1} \in J^{(s)}} (\hat{u}(J_{s:1}) - B'F(A'J_{s:1}))^2 \geq 0. \quad (9)$$

Чтобы построить статистические оценки $\hat{A}$, $\hat{B}$ параметров A , B модели (5), необходимо решить следующую экстремальную задачу:

$$W(B, A) \rightarrow \min_{B, A}. \quad (10)$$

Для того, чтобы оценить вектор параметров B , найдем минимум функции (9) при фиксированной матрице параметров A .

Построим следующие статистики:

$$\begin{aligned} \hat{u}(J_{s:1}) &= F^{-1}(\hat{p}(J_{s:1})) \in \mathbb{R}, \\ E = (e_k) &= \sum_{J_{s:1} \in J^{(s)}} \hat{u}(J_{s:1}) F'(A'J_{s:1}) \in \mathbb{R}^{m \times 1}, \\ D = (d_{ij}) &= \sum_{J_{s:1} \in J^{(s)}} F(A'J_{s:1}) F'(A'J_{s:1}) \in \mathbb{R}^{m \times m}, D^{-1} = (\bar{d}_{ij}) \in \mathbb{R}^{m \times m}. \end{aligned} \quad (11)$$

Т е о р е м а 4. Если в определитель $|D| \neq 0$ (11) и известны параметры $A_i = (a_{i,1}, \dots, a_{i,ns})' \in \mathbb{R}^{N_s}$, $i = 1, \dots, m$, то FBE-оценка для B имеет вид

$$\hat{B} = (b_k) = D^{-1}E; \quad (12)$$

при $T \rightarrow \infty$ она состоятельна, т.е. сходится по вероятности к истинному значению B^0 .

На втором этапе решения задачи (10) для построения статистической оценки $\hat{A}$ подставим оценку $\hat{B}$ в функцию потерь и получим новую экстремальную задачу:

$$w(A) \rightarrow \min_A, w(A) := W(\hat{B}, A) = \sum_{J_{s:1} \in J^{(s)}} \hat{u}^2(J_{s:1}) - E'D^{-1}E. \quad (13)$$

Л е м м а 2. В принятых обозначениях частные производные первого порядка целевой функции (13) вычисляются по формуле (в матричном виде):

$$\frac{dw(A)}{dA} = - \left(\frac{dE'}{dA} + \frac{dE}{dA} \right) \hat{B} - \frac{dD^{-1}}{dA} (E \otimes I_m) E,$$

где для матричных производных в правой части получены явные формулы.

Используя Лемму 2, построим итерационный алгоритм градиентного спуска для статистического оценивания параметров модели:

- 1) инициализируем начальное значение матричного параметра A : $\hat{A}^{(0)}$;
- 2) вычисляем статистическую оценку вектора параметров $\hat{B}$ согласно формуле $\hat{B}^{(\tau-1)} = D^{-1}(\hat{A}^{(\tau-1)})E(\hat{A}^{(\tau-1)})$, $\tau = 1, 2, \dots$, где $D^{-1}(\cdot)$, $E(\cdot)$ вычисляются по формулам (11);
- 3) обновляем значение оценки матрицы параметров $\hat{A}$, используя

$$\hat{A}^{(\tau)} = \hat{A}^{(\tau-1)} - \alpha \frac{dw(\hat{B}^{(\tau-1)}, \hat{A}^{(\tau-1)})}{dA}, \quad (14)$$

где $\hat{B}^{(\tau)}$ – оценка параметра B на τ -ой итерации, $\hat{A}^{(\tau)}$ – оценка параметра A на τ -ой итерации, α – задаваемая величина шага итерации.

Для нахождения глобального минимума функции потерь необходимо осуществить итерационный процесс градиентного спуска (14) для $L \in \mathbb{N}$ различных вариантов задания начального значения $\hat{A}^{(0)}$.

Пусть (B^0, A^0) – истинные значения параметров, (B_π^0, A_π^0) – перестановочный набор параметров: $B_\pi^0 = (b_{\pi_1}^0, \dots, b_{\pi_m}^0)$, $A_\pi^0 = (A_{\pi_1}^0, \dots, A_{\pi_m}^0)$, где $\pi = (\pi_1, \dots, \pi_m) \in \Pi_m$ – произвольная подстановка из множества $m!$ всевозможных подстановок на $\{1, \dots, m\}$.

Т е о р е м а 5. Если имеет место модель (5) с истинными значениями параметров B^0, A^0 , $|D| \neq 0$ и $(\hat{B}, \hat{A})$ – статистические оценки параметров (B^0, A^0) , построенные с помощью алгоритма градиентного спуска (14), то при $L \rightarrow +\infty$ и $T \rightarrow +\infty$ имеет место сходимость по вероятности статистических оценок параметров к истинным значениям $(\hat{B}, \hat{A}) \xrightarrow{P} (B_\pi^0, A_\pi^0)$ для некоторой подстановки $\pi \in \Pi_m$.

4.3. Статистическое прогнозирование

Подстановочный алгоритм оптимального прогнозирования на один шаг с минимальной вероятностью ошибки определяется явным соотношением: $\hat{x}_{il} = \mathbf{1} \left\{ p_l(X_{t-1:t-s}) - \frac{1}{2} > 0 \right\}$, $l = 1, \dots, N$.

Используя статистические оценки параметров моделей, приведем состоятельный при $T \rightarrow \infty$ алгоритм прогнозирования многомерных двоичных временных рядов на один шаг с минимальной вероятностью ошибки:

$$\hat{x}_{il} = \mathbf{1} \left\{ \hat{p}_l(X_{t-1:t-s}) - \frac{1}{2} > 0 \right\}, \quad l = 1, \dots, N.$$

5. Заключение

В статье получены следующие основные результаты: для многомерных двоичных временных рядов предложены малопараметрическая модель на основе базисных функций и малопараметрическая нейросетевая модель; разработаны алгоритм оценивания параметров моделей и подстановочный алгоритм прогнозирования многомерных двоичных временных рядов. Полученные результаты проиллюстрированы решением прикладных задач.

Библиографические ссылки

1. *Anderson T. W.* The Statistical Analysis of Time Series. New York : Wiley, 1971.
2. *Харин Ю. С., Волошко В. А.* Об аппроксимации двоичных цепей Маркова высокого порядка малопараметрическими моделями // Дискрет. математика 2022. Вып. 34, № 3. С. 114–135.
3. *Shibalko S., Kharin Yu.* Parsimonious models of multivariate binary time series: statistical estimation and forecasting // Pattern Recognition and Information Processing (PRIP'2023). Artificial Intelliverse: Expanding Horizons : Proceedings of the 16th International Conference / Belarusian State University : eds. A. Nedzved, A. Belotserkovsky. – Minsk : BSU, 2023. P. 296–300.
4. *Харин Ю. С., Шибалко С. А.* Статистический анализ многомерных двоичных временных рядов на основе нейросетевой модели // Докл. Нац. акад. наук Беларуси. 2024. Т. 68, № 4. С. 271–281
5. *Шибалко С. А.* Малопараметрическая модель на основе базисных функций для статистического анализа двоичных временных рядов // ИКТ-2024: электронный сборник статей III международной практической конференции, Полоцк, 29 марта 2024 г. / Новополоцк : Полоцкий государственный университет имени Евфросинии Полоцкой, 2024. С. 111–115.

РАЗДЕЛ IV

ОПТИМИЗАЦИЯ И НАДЁЖНОСТЬ СИСТЕМ

ПРИМЕНЕНИЕ МЕТОДОВ ОПТИМИЗАЦИИ ДЛЯ КОРРЕКЦИИ ДОЗ ИНСУЛИНА ДЛЯ ПАЦИЕНТОВ С САХАРНЫМ ДИАБЕТОМ 1 ТИПА

Е. С. Акулич

*Белорусский государственный университет,
Минск, Беларусь, fpm.akulichESI@bsu.by*

Рассматривается задача оптимизации инсулинотерапии пациентов с сахарным диабетом 1 типа, использующих болюсные дозы инсулина короткого и длинного действия. В качестве основы используется математическая модель динамики уровня глюкозы в крови [1], модифицированная для учета различных типов болюсного инсулина и многократных приемов пищи. Индивидуальные параметры модели определяются на основе наблюдений за конкретным пациентом. Оптимизация доз проведена для инсулина короткого действия.

Ключевые слова: сахарный диабет 1 типа; инсулинотерапия; математическая модель; идентификация параметров; оптимизация.

OPTIMIZATION OF INSULIN THERAPY FOR PATIENTS WITH TYPE 1 DIABETES MELLITUS

E. S. Akulich

*Belarusian State University,
Minsk, Belarus, fpm.akulichESI@bsu.by*

The paper deals with a problem of optimizing insulin therapy for patients with type 1 diabetes mellitus using bolus doses of short- and long-acting insulin. The mathematical model of blood glucose dynamics from [1] is modified to take into account different types of bolus insulin and multiple meals. Individual parameters of the model are determined based on observations of a specific patient. Optimization is performed for short-acting insulin.

Keywords: type 1 diabetes mellitus; insulin therapy; mathematical model; parameter identification; optimization.

1. Введение

Математическая модель динамики уровня глюкозы в крови пациента с диабетом 1 типа, предложенная в работе [1], имеет восемь переменных состояния, изменение которых описывается системой нелинейных дифференциальных уравнений, и два входа: ввод пациенту инсулина с помощью

инсулиновой помпы и прием пациентом пищи. Первый из них является управлением и может оптимизироваться для достижения целевых показателей уровня сахара в крови пациента. Цель настоящей работы – модифицировать модель из [1] с целью учета следующих двух факторов. Во-первых, пациент не имеет помпы, а использует болюсные инъекции инсулина короткого (Новорапид) и длинного (Левемир) действия. Во-вторых, в работе [1] предполагается однократный прием углеводной пищи, либо равномерное введение углеводов на коротком промежутке времени. В связи с этим в настоящей работе также была проведена модификация подсистемы питания с целью учета многократных приемов пищи пациентом в течение дня. Идея данной модификации основана на [2].

Коэффициенты модели [1] и предлагаемой модификации разбиты на две группы: константы (не зависят от конкретного пациента) и параметры (являются индивидуальными для каждого пациента). Параметры требуют идентификации на основе данных замеров глюкозы в крови на протяжении некоторого промежутка времени. После успешной идентификации параметров ставится и решается задача оптимизации инсулинотерапии – подбор доз инсулина, который использует пациент на протяжении суток. Цель оптимизации – минимизировать квадратичное отклонение уровня сахара от целевого значения, равного 6 ммоль/л.

2. Математическая модель

Математическая модель, исследуемая в данной работе, имеет вид

$$\begin{aligned}
 \frac{dI(t)}{dt} &= u_s(t) + u_l(t) - k_e I(t), \\
 \frac{dx_1(t)}{dt} &= -k_{a1} x_1(t) + k_{b1} I(t), \\
 \frac{dx_2(t)}{dt} &= -k_{a2} x_2(t) + k_{b2} I(t), \\
 \frac{dx_3(t)}{dt} &= -k_{a3} x_3(t) + k_{b3} I(t), \\
 \frac{dQ_1(t)}{dt} &= -x_1(t)Q_1(t) + k_{12}Q_2(t) - \left[F_{01}^c(G(t)) + F_R(G(t)) \right] + \\
 &\quad + U_G(t) + EGP_0(1 - x_3(t)), \\
 \frac{dQ_2(t)}{dt} &= x_1(t)Q_1(t) - (k_{12} + x_2(t))Q_2(t),
 \end{aligned} \tag{1}$$

$$\begin{aligned}\frac{dD_1(t)}{dt} &= A_G D(t) - k_D D_1(t), \\ \frac{dD_2(t)}{dt} &= k_D D_1(t) - k_D D_2(t), \quad t \geq 0.\end{aligned}$$

Здесь $I(t)$ — концентрация инсулина в плазме крови (мМЕ/л); $u_s(t)$ — введение инсулина короткого действия; $u_l(t)$ — введение инсулина длительного действия; k_e — скорость поглощения инсулина (1/мин); $x_1(t)$ — коэффициент влияния инсулина на транспортировку и распространение глюкозы (1/мин); k_{a1} — скорость деактивации глюкозы; k_{b1} — чувствительность инсулина при распределении / транспортировке; $x_2(t)$ — коэффициент влияния инсулина на утилизацию глюкозы (1/мин); k_{a2} — скорость деактивации глюкозы; k_{b2} — чувствительность инсулина при удалении; $x_3(t)$ — коэффициент действия инсулина на выработку эндогенной глюкозы в печени (1/мин); k_{a3} — скорость деактивации глюкозы; k_{b3} — чувствительность инсулина при производстве глюкозы печенью; $Q_1(t)$ — масса глюкозы в кровяном потоке (ммоль); k_{12} — скорость передачи глюкозы из кровяного потока периферийным тканям (1/мин); $Q_2(t)$ — масса глюкозы в периферийных тканях (ммоль); $F_{01}^c(G)$ — скорость потребления глюкозы (ммоль/мин) в центральной нервной системе; $G(t)$ — уровень глюкозы в крови (ммоль/л); $F_R(G)$ — скорость производства глюкозы в почках (ммоль/мин); $U_G(t)$ — скорость абсорбции глюкозы (ммоль/мин); $D(t)$ — количество переваренных углеводов в момент времени t ; $D_1(t)$ — количество переваренных углеводов в кровяном потоке в момент t (г); $D_2(t)$ — количество переваренных углеводов в периферийных тканях в момент t (г); EGP_0 — скорость производства печенью глюкозы при нулевом уровне инсулина; A_G — биодоступность углеводов (безразмерная); k_D — скорость перетекания углеводов между отделами (гр/мин).

Модель (1) содержит три входа: функции $u_s(t)$, $u_l(t)$, $t \geq 0$, связанные с введением пациенту двух типов инсулина и $D(t)$, $t \geq 0$, отвечающая за прием углеводной пищи. Эти функции задаются в следующем виде:

$$u_s(t) = \frac{k}{V_I} \sum_{i=0}^{n_s} k(t - \tau_i) e^{-k(t - \tau_i)} I_s(\tau_i), \quad u_l(t) = \frac{k_l}{V_I} \sum_{i=0}^1 k_l(t - \tau'_i) e^{-k_l(t - \tau'_i)} I_s(\tau'_i),$$

$$D(t) = \frac{1000v(t)}{M_{wG}}, \quad v(t) = \begin{cases} d_i/15, & t \in [t_i, t_i + 15], i = 1, 2, \dots, n_d, \\ 0, & \text{иначе,} \end{cases} \quad t \geq 0,$$

где $\tau_i, I_s(\tau_i), i = \overline{1, n_s}$, – моменты введения и дозы инъекций болюсного инсулина короткого действия (Новорапид), $0 = \tau_0 < \tau_1 < \dots < \tau_{n_s}$; $\tau'_i, I_l(\tau'_i), i = 1, 2$ – моменты введения и дозы длинного инсулина (дважды в день); $v(t), t \geq 0$ – кусочно-постоянная функция, описывающая прием углеводной пищи в объеме d_i в моменты $t_i, i = \overline{1, n_d}$ (5-6 раз в день по 15 минут).

Уровень измеряемой глюкозы (выход модели) вычисляется как

$$G(t) = Q_1(t)/V_G, \quad t \geq 0, \quad (2)$$

где V_G — объем распределения глюкозы (л).

3. Идентификация параметров модели

Идентификации подлежат следующие 7 параметров модели:

$$p_i = \frac{k_{bi}}{k_{ai}}, \quad i = 1, 2, 3; \quad p_4 = EGP_0, \quad p_4 = F_{01}, \quad p_6 = \frac{1}{k_s}, \quad p_7 = \frac{1}{k_l}. \quad (3)$$

Для получения экспериментальных данных были проведены наблюдения (замеры уровня сахара в крови пациента) на протяжении 15 часов. Инсулин вводился по опыту пациента: в начале и конце эксперимента вводился Левемир, перед приемом пищи – Новорапид, доза которого подбиралась под количество углеводов в планируемом приеме пищи. Пациент находился на плане пятиразового питания, принимая разное количество углеводов в граммах в каждый прием пищи (15 минут).

Замеры уровня сахара с помощью глюкометра проводились в моменты времени t_i^d . Измеренный уровень сахара обозначен $G_i^d, i = \overline{1, N_d}$.

Графическая иллюстрация данных эксперимента приведена на рис. 1. Время начала эксперимента принято за 0. На верхнем графике черным цветом указаны инъекции короткого инсулина, красным – длинного. На среднем графике показаны моменты и величины поступающей углеводной пищи. Измерения глюкозы отмечены на нижнем графике символом *, розовая ломаная их соединяющая, приведена для удобства сравнения с модельными результатами в дальнейшем эксперименте. Остальные кривые будут пояснены ниже.

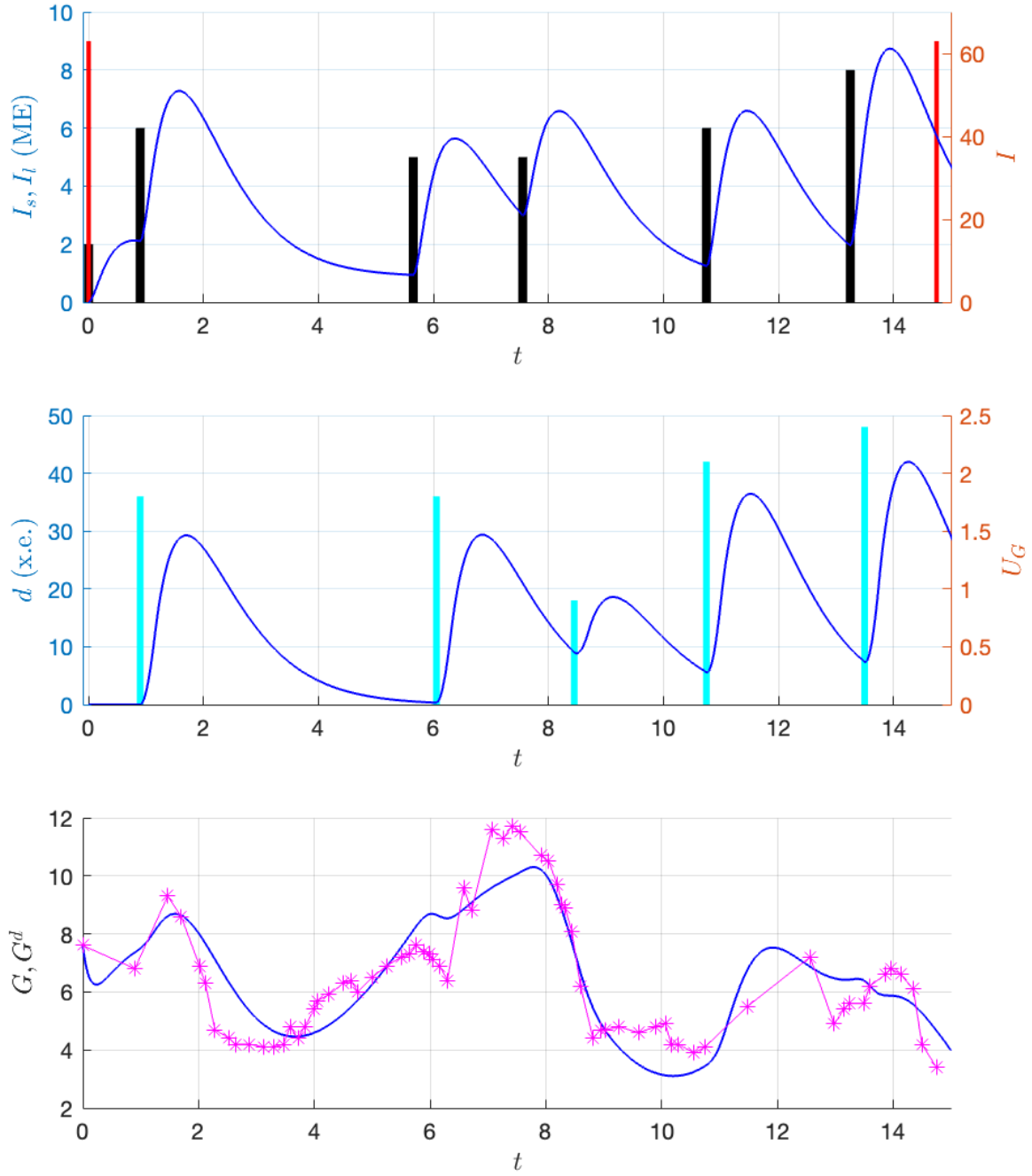


Рис. 1. Данные эксперимента и результаты идентификации

Для идентификации пациентозависимых параметров p минимизируется квадратичное отклонение модельной кривой $G(t)$ от измеренного уровня глюкозы в крови G_i^d , $i = \overline{1, N_d}$:

$$\sum_{i=1}^{N_d} \left(G(t_i^d | p) - G_i^d \right)^2 \rightarrow \min_p. \quad (4)$$

Здесь в $G(t_i^d | p)$ после вертикальной черты явно указана зависимость выхода (2) от вектора параметров (3).

В результате решения задачи (4) был получен следующий вектор оптимальных значений параметров: $p^0 = [3,3998, 2,3772, 1899,9, 0.3364, 0,6145, 38,063, 379,4303]$.

На рис. 1 приведены совмещенные с данными кривые, изображающие траектории инсулина $I(t)$, $t \geq 0$ (верхний график) скорость абсорбции глюкозы $U_G(t)$, $t \geq 0$ (средний график) и уровень сахара в крови $G(t)$ $t \geq 0$ (нижний график).

На нижнем графике можно оценить отклонение измеренной и модельной глюкозы в крови пациента. Квадратичное отклонение (оптимальное значение задачи минимизации (4)) составило 74,6988.

4. Оптимизация доз инъекций короткого инсулина

Сформулируем задачу оптимизации доз инсулина короткого действия, при условии, что в модели установлены оптимальные параметры, полученные в предыдущем эксперименте. Инъекции длинного инсулина остаются прежние, т.е. не подлежат оптимизации. Переменные оптимизации соберем в вектор $u = (I_s(\tau_i), i=1,2,\dots,6)$, где величина измерения для $I_s(\tau_i)$ выбрана в МЕ.

В качестве целевой функции в задаче оптимизации была выбрана сумма квадратичных отклонений уровня глюкозы от желаемого уровня y :

$$\sum_{t=0}^{1458} (G(t | u, p^0) - y)^2 \rightarrow \min_u. \quad (5)$$

Здесь после вертикальной черты в функции G указана зависимость ее от доз короткого инсулина u и пациентозависимых параметров, подобранных ранее, y — норма уровня сахара для больных сахарным диабетом, обычно составляет 6 (ммоль/л). Также были наложены дополнительные ограничения на разовые дозы инъекции, $0 \leq I_s(\tau_i) \leq 9$.

В эксперименте с вектором параметров p^0 , полученным в результате решения задачи (4), были найдены следующие значения вектора оптимальных доз инсулина: $u^0 = [3,0773, 5,0746, 5,5714, 3,7332, 7,0093, 6,4462]$. Общий объем введенного инсулина короткого действия составил 30,9120 (МЕ). На основе опыта пациент в данном эксперименте использовал 32 (МЕ). Значение критерия качества (5) оказалось равным 2086,508.

Динамика переменных входа и выхода модели при использовании доз инсулина из u^0 приведена на рис. 2. Траектории, соответствующие оптимизированным дозам инсулина на графиках представлены красными кривыми. Синими показаны траектории при инъекциях, подобранных пациентом самостоятельно и указанных на рис. 1. Таким образом, можно оценить отклонения траекторий, полученных после оптимизации доз короткого инсулина, от значений без оптимизации.

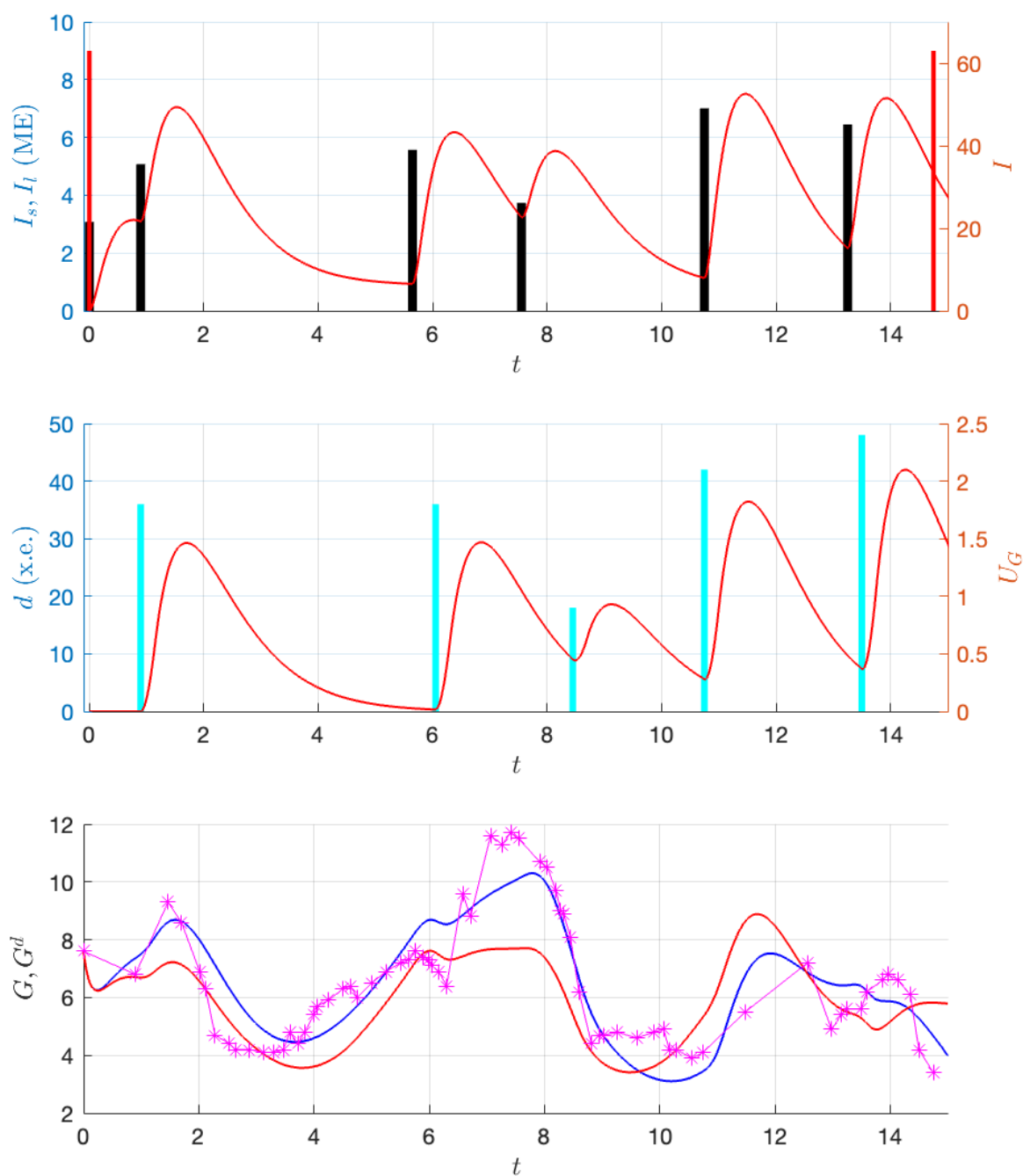


Рис. 2. Результаты оптимизации болюсных инъекций

5. Заключение

Работа посвящена математическому моделированию и оптимизации процессов, протекающих в теле пациента с сахарным диабетом 1 типа. Математическая модель, наиболее полно описывающая процессы в организме, состоит из 8 нелинейных дифференциальных уравнений, которые содержат большое количество параметров, в том числе пациентозависимых. В модифицированной автором модели рассмотрено многократное питание пациента, а также болюсные дозы длинного и короткого инсулина. Применяя метод наименьших квадратов, оценивающий отклонение модельных данных от реальных, была сформулирована и решена задача идентификации пациентозависимых параметров по данным конкретного пациента: его плану питания приема инсулина и замеров уровня сахара на протяжении суток. Затем решена задача оптимизации доз короткого инсулина, по результатам которой удастся избежать резких скачков уровня глюкозы — что является одной из главных задач больного сахарным диабетом 1 типа.

Библиографические ссылки

1. Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes / R. Novorka [et al.] // *Physiological measurement*. 2004. Vol. 25, No. 4. P. 905–920.
2. *Астровский А. И.* Вычислительный эксперимент с математической моделью диабета первого типа // *Труды Института математики НАН Беларуси*. 2021. № 1–2. С. 41–51.

ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ МЕЖДУ ПОКАЗАТЕЛЯМИ СРАВНИТЕЛЬНОЙ СТАТИКИ ТЕОРИИ ФИРМЫ

В. В. Альсевич

*Белорусский государственный университет,
Минск, Беларусь, alsevichvv@bsu.by*

В работе приводится вывод аналога теоремы Слуцкого, полученной в теории потребления, для задачи теории фирмы.

Ключевые слова: задача фирмы; аналог теоремы Слуцкого.

STUDY OF THE DEPENDENCY BETWEEN INDICATORS OF COMPARATIVE STATICS OF THE THEORY OF THE FIRM

V. V. Alseovich

*Belarusian State University,
Minsk, Belarus, alsevichvv@bsu.by*

The paper presents a derivation of an analogue of Slutsky's theorem, obtained in the theory of consumption, for the problem of the theory of the firm.

Keywords: firm problem; analogue of Slutsky's theorem.

1. Введение

Из экономической литературы известно, что в теории потребления одним из основных результатов является теорема Е. Е. Слуцкого, выражающая поведение потребителя в зависимости от одновременного изменения цен и бюджета (см., напр., [1–4]). Независимо от Слуцкого аналогичная теорема была доказана другими авторами, в частности, Дж. Хиксом [5], хотя при этом он подходил из других соображений, чем Слуцкий. Следует отметить, что именно в интерпретации Хикса в основном приводится в учебной литературе эта теорема и фамилия Хикса при этом не упоминается [6].

Если сравнить частные случаи задач фирмы с задачей потребления, они в чем-то схожи. Возникает вопрос: а не справедлива ли подобная теорема в теории производства (фирмы)? Следует подчеркнуть, что сравнение необходимо проводить именно с частными случаями задач фирмы (при фиксированном выпуске или фиксированных издержках), поскольку вывод об оптимальном поведении фирмы, полученный из общепринятой

постановки задачи производства, дает уравнения (см. ниже), из которых решение можно получить лишь при определенном соотношении цен, при котором прибыль равна нулю, что соответствует условиям равновесия, которое в реальной экономике не осуществимо [7, 8].

Возникают вопросы: а) как зависит цена на выпускаемую продукцию при изменении цен на производственные факторы, чтобы не изменилась, например, прибыль фирмы? б) как при этом зависят между собой показатели сравнительной статики теории фирмы?

В предлагаемой работе предпринята попытка получить некоторые ответы на поставленные вопросы.

В силу ограниченности объема данного сообщения приводится только результат для задачи фирмы с фиксированным выпуском.

2. Решение задачи фирмы при фиксированном выпуске

Как известно, основная задача фирмы в долгосрочном периоде в неоклассической теории заключается в максимизации *прибыли* на всем пространстве факторов:

$$\Pi(x) = p_0 f(x) - p'x \rightarrow \max, \quad x \in \mathbf{R}_+^m,$$

где p_0 – цена производимой фирмой продукции; x – m -вектор затрат; $p = (p_1, \dots, p_m)$ – вектор цен факторов производства; $f(x)$ – производственная функция (вогнутая), ставящая в соответствие каждому набору x затрат факторов максимальный объем Q выпуска продукции одного типа и удовлетворяющую определенным свойствам, в частности, в некоторой области пространства затрат (называемой экономической) увеличение использования затрат каждого фактора не приводит к уменьшению выпуска продукции, а в части экономической области (называемой особой) увеличение использования затрат каждого фактора ведет к увеличению выпуска; $R = p_0 f(x)$ – доход; $C = p'x$ – издержки производства. Заметим, что в экономической области функция $f(x)$ вогнутая, а в особой – строго вогнутая.

Предполагаем, что все векторы понимаются как вектор-столбцы, ' (штрих) – знак транспонирования, соотношения $x \geq y$ ($x \leq y$, $x = y$) для векторов одинаковой размерности означают покомпонентные соотношения, а все операции с векторами и матрицами подчиняются правилам векторно-матричного исчисления.

В силу сказанного во введении будем рассматривать задачу с фиксированным выпуском. Пусть в задаче фиксирован выпуск в объеме Q . Тогда задача примет вид:

$$p_0 Q - p'x \rightarrow \max, \quad f(x) = Q, \quad x \geq 0. \quad (1)$$

Введем функцию Лагранжа $L(x, \lambda_Q) = p_0 Q - p'x + \lambda_Q(f(x) - Q)$. Будем считать, что функция $f(x)$ строго вогнутая. Тогда задача (1) – задача выпуклого программирования. Пусть $x^0 = x^0(Q) > 0$ – решение задачи (1), λ_Q^0 – соответствующий множитель Лагранжа. Согласно теореме Куна-Таккера [9] будут выполняться равенства

$$\frac{\partial L(x^0, \lambda_Q^0)}{\partial x_i} = -p_i + \lambda_Q^0 \frac{\partial f(x^0)}{\partial x_i} = 0, i = \overline{1, m}; \quad (2)$$

$$\frac{\partial L(x^0, \lambda_Q^0)}{\partial \lambda_Q} = f(x^0) - Q = 0. \quad (3)$$

Из равенства (2) согласно определению предельного продукта получим

$$\lambda_Q^0 Mf_i(x^0) = p_i, i = \overline{1, m}. \quad (4)$$

Поскольку $p > 0$ и функция $f(x)$ строго вогнутая, то $\lambda_Q^0 > 0$. Очевидно, что если выпуск Q не фиксирован, то к системе (2), (3) добавится равенство

$$\frac{\partial L(x^0, \lambda_Q^0)}{\partial Q} = p_0 - \lambda_Q^0 = 0,$$

откуда $\lambda_Q^0 = p_0$. В этом случае равенство (4) примет известный из литературы вид

$$p_0 Mf_i(x^0) = p_i, i = \overline{1, m},$$

что, как указывалось выше, некорректно для реальной экономики, в которой фирма будет работать с нулевой прибылью. Если же фиксировать Q , то, решив систему (2), (3) относительно затрат, получим $\lambda_Q^0 = \lambda_Q^0(Q)$, $x^0 = x^0(Q)$.

Определив таким образом вектор затрат $x^0 = x^0(Q)$ как функцию выпуска, получим множество $\{x^0 : x^0 = x^0(Q), Q \geq 0\}$, которое называют *долгосрочным путем расширения*.

3. Показатели сравнительной статики теории фирмы

Рассмотрим теперь задачу (1) как семейство задач, зависящих от цен p_0, p . Решением этого семейства будет $x^0 = x^0(p_0, p)$, $p_0 \geq 0, p \geq 0$. Очевидно, в этом случае и выпуск зависит от цен:

$$Q(p_0, p) = f(x^0(p_0, p)) \equiv Q. \quad (5)$$

Как известно, чувствительность оптимальных затрат и выпуска к изменениям цен определяется величинами $\partial x^0 / \partial p_0, \partial x^0 / \partial p, \partial Q / \partial p_0, \partial Q / \partial p$, которые называют *показателями сравнительной статики теории фирмы*.

Будем считать, что функция $f(x)$ строго вогнута в особой области. В экономической литературе для семейства задач вида (1) подробно исследовано поведение спроса на затраты и предложения выпуска при отдельном изменении либо цены p_0 , либо цен $p_i, i = 1, \dots, m$.

К сожалению, автору данного сообщения неизвестны исследования поведения спроса и предложения при одновременном изменении цены на какой-либо фактор и на выпускаемую продукцию, чтобы при этом, например, не изменилась прибыль. Другими словами, как ведут себя спрос и предложение при так называемом компенсированном изменении цен. Обозначим эти величины $(\partial x^0 / \partial p)_{comp}, (\partial Q / \partial p)_{comp}$.

4. Уравнения сравнительной статики для задач с фиксированным выпуском

Используем подход к исследованию по аналогии с подходом Хикса в теории потребления. Рассмотрим следующую ситуацию. Изменяется цена p_i на какой-либо фактор. Как при этом надо изменить цену p_0 на выпускаемую продукцию, чтобы не изменилась прибыль, другими словами, чтобы выполнялось тождество

$$\Pi(p_0, p) = p_0 Q - p' x^0(p_0, p) \equiv const. \quad (6)$$

В этой ситуации $p_0 = p_0(p)$, причем $x^0 = x^0(p_0(p), p)$, а в силу (5) выполняется тождество

$$Q = Q(p_0(p), p) \equiv f(x^0(p_0(p), p)). \quad (7)$$

Заметим, что в таком случае, например, $\left(\partial x_i^0 / \partial p_j\right)_{comp}$ – это с математической точки зрения полная производная по p_j от x^0 , т.е. в матрично-векторной форме получим

$$\left(\frac{\partial x^0}{\partial p}\right)_{comp} = \frac{\partial x^0}{\partial p} + \frac{\partial x^0}{\partial p_0} \left(\frac{\partial p_0}{\partial p}\right)', \quad (8)$$

$$\left(\frac{\partial Q}{\partial p}\right)_{comp} = \frac{\partial Q}{\partial p} + \frac{\partial Q}{\partial p_0} \left(\frac{\partial p_0}{\partial p}\right)'. \quad (9)$$

Найдем величину $\partial p_0 / \partial p$. Дифференцируя тождества (6), (7), получим

$$\left(\frac{\partial p_0}{\partial p}\right)' Q - x^{0'} - p' \left(\frac{\partial x^0}{\partial p}\right)_{comp} = 0, \quad (10)$$

$$\frac{\partial f'}{\partial x} \left(\frac{\partial x^0}{\partial p}\right)_{comp} = 0. \quad (11)$$

Согласно (2) имеем $\partial f / \partial x = p / \lambda_Q^0$. Тогда равенство (11) примет вид

$$\frac{1}{\lambda_Q^0} p' \left(\frac{\partial x^0}{\partial p}\right)_{comp} = 0. \quad (12)$$

В силу (12) из равенства (10) для выражения $\partial p_0 / \partial p$ получим

$$\frac{\partial p_0}{\partial p} = \frac{x^0}{Q}. \quad (13)$$

Подставляя (13) в (8), (9), получим

$$\frac{\partial x^0}{\partial p} = \left(\frac{\partial x^0}{\partial p}\right)_{comp} - \frac{1}{Q} \frac{\partial x^0}{\partial p_0} x^{0'}, \quad (14)$$

$$\frac{\partial Q}{\partial p} = \left(\frac{\partial Q}{\partial p} \right)_{comp} - \frac{1}{Q} \frac{\partial Q}{\partial p_0} x^{0'}. \quad (15)$$

Уравнения (14), (15) и представляют *аналог уравнений Слуцкого в теории производства (фирмы)* в дифференциальной форме. Конечно, здесь использован подход, аналогичный схеме Хикса, а не Слуцкого, и правильнее было бы назвать эти уравнения *аналогом уравнений Хикса*.

Заметим, что уравнение (14) в дискретной постановке можно представить в следующем виде:

$$\Delta x_{(p)} = (\Delta x_{(p)})_{comp} - \Delta x_{(p_0)}. \quad (16)$$

Здесь $\Delta x_{(p)}$ – общий эффект от изменения цен на производственные факторы, $-\Delta x_{(p_0)}$ – эффект от изменения цены на выпускаемую продукцию, $(\Delta x_{(p)})_{comp}$ – эффект замены или компенсации.

Таким образом, уравнения (14), (15) (или (16)) в теории производства можно интерпретировать по аналогии с теоремой Слуцкого в теории потребления в виде следующего утверждения.

Теорема. *Общий эффект от изменения цен на производственные факторы равен сумме эффекта от изменения цены на выпускаемую продукцию и эффекта замены (компенсации).*

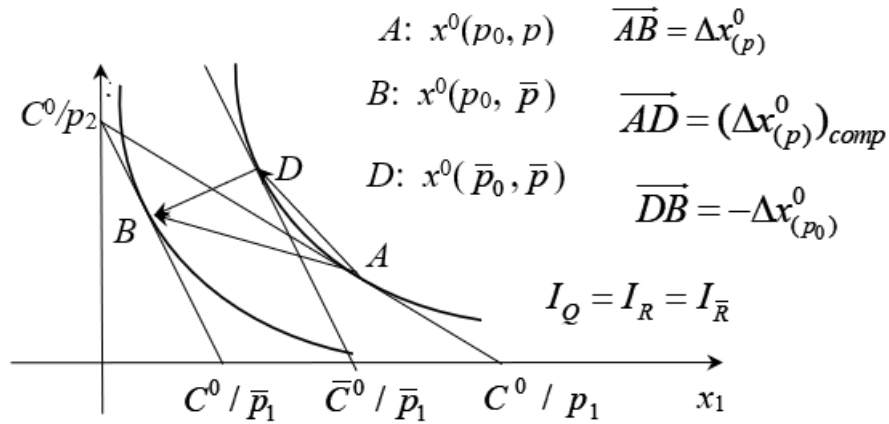
5. Графическая интерпретация основного результата

К сожалению, для семейства задач теории фирмы в общепринятой постановке основной результат в графической интерпретации представить невозможно. Как указывалось выше, в общем случае найти конкретное решение задачи невозможно, а лишь при определенном соотношении цен, при котором прибыль нулевая. Правда, при изучении поведения показателей сравнительной статики постановка семейства общепринятых задач несколько изменяется, т.е. добавляется тождество

$$Q^0 = Q^0(p_0(p), p) \equiv f(x^0(p_0(p), p),$$

которое аналогично тождеству (7), но здесь выпуск Q^0 не фиксирован. Однако в этом случае нельзя получить конкретное решение задачи. Другими словами, решение этих задач заполняет все пространство факторов при изменении цен. Поэтому рассмотрим случай фиксированного выпуска Q .

На рисунке представлена графическая интерпретация уравнения (14) (или (16)) в случае двух факторов при достаточно малых изменениях цен.



Приведем пояснения. Пусть меняется цена на первый фактор: увеличивается от p_1 до $\bar{p}_1$. Поскольку выпуск фиксирован, то получаем изокванту I_Q , любая точка которой соответствует набору факторов, использование которых определяет один и тот же уровень выпуска Q .

Точка A соответствует набору факторов $x^0(p_0, p)$ – решению задачи (1) при заданных ценах p_0, p .

Заметим, что решение задачи (1) не зависит от p_0 [7]. Поэтому, на самом деле, $x^0 = x^0(p, Q)$. Это – точка касания изокванты $I_R = \{x \in \mathbf{R}_+^2 : R = p_0 f(x) = p_0 Q\}$ с изокостой $I_{C^0} = \{x \in \mathbf{R}_+^2 : p_1 x_1 + p_2 x_2 = C^0\}$, где $C^0 = p'x^0(p, Q)$.

Как легко заметить, для изокванты I_Q справедливы соотношения: $I_Q = I_R = I_{\bar{R}}$, где I_R соответствует цене p_0 , $I_{\bar{R}}$ – цене $\bar{p}_0$.

Точка B соответствует точке касания изокосты $\bar{p}_1 x_1 + p_2 x_2 = C^0$ и изокванты $I_{\tilde{Q}} = I_{\tilde{R}}$, где $\tilde{R} = p_0 \tilde{Q}$, $\tilde{Q}$ – решение задачи с фиксированными издержками при $\bar{p} = (\bar{p}_1, p_2)$, $C = C^0$, которая в данном случае не приводится из-за объема сообщения.

Точка D соответствует решению задачи (1) с тем же выпуском Q и новыми ценами $\bar{p}_0, \bar{p}$, причем цена $\bar{p}_0$ такова, что прибыль не меняется. Как указывалось выше, точка D соответствует набору факторов $x^0 = x^0(\bar{p}, Q)$, так что $f(x^0(p, Q)) = f(x^0(\bar{p}, Q)) = Q$. Решив задачу (1) для нового вектора цен $\bar{p}$, найдем издержки $\bar{C}^0 = \bar{p}'x^0(\bar{p}, Q)$. Новая цена $\bar{p}_0$

определяется из условия $\Pi(p_0, p) \equiv \text{const}$, т.е. из равенства $p_0 Q - C^0 = \bar{p}_0 Q - \bar{C}^0$, откуда получим

$$\bar{p}_0 = p_0 + \frac{\bar{C}^0 - C^0}{Q}.$$

К сожалению, из-за ограниченного объема статьи здесь приведены лишь краткие выводы.

Библиографические ссылки

1. Слуцкий Е. К теории сбалансированного бюджета потребителя // Экономико-математические методы. Народнохозяйственные модели. Теоретические вопросы потребления. 1963. № 1. С. 241–277. (Пер. с ит.: *Slutsky E. Sulla teoria del consumatore* // *Giornale degli Economisti*. July, 1915.)
2. Альсевич В. В. Введение в математическую экономику. Кн. 1: Конструктивная теория: Учебное пособие. М. : Книжный дом «ЛИБРОКОМ», 2021.
3. Ашманов С. А. Введение в математическую экономику. М. : URSS, 2020.
4. Интрилигатор М. Математические методы оптимизации и экономическая теория. М. : Айрис-пресс, 2002.
5. Хикс Дж. Р. Стоимость и капитал. М. : Прогресс, 1993.
6. Альсевич В. В., Астровский А. И. К вопросу интерпретации компенсированных показателей сравнительной статики в микроэкономической теории // Анализ, моделирование, управление, развитие экономических систем : сб. науч. трудов IX Междунар. школы-симпоз. АМУР-2015, Севастополь, 12–21 сентября 2015 / Симферополь : КФУ им. В. И. Вернадского, 2015. С. 12–17.
7. Альсевич В. В. О решении задачи фирмы // Анализ, моделирование, управление, развитие экономических систем : сб. науч. трудов X Междунар. школы-симпоз. АМУР-2016, Симферополь-Судак, 12–21 сентября 2016 / Симферополь : КФУ им. В. И. Вернадского, 2016. С. 6–12.
8. Альсевич В. В. К вопросу об основном выводе теории фирмы // Тр. Междунар. научно-практич. конф. «Матем. методы и информ. технологии макроэкон. анализа и эконом. политики», посв. празднов. 80-лет. юбилея акад. НАН РК А. А. Ашимова, Алматы, 11–12 апр. 2017 г. / Алматы : Казах. нац. исследоват. техн. ун-т им. К. И. Сатпаева, 2017. С. 66–69.
9. Методы оптимизации / Р. Габасов [и др.]. Мн. : Четыре четверти, 2011.

ON CHARACTERISTIC PROPERTIES OF FACES OF CONVEX SETS

Valentin V. Gorokhovik

*Institute of Mathematics, The National Academy of Sciences of Belarus,
Minsk, Belarus, gorokh@im.bas-net.by*

In the talk three different characteristic properties of faces of infinite-dimensional convex sets are presented. For their formulations we use generalized semispaces, compatible total preorders, and real-valued step-affine functions. These characteristics extend to infinite-dimensional vector spaces the lexicographical characterization of faces established in finite-dimensional settings by Martinez-Legaz J.-E. (1997).

Keywords: convex sets; faces; semispaces; compatible preorders; step-affine functions.

О ХАРАКТЕРИСТИЧЕСКИХ СВОЙСТВАХ ГРАНЕЙ ВЫПУКЛЫХ МНОЖЕСТВ

Валентин В. Гороховик

*Институт математики Национальной академии наук Беларуси,
Минск, Беларусь, gorokh@im.bas-net.by*

В сообщении представлены три различные характеристические свойства граней бесконечномерных выпуклых множеств. Для их формулировок используются обобщенные семипространства, согласованные полные предпорядки и вещественнозначные ступенчато-аффинные функции. Эти характеристики распространяют на бесконечномерные векторные пространства лексикографическую характеристику граней, установленную в конечномерных рамках Мартинес-Легазом Х.-Е. (1997).

Ключевые слова: выпуклые множества; грани; семипространства; согласованные предпорядки; ступенчато-аффинные функции.

1. The Lexicographical Characterization of Faces of Convex Sets

Let X be a real vector space, Q a convex set in X .

A nonempty convex subset $F \subset Q$ such that $F \neq Q$ is called a *proper face* of Q if every segment of Q , which has an element of F in its relative interior, is entirely contained in F .

We say that a proper face F of a convex set Q *lexicographically exposed* if there exists a finite ordered family $\{l_1, l_2, \dots, l_m\}$ of nonzero linear functions defined on X such that

$$F = \{x \in Q \mid l_i(x) = \min_{y \in F_{i-1}} l_i(y), i = 1, 2, \dots, m\},$$

where $F_0 = Q, F_i = \{x \in F_{i-1} \mid l_i(x) = \min_{y \in F_{i-1}} l_i(y)\}, i = 1, 2, \dots, m$.

J.-E. Martinez-Legaz [1] proved that if a real vector space X is finite dimensional then every proper face of any convex set from X is lexicographically exposed. The authors of the paper [2] put the following question: is it possible to reasonably generalize the lexicographical characterization of faces to convex sets belonging to infinite-dimensional vector spaces? In the talk we give a positive answer to this question. We present three different characterizations of arbitrary proper faces of convex sets in infinite-dimensional spaces, which in the finite-dimensional setting are equivalent to the lexicographical characterization of faces established by J.-E. Martinez-Legaz in [1, Theorem 2]. The first characterization (see Theorem 1 below) is formulated in the terms of generalized semispaces, the second (Theorem 2) — in the terms of compatible total preorders, and the third (Theorem 3) — in the terms of step-affine functions.

2. The Characterization of Faces by Generalized Semispaces

Let M be an affine manifold from X . A *generalized semispace* generated by M is a maximal (by inclusion) convex set, whose intersection with M is empty, or, equivalently, if it is a maximal (by inclusion) convex set in $X \setminus M$. The existence of generalized semispaces, generated by an arbitrary affine manifold, is proved by using the Kuratowski-Zorn lemma. Moreover, it also follows from the Kuratowski-Zorn lemma, that for any convex set $C \subset X \setminus M$, where $M \subset X$, is an affine manifold from there exists a generalized semispace, generated by M , which contains C . If $S \subset X$ is a generalized semispace generated by some affine manifold then $X \setminus S$ also is a convex subset of X .

Theorem 1. Let Q be a convex set in a real vector space X . A subset $F \subset Q$ is a proper face of the set Q if and only if in X there exist a nonempty affine manifold $M \subset X$ and a generalized semispace $S \subset X$ generated by M such that $Q \subset M \cup S$ and $F = M \cap Q$.

3. The Characterization of Faces by Compatible Total Preorders

A preorder relation (a reflexive and transitive binary relation) $\leq$ defined on a real vector space X is said to be *compatible* with algebraic operations of the space X if

$$x \leq y \Rightarrow x + z \leq y + z \text{ for any } x, y, z \in X,$$

and

$$x \leq y \Rightarrow \lambda x \leq \lambda y \text{ for any } x, y \in X \text{ and any } \lambda > 0.$$

If $\leq$ is a compatible preorder relation then

$$x \leq y \Leftrightarrow y - x \in P_{\leq},$$

where $P_{\leq} = \{x \in X \mid 0 \leq x\}$ is the set of positive elements of the relation $\leq$, $P_{\leq}$ is a convex cone such that $0 \in P_{\leq}$.

A preorder $\leq$ defined on X is called *total* if for any $x, y \in X$ at least one of corelations $x \leq y$ or $y \leq x$ holds. A compatible preorder relation $\leq$ is total if and only if $P_{\leq} \cup (-P_{\leq}) = X$.

Theorem 2. Let Q be a convex set in a real vector space X .

A subset $F \subset Q$ is a proper face of the set Q if and only if there exists a compatible total preorder relation $\leq$ defined on X such that $F = \text{Min}(Q \mid \leq) := \{x \in Q \mid x \leq y \text{ for all } y \in Q\}$.

4. The Characterization of Faces by Step-Affine Functions

Let $A(X, R)$ be the vector space of real-valued affine functions defined on a vector space X .

For any family $\mathfrak{F} \subset A(X, R)$, which is linearly ordered by some order relation $\prec_{\mathfrak{F}}$, and for any $f \in \mathfrak{F}$ we define the set (affine manifold)

$$E_f := \{x \in X \mid \bar{f}(x) = 0 \text{ for all } \bar{f} \in \mathfrak{F} \text{ such that } \bar{f} \prec_{\mathfrak{F}} f\}.$$

If f is the smallest element of $\mathfrak{F}$ then we set $E_f = X$.

We say that a family $\mathfrak{F} \subset A(X, R)$ forms a *cortege of affine functions* on X if

- i) $\mathfrak{F}$ is linearly ordered by some order relation $\prec_{\mathfrak{F}}$;
- ii) for every function $f \in \mathfrak{F}$ the corresponding affine manifold E_f is non-empty and $f(E_f) = R$;

iii) for any $x \in X$ the subfamily $\mathfrak{F}_x = \{f \in \mathfrak{F} \mid f(x) \neq 0\}$ is either empty or has a least (with respect to $\prec_{\mathfrak{F}}$) element f_x .

A real-valued function $u: X \rightarrow R$ is called *step-affine* if there exists a cortege of affine functions $\mathfrak{F}$ such that

$$u(x) = u_{\mathfrak{F}}(x) := \begin{cases} 0, & \text{if } \mathfrak{F}_x = \emptyset, \\ f_x(x), & \text{if } \mathfrak{F}_x \neq \emptyset. \end{cases}$$

A step-affine function $u: X \rightarrow R$ is called *regular* if the set of points from X at which it takes the zero value is non-empty.

Theorem 3. Let Q be a convex set in a real vector space X .

A subset $F \subset Q$ is a proper face of the set Q if and only if there exists a regular step-affine function $u: X \rightarrow R$ defined on X such that $Q \subset \{x \in X \mid u(x) \geq 0\}$ and $F = \{x \in X \mid u(x) = 0\}$.

The proofs of the theorems presented here and more detailed discussion of the results can be found in the preprint [3].

Acknowledgments. The research was supported by the State Program for Fundamental Research of Republic of Belarus.

References

1. *Martinez-Legaz J.-E.* Lexicographical characterization of the faces of convex sets // *Acta Mathematica Vietnamica*. 1997. Vol. 22, no.1. P. 207–211.
2. *Diaz Millan R., Roshchina V.* The Intrinsic Core and Minimal Faces of Convex Sets in General Vector Spaces // *Set-Valued and Variational Analysis*. 2023. Article 14.
3. *Gorokhovik V. V.* Characterizations of Faces of Convex Sets in Infinite-dimensional Vector Spaces. URL: <https://arxiv.org/abs/2506.08742> (date of access: 13.09.2025).

СУБОПТИМАЛЬНЫЕ ОБРАТНЫЕ СВЯЗИ В ЛИНЕЙНЫХ ЗАДАЧАХ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ С ВОЗМУЩЕНИЯМИ

М. А. Готовец, Н. М. Дмитрук

*Белорусский государственный университет,
Минск, Беларусь, hatavets@bsu.by, dmitrukn@bsu.by*

Рассматривается задача оптимального управления линейными системами с неизвестными ограниченными возмущениями. Для синтеза субоптимальных обратных связей в исследуемой задаче предлагается использовать методы машинного обучения, в частности метод опорных векторов, который применяется для классификации данных, генерируемых оптимальными регуляторами при управлении в реальном времени.

Ключевые слова: субоптимальная обратная связь; линейные системы; возмущения; метод опорных векторов.

SUBOPTIMAL FEEDBACK CONTROLS IN LINEAR OPTIMAL CONTROL PROBLEMS WITH DISTURBANCES

M. A. Gotovets, N. M. Dmitruk

*Belarusian State University,
Minsk, Belarus, hatavets@bsu.by, dmitrukn@bsu.by*

The problem of optimal control for linear systems with unknown bounded disturbances is considered. To synthesize suboptimal feedback controls in the problem under study, it is proposed to use machine learning methods, specifically the Support Vector Machine (SVM) method, which is applied to classify data generated by optimal controllers during real-time control.

Keywords: suboptimal feedback; linear systems; disturbances; support vector machine.

1. Введение

Проблема синтеза оптимальных обратных связей остается одной из центральных в современной теории управления. Классические методы, такие как принцип максимума Л. С. Понтрягина, дают возможность получать решения в замкнутой форме для относительно простых моделей невысокого порядка. Динамическое программирование представляет универсальный подход для решения задачи синтеза, однако его практическое применение ограничено высокой вычислительной сложностью, известной как

«проклятие размерности». Наличие неопределенностей в системе управления, таких как аддитивные ограниченные возмущения, рассматриваемые в настоящем сообщении, приводят к минимаксным постановкам задач оптимального управления, что еще больше усложняет задачу синтеза.

Одним из современных подходов к решению проблемы синтеза является управление в режиме реального времени [1]. В его основе лежит итеративное решение задачи оптимального управления: на каждом шаге по текущему состоянию системы и моменту времени формируется программное управление на весь будущий промежуток времени, однако используется только ее начальное значение, которое и задает значение оптимальной обратной связи для текущей позиции. Остальные значения оптимальной программы игнорируются.

При реализации подобных схем оптимального управления для систем с возмущениями формируется значительный объем вспомогательных данных, которые напрямую в управление не включаются и обычно остаются невостребованными. Их целенаправленная обработка открывает возможность построения субоптимальных обратных связей, обеспечивающих баланс между качеством управления и вычислительной сложностью. В этой связи представляет интерес применение методов машинного обучения, позволяющих выделять закономерности в накопленных данных и предсказывать законы обратной связи. В настоящей работе основное внимание уделяется методу опорных векторов (Support Vector Machine, SVM, [2]), использование которого дает возможность разработать конструктивный подход к синтезу субоптимальных регуляторов для линейных систем с аддитивными возмущениями [3].

2. Задача оптимального гарантированного управления

Пусть задан объект управления, математическая модель которого на промежутке времени $T = [0, t_f]$ имеет вид

$$\dot{x} = Ax + bu + dw, \quad x(0) = x_0, \quad (1)$$

где $x = x(t) \in R^n$ — состояние в момент времени t ; $u = u(t) \in R$ — значение скалярного управления; $w = w(t) \in R$ — неизвестное ограниченное возмущение; $A \in R^{n \times n}$, $b, d \in R^n$.

Управления u выбираются из класса дискретных управлений [3, 4]:

$$u(t) \equiv u(s), \quad t \in [s, s + h[, \quad s \in T_h = \{0, h, \dots, t_f - h\},$$

где $h = t_f / N$ — период квантования времени, $N \in \mathbb{N}$, $N > 1$.

На абсолютные значения управления наложены ограничения:
 $|u(t)| \leq 1, t \in T$.

Для (1) в классе дискретных управлений будем исследовать терминальную задачу оптимального управления:

$$\begin{aligned} \min_u \max_{w \in W} c^\top x(t_f), \\ \dot{x} = Ax + bu + dw, \quad x(0) = x_0, \quad |u(t)| \leq 1, \quad t \in T, \\ x(t_f) \in X_f \quad \forall w(\cdot) \in W, \end{aligned} \quad (2)$$

где $c \in R^n$; $X_f = \{x \in R^n : Hx \leq g\}$, $H \in R^{m \times n}$, $g \in R^m$, $W = \{w(\cdot) : |w(t)| \leq w^*, t \in T\}$ — множество возможных возмущений.

В задаче (2) требуется выполнить терминальные ограничения с гарантией, т.е. при всех возможных реализациях возмущения $w(\cdot) \in W$, а также минимизировать значение критерия качества при наихудшем возмущении. Такие задачи носят название задач оптимального гарантированного управления [3].

Предлагаемый в работе подход можно развить и на другие задачи оптимального управления линейными системами с линейными или квадратичными критериями качества, см. [5]. Задача (2) выбрана как простейшая для демонстрации предлагаемых идей. Такие задачи хорошо изучены в литературе с точки зрения построения численных методов программного решения и реализации алгоритмов работы оптимальных регуляторов, в которых учитывается структура оптимального программного управления [3, 4].

Понятия допустимой гарантирующей программы и оптимальной гарантирующей программы $u^0(t)$, $t \in T$, для задачи (2) вводятся стандартно [3].

Оптимальную обратную связь (позиционное решение) будем определять в предположении что состояния рассматриваемого объекта доступны для измерения в дискретные моменты $\tau \in T_h$. Из [3] следует, что для задач оптимального управления, содержащих неопределенности, в том числе, возмущения, можно определить различные типы оптимальных обратных связей. Простейшей является так называемая размыкаемая обратная связь. При ее определении не учитывается тот факт, что в будущем возможно поступление дополнительной информации о поведении системы, т.е. обратная связь определяется на основе оптимальных гарантирующих программ [3].

В соответствии с этим погрузим задачу (2) в семейство задач

$$\min_u \max_{w \in W} c^\top x(t_f),$$

$$\begin{aligned}\dot{x} &= Ax + bu + dw, x(\tau) = x_\tau, |u(t)| \leq 1, \quad t \in T(\tau) = [\tau, t_f], \\ x(t_f) &\in X_f \quad \forall w(\cdot) \in W(\tau),\end{aligned}\quad (3)$$

зависящее от позиции (τ, x_τ) , где $\tau \in T_h$ и $x_\tau \in R^n$. Здесь $W(\tau) = \{w(\cdot) : |w(t)| \leq w^*, t \in T(\tau)\}$ — сужение W на промежутки управления $T(\tau)$.

Пусть $u^0(t | \tau, x_\tau)$, $t \in T(\tau)$, — оптимальная гарантирующая программа [3] задачи (3) для позиции (τ, x_τ) ; X_τ — множество векторов $x_\tau \in \mathbb{R}^n$, для которых в момент τ оптимальная гарантирующая программа существует.

Функция

$$u^0(\tau, x) = u^0(\tau | \tau, x), x \in X_\tau, \tau \in T_h, \quad (4)$$

называется [3] оптимальной размыкаемой обратной связью по состоянию.

Построение оптимальной обратной связи (4) в явном виде зачастую невозможно. Для преодоления трудностей, связанных с построением функции (4), в работах [1, 3, 4] предложен и развит метод, получивший название управления в режиме реального времени. В настоящей работе на основе исследования свойств оптимальной обратной связи предлагается подход, основанный на методах классификации данных. Он развивает результаты [5] на случай задач оптимального гарантированного управления линейными системами с возмущениями.

3. Эквивалентная детерминированная задача

Из [3] известно, что каждая задача семейства (3) может быть сведена к детерминированной задаче оптимального управления вида

$$\begin{aligned}\min_u c^\top x_0(t_f), \\ \dot{x}_0 &= Ax_0 + bu, x_0(\tau) = x_\tau, \quad |u(t)| \leq 1, \quad t \in T(\tau), \\ x_0(t_f) &\in X_f^\gamma(\tau) = \{x \in R^n : Hx \leq g - \gamma(\tau)\},\end{aligned}\quad (5)$$

где $\gamma(\tau) = (\gamma_1(\tau), \dots, \gamma_m(\tau)) \in R^m$, $\gamma_i = w^* \int_\tau^{t_f} |h_i^\top e^{A(t_f-t)} b| dt$, h_i — i -ая строка матрицы H .

В классе дискретных управлений задача (5) сводится к задаче линейного программирования

$$\begin{aligned}
& \min_u \sum_{s \in T_h(\tau)} c(s)u(s), \\
& \sum_{s \in T_h(\tau)} d(s)u(s) \leq g_0(\tau), \\
& |u(t)| \leq 1, \quad t \in T_h(\tau),
\end{aligned} \tag{6}$$

где $g_0(\tau) = g - \gamma(\tau) - He^{A(t_f - \tau)}x_\tau$,

$$c(s) = \int_s^{s+h} c^\top e^{A(t_f - t)} b dt, \quad d(s) = \int_s^{s+h} He^{A(t_f - t)} b dt, \quad s \in T_h(\tau) = T(\tau) \cap T_h.$$

4. Субоптимальная обратная связь

Рассматривая задачу (6) как задачу многопараметрического линейного программирования относительно параметра x_τ и применяя результаты монографии [6], нетрудно показать, что в задаче оптимального управления (2) оптимальная обратная связь $u^0(\tau, x)$, $x \in X_\tau$, $\tau \in T_h$, является нестационарной кусочно-аффинной. Для ее построения необходимо найти многогранные критические области [6]

$$K_\tau^l = \{x \in X_\tau : H_\tau^l x \leq g_\tau^l\}, \quad l \in L_\tau, \tag{7}$$

и для каждой области указать аффинный закон обратной связи:

$$u^0(\tau, x) = (k_\tau^l)^\top x + a_\tau^l, \quad x \in K_\tau^l, \quad l \in L_\tau, \tag{8}$$

где $H_\tau^l \in R^{m_\tau^l \times n}$, $g_\tau^l \in R^{m_\tau^l}$, $k_\tau^l \in R^n$, $a_\tau^l \in R$.

Далее будем строить субоптимальную обратную связь, применяя метод опорных векторов [2], для классификации (для каждого $\tau \in T_h$) данных выборки

$$\{x_\tau^i, y_\tau^i\}_{i=1}^{M_\tau}, \tag{9}$$

где x_τ^i — состояние из X_τ , $y_\tau^i \in L_\tau$ — метка класса, которому принадлежит точка x_τ^i , M_τ — объем выборки.

Опишем процесс получения выборки (9).

Точки $x_\tau^i \in X_\tau$, $i = \overline{1, M_\tau}$ в простейшем случае могут быть выбраны случайным образом или на основе результатов управления в реальном времени в ряде конкретных процессов управления динамическим объектом (1).

Далее для всех позиций (τ, x_τ^i) , $i = \overline{1, M_\tau}$, $\tau \in T_h$, решается задача (6). Если оптимальная программа $u^0(t | \tau, x_\tau^i)$, $t \in T(\tau)$, имеет ту же структуру, что и программное решение для некоторой другой точки x_τ^j , то эти две точки принадлежат одной критической области в X_τ и точке x_τ^i присваивается та же метка, что и у точки x_τ^j : $y_\tau^i := y_\tau^j$. Иначе заводится новая уникальная метка y_τ^i . Будем считать, что после решения задач (6) для всех точек x_τ^i , $i = \overline{1, M_\tau}$, метки пронумерованы от 1 до L_τ , т.е. $y_\tau^i \in \{1, \dots, L_\tau\}$. Таким образом будет построена требуемая выборка вида (9).

В процессе построения выборки (9) будем также запоминать первое значение оптимальной программы:

$$u_\tau^i := u^0(\tau | \tau, x_\tau^i).$$

Таким образом, наряду с (9) будет построена выборка

$$\{x_\tau^i, u_\tau^i\}_{i=1}^{M_\tau}. \quad (10)$$

Точки $x_\tau^i, x_\tau^{i'}$, имеющие одинаковую метку (для определенности будем считать, что $y_\tau^i = y_\tau^{i'} = l$), принадлежат одной критической области K_τ^l и одному классу l . Множество индексов точек, принадлежащих классу l для момента времени τ будем обозначать $I_\tau(l) = \{i : y_\tau^i = l\}$. Не ограничивая общности, будем предполагать, что каждый класс содержит не менее $n + 1$ точки.

Поскольку все критические области являются выпуклыми многогранниками, существует разделяющая гиперплоскость $\Gamma = \{x \in R^n : \omega^\top x + b = 0\}$, $\omega \neq 0$, такая, что $\omega^\top x_\tau^i + b \leq 0$ при $i \in I_\tau(l)$, и $\omega^\top x_\tau^i + b \geq 0$ при $i \in I_\tau(j)$, $j \neq l$.

Применяя метод опорных векторов [2] для каждого класса l строится набор гиперплоскостей, разделяющих классы l и $j \in C_\tau(l)$, где $C_\tau(l)$ — совокупность номеров классов, соседних с l . В результате будут получены аппроксимации критической области K_τ^l и оптимальной обратной связи (8).

5. Заключение

В работе исследована терминальная задача оптимального управления линейными системами с возмущениями с терминальными ограничениями и ограниченными дискретными управлениями. Предложен метод построения субоптимальной размыкаемой обратной связи, использующий метод опорных векторов для классификации состояний и соответствующих им ответов о принадлежности состояния некоторой критической области. Линейность обратной связи и тот факт, что критические области являются выпуклыми многогранниками, позволяет гарантировать линейную разделимость данных при классификации.

Библиографические ссылки

1. Кириллова Ф. М., Дмитрук Н. М., Габасов Р. Синтез оптимальных систем — оптимальное управление в реальном времени // Динамика систем и процессы управления. 2015. С. 208–219.
2. Дайзенрот М. П., Альдо Ф. А., Чен С. О. Математика в машинном обучении. СПб. : Питер, 2024.
3. Балашевич Н. В., Габасов Р., Кириллова Ф. М. Построение оптимальных обратных связей по математическим моделям с неопределенностью // Ж. выч. матем. и матем. физики. 2004. Т. 44, № 2. С. 265–286.
4. Балашевич Н. В., Габасов Р., Кириллова Ф. М. Численные методы программной и позиционной оптимизации линейных систем управления // Ж. выч. матем. и матем. физики. 2000. Т. 40, № 6. С. 838–859.
5. Дмитрук Н. М., Готовец М. А. Применение метода опорных векторов для синтеза субоптимальных обратных связей в линейной задаче терминального управления // Изв. Иркутского гос. ун-та. Сер. «Математика». 2023. Т.46. С. 19–34.
6. Borrelli, F. Constrained Optimal Control of Linear and Hybrid Systems. Berlin, Heidelberg : Springer Berlin Heidelberg, 2003.

СИСТЕМА ОБСЛУЖИВАНИЯ С РЕЗЕРВИРОВАНИЕМ ПРИБОРОВ ДЛЯ ПОВТОРНЫХ ЗАПРОСОВ

А. Н. Дудин, С. А. Дудин

*Белорусский государственный университет,
Минск, Беларусь, dudin@bsu.by, dudins@bsu.by*

Рассматривается система типа *MAP/PH/N* с возможностью повторных обслуживаний. Считается, что после обслуживания запрос может покинуть систему или отправиться на орбиту и совершать попытки получить повторное обслуживание. В зависимости от числа запросов на орбите некоторое число приборов может быть зарезервировано для обслуживания повторных запросов. Поведение системы описано многомерной цепью Маркова. Найдены основные характеристики производительности системы.

Ключевые слова: повторное обслуживание; резервирование приборов; стационарные вероятности; цепь Маркова; вероятность потери.

QUEUEING SYSTEM WITH SERVER RESERVATION FOR REPEAT CUSTOMERS

A. N. Dudin, S. A. Dudin

*Belarusian State University,
Minsk, Belarus, dudin@bsu.by, dudins@bsu.by*

A *MAP/PH/N*-type queueing system with the possibility of repeated services is considered. It is assumed that after the service, a request can leave the system or go to orbit and make attempts to receive service again. Depending on the number of requests in orbit, some number of servers can be reserved for repeated requests. The behavior of the system is described by a multidimensional Markov chain. The main characteristics of the system performance are obtained.

Keywords: repeated service; server reservation; stationary probabilities; Markov chain; loss probability.

1. Введение

Системы массового обслуживания с повторными вызовами характеризуются тем, что в них нет места для ожидания прибывающих запросов. В случае недоступности приборов в момент прибытия запрос может переместиться в виртуальное место, известное как орбита, и попытаться получить обслуживание через случайные интервалы времени. Монография [1]

содержит примеры применения теории систем с повторными вызовами для анализа реальных систем. Обзоры исследований систем с повторными вызовами приведены в [2–4].

В отличие от классических систем с повторными вызовами в данной работе исследуется система с повторными обслуживаниями. Другими словами, на орбиту идут не запросы, которые не получили обслуживание из-за занятости всех приборов в момент их поступления, а запросы, получившие обслуживание. Считаем, что данные запросы являются своего рода постоянными клиентами и они более ценны для системы, чем первичные запросы. Чтобы обеспечить приоритет для обслуживания таких запросов, предполагается резервирование части приборов для обслуживания исключительно повторных запросов. Резервирование является динамическим, то есть число зарезервированных приборов зависит от числа повторных запросов в системе.

Процесс поступления запросов описывается марковским входным потоком, что позволяет учитывать возможную корреляцию длин интервалов между моментами поступления запросов, свойственную входным потокам в современных системах и сетях. Также предполагается, что времена обслуживания первичных и вторичных запросов имеют распределение фазового типа, что позволяет учитывать коэффициент вариации времен обслуживания.

Для большей адекватности реальным системам в данной работе считается, что повторные запросы являются нетерпеливыми, то есть они могут покинуть систему без обслуживания, а также ненастойчивыми, то есть они могут покинуть систему после неудачной попытки попасть на обслуживание.

Целью исследования является разработка методов для нахождения стационарных характеристик систем с повторными обслуживаниями, что позволило бы определять оптимальные стратегии динамического резервирования приборов.

2. Математическая модель

Рассмотрим многолинейную систему массового обслуживания без буфера, состоящую из N приборов.

Предполагается, что входящий поток первичных запросов в систему является марковским входным потоком (*МАР*-потоком), который управляется цепью Маркова $v_t, t \geq 0$, с конечным пространством состояний $\{1, 2, \dots, W\}$. Интенсивности переходов этого процесса в пространстве состояний определяются квадратными матрицами D_0 и D_1 размера W . Элементы матрицы D_1 определяют интенсивности переходов, во время

которых происходит поступление запроса. Недиagonальные элементы матрицы D_0 определяют интенсивности переходов между состояниями, во время которых поступление запросов не происходит. Диагональные элементы матрицы D_0 отрицательны. Их модули определяют интенсивность выхода из соответствующего состояния. Сумма матриц $D_0 + D_1$ является генератором цепи Маркова $v_t, t \geq 0$.

Средняя интенсивность поступления запросов λ определяется по формуле $\lambda = \theta D_1 e$, где θ – вектор-строка стационарных вероятностей цепи Маркова $v_t, t \geq 0$. Этот вектор является единственным решением системы $\theta(D_0 + D_1) = 0, \theta e = 1$. Здесь и далее 0 – вектор-строка соответствующего размера, состоящая из нулей, а e – вектор-столбец соответствующего размера, состоящий из единиц.

Более подробную информацию о *МАР*-потоке можно найти в [5–7].

Если в момент прибытия все приборы заняты, запрос теряется, в противном случае он начинает обслуживание.

После обслуживания запрос покидает систему с вероятностью p , а с дополнительной вероятностью переходит на орбиту неограниченного размера (становится вторичным запросом), откуда пытается снова обслужиться независимо от других запросов на орбите через случайные промежутки времени. Если в произвольный момент времени количество запросов на орбите составляет $i, i > 0$, то общая интенсивность повторных попыток равняется $i\alpha, \alpha > 0$. Попытка считается успешной, если есть хотя бы один свободный прибор. Если попытка неуспешна, вторичный запрос покидает систему с вероятностью q , и с дополнительной вероятностью $1 - q$ остается на орбите.

Мы предполагаем, что некоторое количество приборов (до K) зарезервировано для обслуживания вторичных запросов. Оно зависит от количества запросов на орбите. Зафиксируем некоторые пороги $G_1, G_2, \dots, G_K$. Если число запросов на орбите находится в интервале $[G_k, G_{k+1})$, то $k, k = \overline{0, K}$, приборов зарезервированы для вторичных запросов, $G_0 = 0, G_{K+1} = \infty$.

Время обслуживания первичного и вторичного запроса любым прибором имеет распределение фазового типа с неприводимым представлением (β_1, S) и (β_2, S) соответственно. Это распределение задается цепью Маркова $m_t, t \geq 0$, с пространством переходных состояний $\{1, 2, \dots, M\}$ и поглощающим состоянием $M + 1$. Время, имеющее такое распределение, можно рассматривать как время блуждания некоторого запроса в сети, состоящей из M узлов. Стохастический вектор-строка $\beta_l, l = 1, 2,$

размерности M определяет начальное положение первичного или вторичного запроса в сети в момент начала блуждания. Квадратная матрица S размерности M является субгенератором, определяющим интенсивности переходов цепи m_t , $t \geq 0$, внутри множества $\{1, 2, \dots, M\}$. Переход в поглощающее состояние означает окончание обслуживания. Интенсивности перехода в поглощающие состояния определяются вектором-столбцом $S_0 = -Se$. Среднее время обслуживания первичного запроса определяется как $b_1^{(1)} = \beta_1(-S)^{-1}e$, среднее время обслуживания вторичного запроса вычисляется как $b_1^{(2)} = \beta_2(-S)^{-1}e$.

Вторичные запросы могут быть нетерпеливыми и покидать орбиту и систему без обслуживания, независимо друг от друга, через экспоненциально распределенное с параметром γ , $\gamma > 0$, время.

3. Процесс состояний системы и его генератор

Поведение исследуемой системы описывается регулярной неприводимой цепью Маркова с непрерывным временем

$$\xi_t = \{i_t, n_t, v_t, m_t^{(1)}, \dots, m_t^{(M)}\}, \quad t \geq 0,$$

где

i_t — количество вторичных запросов, $i_t \geq 0$,

n_t — количество занятых приборов, $n_t = \overline{0, N}$,

v_t — состояние управляющего процесса *МАР*-потока первичных запросов, $v_t = \overline{1, W}$,

$m_t^{(l)}$ — количество приборов на l -й фазе обслуживания, $m_t^{(l)} = \overline{0, n_t}$,

$l = \overline{1, M}$, $\sum_{l=1}^M m_t^{(l)} = n_t$, в момент времени $t, t \geq 0$.

Теорема 1. Инфинитезимальный генератор Q цепи Маркова ξ_t , $t \geq 0$, имеет блочно-тридиагональную структуру

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & O & O & \dots \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & O & \dots \\ O & Q_{2,1} & Q_{2,2} & Q_{2,3} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

где его ненулевые блоки $Q_{i,j}$, $|i-j| \leq 1$, $i, j \geq 0$, определяются как

• диагональные блоки $Q_{i,i}$ имеют блочно-тридиагональную структуру $Q_{i,i} = (Q_{i,i}^{(n,n')})$, $|n - n'| \leq 1$, $n, n' = \overline{0, N}$, с ненулевыми блоками:

$$\begin{aligned} Q_{i,i}^{(n,n)} &= D_0 \oplus (A_n + \Delta_n) + \delta_{n \geq N-k} D_1 \otimes I_{T_n} - \\ &- i(\delta_{n < N} \alpha + q \delta_{n=N} \alpha + \gamma) I_{W_{T_n}}, n = \overline{0, N}, i \in [G_k, G_{k+1}), k = \overline{0, K}, \\ Q_{i,i}^{(n,n+1)} &= D_1 \otimes P_n(\beta_1), n = \overline{0, N-k-1}, i \in [G_k, G_{k+1}), k = \overline{0, K}, \\ Q_{i,i}^{(n,n-1)} &= p I_W \otimes L_n, n = \overline{1, N}, i \geq 0; \end{aligned}$$

• наддиагональные блоки $Q_{i,i+1}$ имеют только ненулевые поддиагональные блоки $Q_{i,i+1}^{(n,n-1)}$, определяемые как

$$Q_{i,i+1}^{(n,n-1)} = (1-p) I_W \otimes L_n, n = \overline{1, N}, i \geq 0;$$

• поддиагональные блоки $Q_{i,i-1}$ имеют ненулевые наддиагональные блоки $Q_{i,i-1}^{(n,n+1)}$, заданные формулой

$$Q_{i,i-1}^{(n,n+1)} = i \alpha I_W \otimes P_n(\beta_2), n = \overline{0, N-1}, i \geq 1,$$

и диагональные блоки

$$Q_{i,i-1}^{(n,n)} = i(\gamma + \delta_{n=N} q \alpha) I_{W_{T_n}}, n = \overline{0, N}, i \geq 1.$$

Здесь

$\otimes$ и $\oplus$ – символы Кронекеровых произведения и суммы матриц, см., например, [8];

I – единичная матрица, а O – нулевая матрица, размерность которой указывается нижним индексом при необходимости;

$$\delta_a = \begin{cases} 1, & \text{если } a \text{ верно;} \\ 0, & \text{в обратном случае.} \end{cases}$$

Вспомогательные матрицы L_n , $n = \overline{1, N}$, Δ_n , A_n , $n = \overline{0, N}$, и $P_n(\beta_l)$, $l = 1, 2$, $n = \overline{0, N-1}$, вычисляются с помощью алгоритмов для матриц с тем же названием и смыслом, представленных в [9].

Размер T_n матриц, представленных выше, определяется как

$$T_n = \frac{(n + M - 1)!}{n!(M - 1)!}, \quad n = \overline{1, N}.$$

Доказательство теоремы проводится путем анализа интенсивностей всех возможных переходов цепи Маркова ξ_t , $t \geq 0$, за бесконечно малый интервал времени.

Следствие 1. Цепь Маркова ξ_t , $t \geq 0$, принадлежит классу асимптотически квазитеплицевых цепей Маркова (см. [10]).

Используя результаты [11], легко показать, что цепь Маркова ξ_t , $t \geq 0$, является эргодической, если вторичные запросы являются нетерпеливыми ($\gamma > 0$) и/или ненастойчивыми ($q > 0$). Значит стационарные вероятности состояний цепи

$$\pi(i, n, v, m^{(1)}, \dots, m^{(M)}) = \lim_{t \rightarrow \infty} P\{i_t = i, n_t = n, v_t = v, m_t^{(1)} = m^{(1)}, \dots, m_t^{(M)} = m^{(M)}\},$$

$$i \geq 0, n = \overline{0, N}, v = \overline{1, W}, m^{(l)} = \overline{0, n}, l = \overline{1, M}, \sum_{l=1}^M m^{(l)} = n,$$

существуют для всех возможных значений параметров системы.

Сформируем векторы-строки π_i из этих вероятностей, перенумерованных в прямом лексикографическом порядке компонент n, v и обратном лексикографическом порядке компонент $m^{(1)}, \dots, m^{(M)}$.

Хорошо известно, что эти векторы удовлетворяют следующей системе линейных алгебраических уравнений:

$$(\pi_0, \pi_1, \dots)Q = 0, \quad (\pi_0, \pi_1, \dots)e = 1,$$

где Q – инфинитезимальный генератор цепи Маркова ξ_t , $t \geq 0$. Для решения этой системы рекомендуется использовать эффективный и численно устойчивый алгоритм, разработанный в [12].

4. Характеристики производительности

Среднее количество вторичных запросов рассчитывается по формуле

$$L^{retr} = \sum_{i=1}^{\infty} i \pi_i e.$$

Среднее количество занятых приборов рассчитывается по формуле

$$N^{serv} = \sum_{i=0}^{\infty} \sum_{n=1}^N n \pi(i, n) e.$$

Среднее общее количество запросов в системе (первичных и вторичных) рассчитывается по формуле

$$L^{sys} = \sum_{i=0}^{\infty} \sum_{n=0}^N (i + n) \pi(i, n) e = L^{retr} + N^{serv}.$$

Средняя интенсивность выходного потока успешно обслуженных запросов рассчитывается по формуле

$$\mu^{out} = \sum_{i=0}^{\infty} \sum_{n=1}^N \pi(i, n) (I_W \otimes L_n) e.$$

Вероятность того, что произвольная повторная попытка будет успешной, рассчитывается как

$$p^{suc-ret} = \frac{\sum_{i=1}^{\infty} \sum_{n=0}^{N-1} i \pi(i, n) e}{\sum_{i=1}^{\infty} \sum_{n=0}^N i \pi(i, n) e}.$$

Вероятность потери первичного запроса по прибытии рассчитывается как

$$p^{primary-loss} = \frac{1}{\lambda} \sum_{k=0}^K \sum_{i=G_k}^{G_{k+1}-1} \sum_{n=N-k}^N \pi(i, n) (D_1 \otimes I_{T_n}) e.$$

Вероятность потери вторичного запроса из-за нетерпеливости рассчитывается как

$$p^{retr-imp-loss} = \frac{1}{(1-p)\mu^{out}} \sum_{i=1}^{\infty} i \gamma \pi_i e.$$

Вероятность потери вторичного запроса из-за нетерпеливости рассчитывается как

$$P^{retr-non-pers-loss} = \frac{q \sum_{i=1}^{\infty} i \alpha \pi(i, N) e}{(1-p) \mu^{out}}.$$

Вероятность потери вторичного запроса рассчитывается как

$$P^{retr-loss} = P^{retr-imp-loss} + P^{retr-non-pers-loss}.$$

Вероятность потери произвольного запроса рассчитывается как

$$\begin{aligned} P^{loss} &= \frac{1}{\lambda + (1-p) \mu^{out}} (\lambda P^{primary-loss} + (1-p) \mu^{out} P^{retr-loss}) = \\ &= 1 - \frac{\mu^{out}}{\lambda + (1-p) \mu^{out}}. \end{aligned}$$

5. Заключение

Исследована система массового обслуживания с резервированием приборов для запросов, осуществляющих повторное обслуживание. В систему поступает коррелированный входной поток первичных запросов. После обслуживания запрос может покинуть систему или отправиться на орбиту с целью повторения обслуживания через некоторое время. Запросы на орбите являются ненастойчивыми и нетерпеливыми. Поведение системы описано многомерной цепью Маркова с пространственно-неоднородными переходами. Найден инфинитезимальный генератор этой цепи, получены формулы для вычисления основных характеристик производительности системы. Результаты работы могут быть использованы для определения оптимальной политики резервирования приборов в зависимости от числа повторных запросов в системе.

Библиографические ссылки

1. Artalejo J. R., Gomez-Corral A. Retrial Queueing Systems: A Computational Approach. Springer-Verla, 2008.
2. Falin G. A survey of retrial queues // Queueing systems. 1990. Vol. 7, iss. 2. P. 127–167.
3. Templeton J. G., Falin G. I. Retrial queues. Routledge, 2023.
4. Gomez-Corral A. A bibliographical guide to the analysis of retrial queues through matrix analytic techniques // Annals of Operations Research. 2006. Vol. 141, iss. 1. P. 163–191.
5. Chakravarthy S. R. Introduction to Matrix-Analytic Methods in Queues 1: Analytical and Simulation Approach – Basics // New York : ISTE Ltd, London and John Wiley and Sons, 2022.
6. Dudin A. N., Klimenok V. I., Vishnevsky V. M. The theory of queueing systems with correlated flows // Springer Nature, 2020.

7. *Lucantoni D.* New results on the single server queue with a batch Markovian arrival process // *Communication in Statistics-Stochastic Models*. 1991. Vol. 7. P. 1–46.
8. *Graham A.* Kronecker products and matrix calculus with applications. Courier Dover Publications, 2018.
9. Mathematical model of operation of a cell of a mobile communication network with adaptive modulation schemes and handover of mobile users / C. Kim [et al.] // *IEEE Access*. 2021. Vol. 9. P. 106933–106946.
10. *Klimenok V. I., Dudin A. N.* Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory // *Queueing Systems*. 2006. Vol. 54. P. 245–259.
11. Stability of queueing systems with impatience, balking and non-persistence of customers / A. N. Dudin [et al.] // *Mathematics*. 2024. Vol. 12, iss. 14. Article no. 2214.
12. *Dudin S., Dudina O.* Retrial multi-server queueing system with *PHF* service time distribution as a model of a channel with unreliable transmission of information // *Applied Mathematical Modelling*. 2019. Vol. 65. P. 676–695.

АНАЛИЗ СИСТЕМЫ ОБСЛУЖИВАНИЯ С ПРОЦЕССОМ ПОСТУПЛЕНИЯ, ЗАВИСЯЩИМ ОТ КАЧЕСТВА ФУНКЦИОНИРОВАНИЯ

О. С. Дудина, А. Н. Дудин

*Белорусский государственный университет,
Минск, Беларусь, dudina@bsu.by, dudin@bsu.by*

Рассматривается однолинейная система массового обслуживания с коррелированным процессом поступления клиентов, зависящим от рейтинга системы. Построен процесс, описывающий функционирование системы, в виде цепи Маркова с непрерывным временем и счетной первой компонентой. Найден генератор данной цепи, а также приведен способ нахождения ее стационарного распределения. Получены формулы для нахождения основных характеристик производительности системы.

Ключевые слова: коррелированный поток; рейтинг системы; цепь Маркова; стационарное распределение; характеристики производительности.

ANALYSIS OF A QUEUEING SYSTEM WITH AN ARRIVAL PROCESS, DEPENDING ON THE QUALITY OF SERVICE

O. S. Dudina, A. N. Dudin

*Belarusian State University,
Minsk, Belarus, dudina@bsu.by, dudin@bsu.by*

A single-server queueing system with a correlated arrival process, depending on the system rating, is considered. A process describing the functioning of the system is constructed in the form of a Markov chain with continuous time and a countable first component. A generator of this chain is found, and a method for finding its stationary distribution is given. Formulas for calculating the main characteristics of the system's performance are obtained.

Keywords: correlated flow; system rating; Markov chain; stationary distribution; performance measures.

1. Введение

Теория массового обслуживания представляет собой мощный инструмент для решения задач принятия решений, обеспечивающих наилучшее использование ограниченных ресурсов для обслуживания случайных потоков запросов в телекоммуникационных, производственных,

логистических системах и сетях, а также в сфере сервиса. В классических системах обслуживания предполагалось, что потоки поступающих запросов описываются стационарными пуассоновскими потоками.

По мере развития технологий, потоки запросов, особенно в телекоммуникационных сетях, стали все больше усложняться, что вызвало необходимость пересмотреть традиционную модель пуассоновского потока как подходящую для описания поступления запросов. Реальные потоки в современных системах являются коррелированными и могут иметь высокий коэффициент вариации длин интервалов между моментами поступления запросов, в то время как стационарный пуассоновский поток характеризуется отсутствием корреляции и имеет коэффициент вариации, равный единице.

Исследования привели к разработке новой модели входного потока, получившей название марковского входного потока. Она превосходила своих предшественников за счет учета коррелированности и вариации входного потока, как показано в [1, 2]. Тем не менее и эта модель имеет свои ограничения. Главным из них является отсутствие корреляции между интенсивностью входного потока и качеством работы системы. Другими словами, данная модель исходит из предположения, что независимо от эффективности работы системы, она будет непрерывно получать одинаковый входной поток с постоянной средней интенсивностью.

В современной сфере оказания услуг существует высокая конкуренция, и очевидно, что при неэффективной работе системы пользователи предпочтут обратиться к конкурентам. Это, без сомнения, приведет к снижению количества поступающих клиентов. Следует отметить, что в современном мире любой желающий может ознакомиться с рейтингами практически всех поставщиков услуг в интернете, поэтому важно учитывать влияние изменений потока клиентов в зависимости от качества работы системы.

В данной статье исследуется система массового обслуживания с коррелированным процессом поступления клиентов, зависящим от рейтинга системы. Время обслуживания клиента имеет фазовое распределение. Клиенты, ожидающие обслуживание в буфере бесконечной емкости, могут проявлять нетерпеливость и покидать систему. Рейтинг системы изменяется с определенными вероятностями в момент поступления клиента или его ухода из буфера из-за нетерпеливости. Более высокие рейтинги подразумевают более высокие средние интенсивности поступления клиентов. Когда рейтинг имеет минимально допустимый уровень и должен быть дополнительно уменьшен, объявляется проведение промо акций, что приводит к увеличению рейтинга на фиксированное число пунктов.

2. Математическая модель

Рассмотрим систему массового обслуживания, состоящую из бесконечного буфера и одного прибора.

Клиенты поступают в зависящем от рейтинга марковском входном потоке (*RMAP*-потоке). Рейтинг системы определяется качеством ее работы и принадлежит интервалу $[1, R]$. Более низкий рейтинг соответствует худшей работе системы. Поступления в *RMAP*-потоке регулируются управляющей цепью Маркова v_t , $t \geq 0$, с конечным пространством состояний $\{1, 2, \dots, W\}$. Интенсивности переходов этого процесса зависят от текущего рейтинга. В случае рейтинга r , $r \in \{1, 2, \dots, R\}$, *RMAP*-поток определяется матрицами $D_0^{(r)}$ и $D_1^{(r)}$, которые задают стандартный *MAP*-поток. При изменении рейтинга текущее состояние *RMAP*-потока не меняется, но интенсивности переходов мгновенно изменяются. Обозначим как λ_r среднюю интенсивность прибытия в *RMAP*-потоке при рейтинге r , $r \in \{1, 2, \dots, R\}$. Эту величину можно вычислить с помощью стандартной формулы для средней входной интенсивности *MAP*-потока:

$$\lambda_r = \theta^{(r)} D_1^{(r)} e, \quad r \in \{1, 2, \dots, R\},$$

где $\theta^{(r)}$ – вектор стационарного распределения состояний *MAP*-потока, определяемый матрицами $D_0^{(r)}$ и $D_1^{(r)}$, а $e = (1, 1, \dots, 1)^T$. Здесь мы предполагаем, что интенсивности поступления перечислены в порядке возрастания. Это означает, что $\lambda_r > \lambda_{r'}$, если $r > r'$.

Распределение времени обслуживания клиентов имеет фазовый тип. Это распределение определяется цепью Маркова m_t , $t > 0$, с конечным пространством состояний $\{1, 2, \dots, M\}$ и неприводимым представлением (β, S) , где вектор-строка β определяет распределение состояния процесса m_t , $t > 0$, в момент начала обслуживания, а субгенератор S определяет интенсивности переходов процесса m_t . Более подробную информацию о распределении фазового типа можно найти в [3].

Клиенты могут быть нетерпеливыми и покидать буфер (систему) через экспоненциально распределенное время с параметром α , $\alpha > 0$.

Используется следующий механизм изменения рейтинга. Если прибывший клиент застает в системе не более K клиентов, рейтинг системы увеличивается с вероятностью p , а с дополнительной вероятностью $1 - p$ он остается прежним. Если текущий рейтинг равен R , он не может быть увеличен. Если клиент покидает систему без обслуживания из-за

нетерпеливости, рейтинг уменьшается с вероятностью q , а с дополнительной вероятностью $1 - q$ он не изменяется.

Если текущий рейтинг равен единице и должен быть уменьшен, система объявляет проведение промо акций. Это означает, что одномоментно рейтинг увеличивается на $\bar{r}$ пунктов $\bar{r} \in \{0, 1, \dots, R-1\}$. Отметим, что $\bar{r} = 0$ соответствует случаю, когда рейтинг не повышается и остается на самом низком уровне.

Исследуем описанную выше систему.

3. Процесс состояний системы и его генератор

Пусть

$i_t, i_t \geq 0$, – число клиентов в системе;

$r_t, r_t \in \{1, 2, \dots, R\}$, – рейтинг системы;

$v_t, v_t \in \{1, 2, \dots, W\}$, – состояние управляющего процесса *RMAP*-потока;

$m_t, m_t \in \{1, 2, \dots, M\}$, – состоянием процесса обслуживания в момент времени $t, t \geq 0$.

Поведение исследуемой системы описывается регулярной неприводимой цепью Маркова с непрерывным временем

$$\xi_t = \{i_t, r_t, v_t, m_t\}, \quad t \geq 0.$$

Перенумеруем состояния цепи Маркова $\xi_t, t \geq 0$, в прямом лексикографическом порядке компонент (i_t, r_t, v_t, m_t) и назовем множество состояний цепи, имеющих значение i первой компоненты цепи Маркова, уровнем $i, i \geq 0$. Множество состояний цепи, имеющих значения (i, r) первой и второй компонент цепи Маркова, назовем макросостоянием $(i, r), i \geq 0, r \in \{1, 2, \dots, R\}$.

Теорема 1. Генератор цепи Маркова $\xi_t, t \geq 0$, имеет следующую блочно-трехдиагональную структуру

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & O & O & O & \dots \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & O & O & \dots \\ O & Q_{2,1} & Q_{2,2} & Q_{2,3} & O & \dots \\ O & O & Q_{3,2} & Q_{3,3} & Q_{3,4} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

где ненулевые блоки $Q_{i,j}, |i - j| \leq 1, i, j \geq 0$, содержащие интенсивности переходов с уровня i на уровень j , определяются следующим образом:

$$\begin{aligned}
Q_{0,0} &= \text{diag}\{D_0^{(r)}, r \in \{1, 2, \dots, R\}\}, \\
Q_{i,i} &= \text{diag}\{D_0^{(r)} \otimes I_M, r \in \{1, 2, \dots, R\}\} + I_{RW} \otimes S - i\alpha I_{RWM}, i \geq 1, \\
Q_{0,1} &= (1-p)\text{diag}\{D_1^{(r)} \otimes \beta, r \in \{1, 2, \dots, R\}\} + p\Phi, \\
Q_{i,i+1} &= (1-p)\text{diag}\{D_1^{(r)} \otimes I_M, r \in \{1, 2, \dots, R\}\} + p\Psi, i = \overline{1, K}, \\
Q_{i,i+1} &= \text{diag}\{D_1^{(r)} \otimes I_M, r \in \{1, 2, \dots, R\}\}, i > K, \\
Q_{1,0} &= I_{RW} \otimes S_0, \\
Q_{i,i-1} &= I_{RW} \otimes S_0\beta + (1-p)\alpha(i-1)I_{RWM} + q\alpha(i-1)\Lambda \otimes I_{WM}, i > 1.
\end{aligned}$$

Здесь

$\otimes$ – символ Кронекерова произведения матриц; см., например, [4];

I_n – единичная матрица, а O_n – нулевая матрица, размерность которой указывается нижним индексом n при необходимости;

$\text{diag}\{c_1, c_2, \dots, c_n\}$ – диагональная матрица с диагональными элементами $c_1, c_2, \dots, c_n$;

$$\begin{aligned}
\Phi &= \begin{pmatrix} O_{W \times WM} & D_1^{(1)} \otimes \beta & O & \dots & O & O \\ O & O & D_1^{(2)} \otimes \beta & \dots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \dots & O & D_1^{(R-1)} \otimes \beta \\ O & O & O & \dots & O & D_1^{(R)} \otimes \beta \end{pmatrix}, \\
\Psi &= \begin{pmatrix} O_{WM} & D_1^{(1)} \otimes I_M & O & \dots & O & O \\ O & O & D_1^{(2)} \otimes I_M & \dots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \dots & O & D_1^{(R-1)} \otimes I_M \\ O & O & O & \dots & O & D_1^{(R)} \otimes I_M \end{pmatrix},
\end{aligned}$$

Λ – квадратная матрица размера R со всеми нулевыми элементами, за исключением элементов $(\Lambda)_{i,i-1}$, $i \in \{1, 2, \dots, R\}$, и $(\Lambda)_{0,\bar{r}}$, которые равны 1.

Теорема доказывается путем анализа интенсивностей всех возможных переходов цепи Маркова ξ_t , $t \geq 0$, за бесконечно малый интервал времени.

Легко видеть, что в случае терпеливых клиентов ($\alpha = 0$) блоки генератора не зависят от компоненты i , и цепь Маркова ξ_t , $t \geq 0$, принадлежит классу квазипроцессов гибели и размножения. Поскольку в данной модели

предполагается, что клиенты являются нетерпеливыми ($\alpha > 0$), то согласно теореме 1 из [5], цепь Маркова ξ_t , $t \geq 0$, является эргодичной при любых параметрах системы. Следовательно, существует стационарное распределение данной цепи

$$\pi(i, r, v, m) = \lim_{t \rightarrow \infty} P\{i_t = i, r_t = r, v_t = v, m_t = m\},$$

$$i_t \geq 0, r \in \{1, 2, \dots, R\}, v \in \{1, 2, \dots, W\}, m \in \{1, 2, \dots, M\}.$$

Перенумеруем данные вероятности в прямом лексикографическом порядке и сформируем векторы

$$\pi(i, r, v) = (\pi(i, r, v, 1), \pi(i, r, v, 2), \dots, \pi(i, r, v, M)),$$

$$\pi(i, r) = (\pi(i, r, 1), \pi(i, r, 2), \dots, \pi(i, r, W)),$$

$$\pi_i = \pi(i) = (\pi(i, 1), \pi(i, 2), \dots, \pi(i, R)).$$

Хорошо известно, что векторы π_i , $i \geq 0$, удовлетворяют следующей системе линейных алгебраических уравнений:

$$(\pi_0, \pi_1, \dots, \pi_i, \dots)Q = 0, (\pi_0, \pi_1, \dots, \pi_i, \dots)e = 1, \quad (1)$$

где 0 – вектор-строка, состоящая из нулей.

Система (1) имеет бесконечное число уравнений и не может быть решена стандартными методами. Для ее решения может быть использован эффективный алгоритм, представленный в работе [6]. Найдя стационарное распределение системы, можно определить ее основные характеристики производительности.

4. Характеристики производительности

Среднее количество клиентов в системе рассчитывается по формуле

$$N_{sys} = \sum_{i=1}^{\infty} i \pi_i e.$$

Среднее количество клиентов в буфере рассчитывается по формуле

$$N_{buf} = \sum_{i=2}^{\infty} (i-1) \pi_i e = N_{sys} - N_{serv}.$$

Средний рейтинг системы рассчитывается по формуле

$$\bar{R} = \sum_{i=0}^{\infty} \sum_{r=1}^R r \pi(i, r) \mathbf{e}.$$

Средняя интенсивность поступления клиентов рассчитывается как

$$\lambda_{in} = \sum_{i=1}^{\infty} \sum_{r=1}^R \pi(i, r) (D_1^{(r)} \otimes I_M) \mathbf{e} + \sum_{r=1}^R \pi(0, r) D_1^{(r)} \mathbf{e}.$$

Средняя интенсивность выходного потока успешно обслуженных клиентов рассчитывается по формуле

$$\lambda_{out} = \sum_{i=1}^{\infty} \pi_i (I_{RW} \otimes S_0) \mathbf{e}.$$

Вероятность немедленного доступа рассчитывается как

$$P_{imm} = \frac{1}{\lambda_{in}} \sum_{r=1}^R \pi(0, r) D_1^{(r)} \mathbf{e}.$$

Средняя интенсивность акций рассчитывается как

$$\gamma_{prom} = q\alpha \sum_{i=2}^{\infty} (i-1) \pi(i, 1) \mathbf{e}.$$

Средняя интенсивность повышения рейтинга без рекламы рассчитывается как

$$\gamma^+ = p \sum_{i=1}^{i_1} \sum_{r=1}^{R-1} \pi(i, r) (D_1^{(r)} \otimes I_M) \mathbf{e} + \sum_{r=1}^{R-1} \pi(0, r) D_1^{(r)} \mathbf{e}.$$

Средняя интенсивность снижения рейтинга без рекламы рассчитывается как

$$\gamma^- = q\alpha \sum_{i=2}^{\infty} (i-1) \pi_i \mathbf{e}.$$

Вероятность потери произвольного клиента рассчитывается по формуле

$$P_{loss} = \frac{\alpha \sum_{i=2}^{\infty} (i-1) \pi_i e}{\lambda_{in}} = 1 - \frac{\lambda_{out}}{\lambda_{in}}.$$

5. Заключение

В статье исследована система массового обслуживания с коррелированным входным потоком, зависящим от качества функционирования. Клиенты поступают в *RMAP* потоке. Время обслуживания клиента имеет фазовое распределение. Клиенты, ожидающие обслуживание в буфере, могут проявлять нетерпеливость и покидать систему. Рейтинг системы изменяется с определенными вероятностями в момент поступления клиента или его ухода из буфера из-за нетерпеливости. Построена цепь Маркова, которая описывает процесс функционирования системы. Приведен ее генератор и найдено стационарное распределение. Получены формулы для вычисления важных характеристик производительности системы.

Библиографические ссылки

1. *Lucantoni D.* New results on the single server queue with a batch Markovian arrival process // *Communication in Statistics-Stochastic Models*. 1991. Vol. 7. P. 1–46.
2. *Dudin A. N., Klimenok V. I., Vishnevsky V. M.* The theory of queueing systems with correlated flows // Springer Nature, 2020.
3. *Neuts M. F.* Matrix-analytic methods in queueing theory // *European Journal of Operational Research*. 1984. Vol. 15, iss. 1. P. 2–12.
4. *Graham A.* Kronecker products and matrix calculus with applications. Courier Dover Publications, 2018.
5. Stability of queueing systems with impatience, balking and non-persistence of customers / A. N. Dudin [et al.] // *Mathematics*. 2024. Vol. 12, iss. 14. Article no. 2214.
6. *Dudin S., Dudina O.* Retrial multi-server queueing system with *PHF* service time distribution as a model of a channel with unreliable transmission of information // *Applied Mathematical Modelling*. 2019. Vol. 65. P. 676–695.

**АСИМПТОТИКА РЕШЕНИЯ ЗАДАЧИ МИНИМИЗАЦИИ
ЭНЕРГЕТИЧЕСКИХ ЗАТРАТ ПРИ УПРАВЛЕНИИ
ТРЕХТЕМПОВОЙ СИНГУЛЯРНО ВОЗМУЩЕННОЙ
ЛИНЕЙНОЙ СИСТЕМОЙ**

А. И. Калинин¹⁾, Л. И. Лавринович²⁾

¹⁾ *Белорусский государственный университет,
Минск, Беларусь, kalininai@bsu.by*

²⁾ *Белорусский государственный университет,
Минск, Беларусь, lavrinovich@bsu.by*

Рассматривается задача минимизации энергетических затрат для линейной сингулярно возмущенной системы, содержащей три группы переменных с существенно различными скоростями изменения. Строятся асимптотические приближения к решению этой задачи в виде программы и обратной связи. Основное достоинство предлагаемых вычислительных процедур состоит в том, что при их применении исходная задача распадается на три невозмущенные задачи оптимального управления меньшей размерности.

Ключевые слова: малый параметр; сингулярно возмущенная система; квадратичный функционал; оптимальное управление; обратная связь; асимптотические приближения.

**ASYMPTOTICS OF THE SOLUTION OF THE PROBLEM
OF MINIMIZING ENERGY COSTS IN CONTROLLING
A THREE-TEMPO SINGULARLY PERTURBED LINEAR SYSTEM**

A. I. Kalinin^{a)}, L. I. Lavrinovich^{b)}

^{a)} *Belarusian State University,
Minsk, Belarus, kalininai@bsu.by*

^{b)} *Belarusian State University,
Minsk, Belarus, lavrinovich@bsu.by*

The problem of minimizing energy costs for a linear singularly perturbed system containing three groups of variables with significantly different rates of change is considered. Asymptotic approximations to the solution of this problem in the form of a program and feedback are constructed. The main advantage of the proposed computational procedures is that when they are applied, the original problem is divided into three unperturbed optimal control problems of lower dimension.

Keywords: small parameter; singularly perturbed system; quadratic functional; optimal control; feedback; asymptotic approximations.

1. Введение

В математической теории оптимальных процессов значительное внимание уделяется задачам оптимизации сингулярно возмущенных систем, содержащих малые параметры при части производных. Это вызвано эффективностью асимптотических методов решения таких задач, при применении которых исходная задача распадается на задачи оптимального управления меньшей размерности. Кроме того, асимптотический подход позволяет избежать интегрирования сингулярно возмущенных систем, которые являются жесткими.

В первых работах, посвященных задачам оптимального управления с сингулярными возмущениями, рассматривались динамические системы, содержащие только две группы переменных с существенно различными скоростями изменения, которые были названы медленными и быстрыми переменными (см. обзоры [1] – [4]). В дальнейшем появились работы по оптимизации систем с несколькими группами быстрых переменных, которые содержали при производных параметры различных порядков малости. Их обзор содержится в [5]. Представленный доклад посвящен построению асимптотических приближений в виде программы и обратной связи к решению задачи об управлении с минимальными энергетическими затратами сингулярно возмущенной системой с двумя группами быстрых переменных.

2. Постановка задачи

В классе r -мерных управляющих воздействий $u(t) = (u_1(t), \dots, u_r(t))$, $t \in T = [t_*, t^*]$, с кусочно-непрерывными компонентами рассмотрим следующую задачу оптимального управления:

$$\begin{aligned} \dot{x} &= A_{11}x + A_{12}y + A_{13}z + B_1u, \quad x(t_*) = x_*, \\ \mu \dot{y} &= A_{21}x + A_{22}y + A_{23}z + B_2u, \quad y(t_*) = y_*, \end{aligned} \quad (1)$$

$$\begin{aligned} \mu^2 \dot{z} &= A_{31}x + A_{32}y + A_{33}z + B_3u, \quad z(t_*) = z_*, \\ x(t^*) &= 0, \quad y(t^*) = 0, \quad z(t^*) = 0, \end{aligned} \quad (2)$$

$$J(u) = \frac{1}{2} \int_{t_*}^{t^*} u^T P u dt \rightarrow \min, \quad (3)$$

где μ – малый положительный параметр, t_* , t^* – заданные моменты времени ($t_* < t^*$), x – n_1 -вектор медленных переменных, y, z – векторы быстрых переменных размерности n_2 и n_3 соответственно. Остальные элементы

задачи имеют соответствующие размеры. В критерии качества P – положительно-определенная симметрическая матрица.

Предположение 1. Матрицы A_{33} , $C = A_{22} - A_{23}A_{33}^{-1}A_{32}$ устойчивые, т.е. действительные части всех их собственных значений отрицательны.

Определение 1. Управление $u^{(N)}(t, \mu)$, $t \in T$, с кусочно-непрерывными компонентами назовем (программным) асимптотически субоптимальным управлением N -го порядка ($N = 0, 1, 2, \dots$), если оно переводит динамическую систему (1) в состояние $O(\mu^{N+1})$ и отклоняется по критерию качества (3) от оптимального управления на величину того же порядка малости.

Определение 2. Вектор-функцию $u^{(N)}(x, y, z, t, \mu)$ назовем асимптотически субоптимальной обратной связью N -го порядка, если для любого начального состояния (x_*, y_*, z_*, t_*) , $t_* < t^*$, имеет место $u^{(N)}(x_*, y_*, z_*, t_*, \mu) = u^{(N)}(t_*, \mu)$, где $u^{(N)}(t, \mu)$, $t \in T$, – асимптотически субоптимальное управление N -го порядка в задаче (1) – (3).

3. Базовые задачи

Основная идея применяемого подхода состоит в асимптотическом разложении по целым степеням малого параметра начальных значений (в момент времени t^*) сопряженных переменных – конечномерных элементов, по которым можно легко восстановить решение задачи. Старшие коэффициенты этих разложений могут быть найдены в результате решения трех базовых невозмущенных задач оптимального управления с n_1 , n_2 , n_3 фазовыми переменными соответственно. Первой из них является вырожденная задача

$$\dot{x} = A_0 x + B_0 u, \quad x(t_*) = x_*, \quad x(t^*) = 0,$$

$$J_1(u) = \frac{1}{2} \int_{t_*}^{t^*} u^T P u dt \rightarrow \min,$$

где

$$A_0 = A_{11} - A_{13}A_{33}^{-1}A_{31} - (A_{21} - A_{13}A_{33}^{-1}A_{32})C^{-1}(A_{21} - A_{23}A_{33}^{-1}A_{31}),$$

$$B_0 = B_1 - A_{13}A_{33}^{-1}B_3 - (A_{12} - A_{13}A_{33}^{-1}A_{32})C^{-1}D,$$

$$D = B_2 - A_{23}A_{33}^{-1}B_3.$$

В дальнейшем эту задачу будем называть первой базовой.

Во второй базовой задаче

$$\dot{y} = Cy + Du, y(0) = C^{-1}Du^0(t^*),$$

$$y(-\infty) = 0, J_2(u) = \frac{1}{2} \int_{-\infty}^0 u^T P u ds \rightarrow \min,$$

$u^0(t)$, $t \in T$, – решение первой базовой задачи.

Третья базовая задача имеет вид

$$\dot{z} = A_{33}z + B_3u, z(0) = A_{33}^{-1}B_3(u^0(t^*) + u^*(0)),$$

$$z(-\infty) = 0, J_3(u) = \frac{1}{2} \int_{-\infty}^0 u^T P u d\tau \rightarrow \min,$$

где $u^*(s)$, $s \leq 0$, – решение второй базовой задачи. Обозначим через $u_*(\tau)$, $\tau \leq 0$, оптимальное управление в третьей базовой задаче.

Предположение 2. Динамические системы в базовых задачах являются вполне управляемыми.

Сделанные предположения гарантируют существование и единственность решений базовых задач, которые являются нормальными экстремалиями.

Теорема. При выполнении предположений 1–2 в задаче (1) – (3) с достаточно малым μ существует единственное оптимальное управление, представимое в виде

$$u^0(t, \mu) = P^{-1} (B_1^T \psi_1^0(t, \mu) + B_2^T \psi_2^0(t, \mu) + B_3^T \psi_3^0(t, \mu)), t \in T.$$

Начальные значения $\lambda(\mu) = \psi_1^0(t^*, \mu)$, $v(\mu) = \psi_2^0(t^*, \mu)$, $\eta(\mu) = \psi_3^0(t^*, \mu)$ вектора сопряженных переменных $(\psi_1^0(t, \mu), \psi_2^0(t, \mu), \psi_3^0(t, \mu))$, $t \in T$, который в силу принципа максимума [6] соответствует оптимальному управлению, допускают асимптотические разложения

$$\lambda(\mu) \sim \lambda_0 + \sum_{k=1}^{\infty} \mu^k \lambda_k, v(\mu) \sim v_0 + \sum_{k=1}^{\infty} \mu^k v_k, \eta(\mu) \sim \eta_0 + \sum_{k=1}^{\infty} \mu^k \eta_k,$$

в которых

$$v_0 = \sigma_0 - \left((A_{12} - A_{13}A_{33}^{-1}A_{32})C^{-1} \right)^T \lambda_0, \quad \eta_0 = \chi_0 - (A_{13}A_{33}^{-1})^T \lambda_0 + (A_{23}A_{33}^{-1})^T v_0,$$

а λ_0 , σ_0 , χ_0 – начальные значения сопряженных переменных соответственно в первой, второй и третьей базовых задачах.

Асимптотически субоптимальное управление нулевого порядка представимо в виде

$$u^{(0)}(t, \mu) = u^0(t) + u^* \left((t - t^*) / \mu \right) + u_* \left((t - t^*) / \mu^2 \right), \quad t \in T,$$

и может быть сформировано непосредственно после решения базовых задач. Для построения асимптотически субоптимальных управлений более высокого порядка нужно дополнительно интегрировать невозмущенные системы линейных дифференциальных уравнений и находить решения невырожденных линейных алгебраических систем.

4. Заключение

В докладе предложены вычислительные процедуры построения асимптотических приближений к решению рассмотренной задачи в виде программы и обратной связи. При реализации предлагаемых алгоритмов исходная задача оптимального управления распадается на три невозмущенные задачи меньшей размерности. Такая декомпозиция позволяет эффективно решать задачи оптимизации динамических систем с большим числом фазовых переменных. Кроме того, вычислительные процедуры алгоритмов не содержат интегрирований жестких систем.

Представленная в докладе методика решения задачи является естественным развитием подхода, изложенного в [7], [8]. При этом предложенная методика позволяет исследовать задачи, в которых имеется несколько групп быстрых переменных с иерархией скоростей по целым степеням малого параметра. При этом количество задач, на которые распадается исходная задача, равно количеству групп разнотемповых переменных. Такое обобщение вносит в алгоритм не принципиальные изменения, которые легко прослеживаются на примере рассмотренной задачи. В то же время это приводит к громоздким формулам.

Отметим также, что развитие полученных результатов на нестационарные системы с достаточно гладкими коэффициентами не вызывает принципиальных трудностей.

Библиографические ссылки

1. *Дмитриев М. Г., Курина Г. А.* Сингулярные возмущения в задачах управления // Автоматика и телемеханика. 2006. № 1. С. 3–51.
2. *Калинин А. И.* Асимптотика решений возмущенных задач оптимального управления // Изв. РАН. Техн. кибернетика. 1994. № 3. С. 104–114.
3. Singular perturbation and time scales in control theories and applications. An overview 2002 – 2012 / Y. Zhang [et al.] // Int. J. Information and Systems Sciences. 2014. Vol. 9, no. 1, P. 1–36.
4. *Kokotovic P. V., Khalil H. K.* Singular perturbations in systems and control. New York : IEEE Press, 1986.
5. *Курина Г. А., Калашникова М. А.* Сингулярно возмущенные задачи с разнотемповыми быстрыми переменными // Автоматика и телемеханика. 2022. № 11. С. 3–61.
6. Математическая теория оптимальных процессов / Л. С. Понтрягин [и др.]. М. : Наука, 1983.
7. *Калинин А. И., Лавринович Л. И.* Асимптотика решения сингулярно возмущенной линейно-квадратичной задачи оптимального управления // Журнал вычислительной математики и математической физики. 2015. Т. 55, № 2. С. 194–206.
8. *Калинин А. И., Лавринович Л. И.* Применение метода малого параметра к сингулярно возмущенной линейно-квадратичной задаче оптимального управления // Автоматика и телемеханика. 2016. № 5. С. 3–18.

ON FACIAL STRUCTURE OF THE COPOSITIVE AND COMPLETELY POSITIVE CONES

O. I. Kostyukova^{a)}, T. V. Tchemisova^{b)}

^{a)} *Institute of Mathematics, National Academy of Sciences of Belarus,
Minsk, Belarus, kostyukova@im.bas-net.by*

^{b)} *Mathematical Department, University of Aveiro, Campus Universitario Santiago,
3810-193, Aveiro, Portugal, tatiana@ua.pt*

This paper investigates the cones of copositive and completely positive matrices, mathematical structures that have garnered significant attention due to their diverse applications. Despite their importance, the intricate geometric properties of these cones, particularly their facial structures, remain only partially understood. A key aspect of studying cones, and matrix cones in particular, lies in analyzing their boundary structure, defined by their faces. This work aims to examine new properties of faces of these cones.

Keywords: conic optimization; copositive and completely positive matrices; singularity degree.

1. Introduction

Linear conic optimization, which involves optimizing linear functions over feasible sets defined by the intersection of an affine space and a convex cone, has emerged as a significant subfield of convex optimization. The conical structure of feasible sets facilitates the modeling of numerous practical nonlinearities, broadening the scope of applications in this domain. With the rapid growth of such applications and the advancement of supporting computational tools (see, e.g., [1, 2]), conic optimization continues to gain prominence in both theoretical and applied contexts.

As the feasible set of a linear conic program is defined by the intersection of an affine space and a convex cone, this cone plays the pivotal role. While the objective function and the constraints functions are linear, the geometry and structural complexity of the cone under consideration often present the main challenges. Understanding the construction of the cone, particularly through its local description, can significantly enhance the efficiency of solving conic optimization problems.

A fundamental aspect of studying convex cones lies in their boundary structure, which is determined by their faces. The fact that these faces either possess or lack the facial exposedness, offers valuable insights into the cone

properties. This makes the study of cone faces and their features a central focus in conic optimization research.

This paper focuses on a pair of dual cones — the cone of copositive and completely positive matrices, which have been extensively investigated due to their wide range of applications (see [3–8]). However, despite their importance, the complex geometric structures of these cones, particularly their facial structures, remain only partially understood.

In this study, we aim to deepen the understanding of the geometric properties of the copositive and completely positive cones and contribute to the broader exploration of conic optimization.

2. Definitions and notation

In this section, we define terms and notation to be used henceforth.

2.1. General cones and their faces

Given a finite dimensional space $\mathfrak{X}$, let us remind some general definitions.

A set $C \subseteq \mathfrak{X}$ is convex if for any $x, y \in C$ and any $\alpha \in [0, 1]$, it holds $\alpha x + (1 - \alpha)y \in C$. Given a set $\mathcal{B} \subset \mathfrak{X}$, denote by $\text{conv}\{\mathcal{B}\}$ its *convex hull*, i.e. the minimal (by inclusion) convex set, containing $\mathcal{B}$, by $\text{cone}\{\mathcal{B}\}$ its *conic hull*, which is the smallest cone containing $\mathcal{B}$, and by $\text{ri}\{\mathcal{B}\}$ its relative interior.

A set $K \subseteq \mathfrak{X}$ is called a *cone* if for any $x \in K$ and any $\alpha > 0$, it holds $\alpha x \in K$. A cone $K \subseteq \mathfrak{X}$ is called *polyhedral* if there exists a finite set $\{x(i), i \in I\} \subset \mathfrak{X}$ such that $K = \text{cone}\{x(i), i \in I\}$. A nonempty convex subset F of a convex closed cone $K \subset \mathfrak{X}$ is its *face* if an inclusion $\alpha x + (1 - \alpha)y \in F$ with $x, y \in K$ and $\alpha \in (0, 1)$ implies $x, y \in F$. A face F of a cone K is *proper* if $F \neq \emptyset$ and $F \neq K$.

A face F of a closed convex cone $K \subset \mathfrak{X}$ is called *exposed* if it can be presented as the intersection of K with a supporting hyperplane, i.e., there exist $y \in \mathfrak{X}$ and $d \in \mathbb{R}$ such that for all $x \in K$, it holds $\langle y, x \rangle \geq d$ and $\langle y, x \rangle = d$ iff $x \in F$. Every exposed face is, evidently, a face. A cone K is called *facially exposed* if all its faces are exposed.

Given a cone $K \subset \mathfrak{X}$, its dual cone K^* is defined as

$$K^* = \{x \in \mathfrak{X} : \langle x, y \rangle \geq 0 \ \forall y \in K\}.$$

Let F_* be a proper face of a cone K . Applying Algorithm 2 from [9], one can conclude that there exist an integer $k_0 > 0$ and a set of cones $\{F_k, k = 0, 1, \dots, k_0\}$ such that

$$F_0 := K, \\ F_k := F_{k-1} \cap U_{k-1}^\perp \text{ with some } U_{k-1} \in (F_{k-1})^*, \quad k = 1, \dots, k_0; \text{ and } F_{k_0} = F_*. \quad (1)$$

It follows from these relations that for $k = 1, \dots, k_0$, the cone F_k is an exposed face of the cone F_{k-1} and the face F_* of K admits a representation

$$F_* = K \cap U_0^\perp \cap U_1^\perp \cap \dots \cap U_{k_0}^\perp.$$

The following definition is provided in [9].

Definition. For a given proper face F of a cone K , the *singularity degree* of F over K is the minimal number k_0 such that there exists a set of cones $\{F_k, k = 0, \dots, k_0\}$ satisfying (1).

Let us define the *singularity degree of a cone* as the maximal singularity degree among all its faces. It is evident that the singularity degree of a cone characterizes the complexity of the cone.

It is evident that the singularity degree of a face of a cone characterizes the complexity of both, the structure of this face and the cone as a whole.

Given a facially exposed cone, it is easy to verify that all its proper faces have singularity degree equal to one. It is known that second order- and p -cones, polyhedral and semidefinite cones are facially exposed (see, for instance, [10]).

At the same time, there are cones that are not facially exposed. Non-facially exposed cones may have faces with a singularity degree greater than or equal to two. This paper aims to explore facial structure and the singularity degrees of the faces of the cones of copositive and completely positive matrices.

2.2. The cones of copositive and completely positive matrices

Given an integer $p > 1$, denote by $\mathbb{R}_+^p$ the set of all p -vectors with non-negative components (the non-negative orthant in $\mathbb{R}^p$) and by e_k , $k = 1, \dots, p$, the standard basis in $\mathbb{R}^p$. For a vector $t = (t_k, k = 1, \dots, p)^\top \in \mathbb{R}^p$, denote its norm by $\|t\|_1 := |t_1| + \dots + |t_p|$ and its support by $\text{supp}(t) := \{k \in \{1, \dots, p\} : t_k \neq 0\}$.

Let $\mathbb{S}(p)$ be the space of real symmetric $p \times p$ matrices. The inner product of two matrices, A and U , in the matrix space $\mathbb{S}(p)$ is defined as follows: $A \bullet U := \text{trace}(AU)$.

For a given closed subset $K \subset \mathbb{R}^p$ denote by $COP(K)$ the cone of symmetric K -semidefinite matrices [11]

$$COP(K) := \{D \in \mathbb{S}(p) : t^\top D t \geq 0 \ \forall t \in K\}.$$

It is evident that $COP(K) = COP(K_{norm})$ where $K_{norm} = \{t \in K : \|t\|_1 = 1\}$.

For $K = \mathbb{R}_+^p$, we obtain the cone of copositive matrices [5]

$$COP(\mathbb{R}_+^p) := \{D \in \mathbb{S}(p) : t^\top D t \geq 0 \ \forall t \in \mathbb{R}_+^p\} = COP(T),$$

where

$$T := \{t \in \mathbb{R}_+^p : \|t\|_1 = 1\}$$

is the simplex in $\mathbb{R}^p$.

In what follows, we will denote the copositive cone $COP(T)$ as $\mathbb{COP}(p)$.

For the cone $\mathbb{COP}(p)$, the dual cone is the cone of completely positive matrices

$$\mathbb{CP}(p) := (\mathbb{COP}(p))^* = \text{cone}\{tt^\top, t \in T\}.$$

Recall that a completely positive matrix $U \in \mathbb{CP}(p)$ can be written as $U = \sum_{i \in I} t(i)t(i)^\top$ with an index set I such that $|I| \leq p(p+1)/2$ and some vectors $t(i) \in \mathbb{R}_+^p$, $i \in I$.

Suppose that a matrix $X \in \mathbb{COP}(p)$ is given. Denote by $T_0(X)$ the set of all *normalized zeros* of X :

$$T_0(X) := \{t \in T : t^\top X t = 0\}.$$

A zero $\tau \in T_0(X)$ of the matrix X is called *minimal* if there does not exist a zero $t \in T_0(X)$ such that the strict inclusion $\text{supp}(t) \subset \text{supp}(\tau)$ holds. It is a known fact that for a copositive matrix, there is a finite number of normalized minimal zeros.

3. On the properties of faces of the cone $\mathbb{COP}(p)$

In this section, we will establish some important properties of faces of the cone $\mathbb{COP}(p)$. It is known that the cone $\mathbb{COP}(p)$ is not facially exposed, but it was not known what is a singularity degree of this cone. In this section to prove

that any proper face of $\mathbb{COP}(p)$ has the singularity degree less or equal to two, and hence the singularity degree of $\mathbb{COP}(p)$ is equal to two.

Let F_* be a proper face of the cone $\mathbb{COP}(p)$. Consider a fixed matrix $X_* \in ri(F_*)$. Denote by $T_0(X_*)$ the set of all normalized zeros of this matrix:

$$T_0(X_*) := \{t \in T : t^\top X_* t = 0\}.$$

Applying the algorithms from [12], one can find for matrix X_* the corresponding set $\{\tau(j), j \in J\}$ of all normalized minimal zeros. One can also determine the set $\{J(s), s \in S\}$ of subsets of J such that the set $T_0(X_*)$ of all normalized zeros of X_* admits the representation

$$T_0(X_*) = \bigcup_{s \in S} T_*(s), \text{ where } T_*(s) := \text{conv}\{\tau(j), j \in J(s)\} \forall s \in S. \quad (2)$$

Consider the following sets:

$$M(j) := \{k \in \{1, \dots, p\} : e_k^\top X_* \tau(j) = 0\} \forall j \in J. \quad (3)$$

It is easy to show that

$$\text{supp}(\tau(j)) \subseteq M(j) \forall j \in J. \quad (4)$$

Using the set $\{\tau(j), j \in J\}$ of minimal normalized zeros of X_* and the sets $M(j)$, $j \in J$, defined in (3), let us construct the corresponding set of matrices

$$F_* := \{X \in \mathbb{COP}(p) : e_k^\top X \tau(j) = 0 \forall k \in M(j), \forall j \in J\}. \quad (5)$$

Using representation (2) of the set of all normalized zeros of X_* and inclusions (3), one can prove the following proposition.

Proposition 1. Let F_* be a face of $\mathbb{COP}(p)$, $X_* \in ri(F_*)$ and F^* be the corresponding set defined in (5). Then $F_* = F^*$.

Based on this proposition we prove the following theorem which is the main result of this section.

Theorem 1. *Let F_* be a proper face of the cone $\mathbb{COP}(p)$. Then the singularity degree of F_* over $\mathbb{COP}(p)$ is not greater than two.*

To prove this theorem we show that, for $k_0 = 2$, faces F_k , $k = 0, \dots, k_0$, in (1) can be constructed as follows.

1. Set $F_0 := \mathbb{COP}(p)$ and $U_0 := \sum_{j \in J} \tau(j)(\tau(j))^\top$. By construction, $U_0 \in \mathbb{COP}(p) = F_0^*$.

2. Consider an exposed face F_1 of $\mathbb{COP}(p)$ defined as

$$F_1 := F_0 \cap U_0^\perp = \{X \in \mathbb{COP}(p) : (\tau(j))^\top X \tau(j) = 0 \ \forall j \in J\}$$

and a matrix $U_1 := \sum_{j \in J} \sum_{k \in M(j)} (e_k(\tau(j))^\top + \tau(j)e_k^\top) \in \mathbb{S}(p)$. One can show that

$$U_1 \in F_1^*.$$

3. Then we consider an exposed face F_2 of the cone F_1 : $F_2 := F_1 \cap U_1^\perp$. By construction,

$$F_2 = \{X \in \mathbb{COP}(p) : (\tau(j))^\top X \tau(j) = 0 \ \forall j \in J; \ X \bullet U_1 = 0\}.$$

Using Proposition 1, one can prove that $F_2 = F_*$.

This implies that singularity degree of any face of $\mathbb{COP}(p)$ cannot be greater than two.

Note that it is known that the cone $\mathbb{COP}(p)$ is not facially exposed, and consequently there exist faces with singularity degree greater than one. Hence, it follows from Theorem 1 that singularity degree of the cone $\mathbb{COP}(p)$ is equal to two.

These results provide new insights into the facial structure of the cone of copositive matrices, offering valuable information for further investigations in this area. Moreover, the approach developed in this work can be extended to the study of other matrix cones, broadening its applicability and significance.

4. On the properties of faces of the cone $\mathbb{CP}(p)$

In this section, we consider the cone of completely positive matrices.

Some open problems, which are currently of interest in the theory of copositive and completely positive matrices and highly useful for optimization, were outlined by Berman et al. in [13]. One of the stated open problems was: Is the $p \times p$ completely positive cone facially exposed?

In paper [14], the author answers this question providing a parametric family of non-exposed faces of the 5×5 completely positive cone. At the end of this paper the author asked “Whether there is another type of non-exposed faces in the 5×5 completely positive cone, or specifically in the 5×5 completely positive cones, will be an interesting research topic.”

In our study, we explore non-exposed polyhedral and non-polyhedral faces of the cone of completely positive matrices. We prove a criterion for a face of this cone to be non-exposed. We also provide sufficient conditions that can be easily checked numerically. We show that for any $p \geq 6$, there exist non-exposed (polyhedral and non-polyhedral) faces of the cone of $p \times p$ completely positive matrices.

Let F be a proper face of $\mathbb{CP}(p)$. Then applying Algorithm 2 from [9], we conclude that there exists a finite integer $k_0 \geq 0$ such that $F = F_{k_0+1}$, where

$$F_0 = \mathbb{CP}(p), \quad X_k \in F_k^*, \quad X_k \notin F_k^\perp, \quad F_{k+1} = F_k \cap X_k^\perp, \quad k = 0, \dots, k_0.$$

It follows from these relations that

$$X_0 \in \mathbb{COP}(p), \quad Z_0(0) := T_0(X_0) \neq \emptyset, \quad Z_0(0) \neq T, \quad (6)$$

$$X_{k+1} \in \mathbb{COP}(Z_0(k)), \quad Z_0(k+1) := \{t \in Z_0(k) : t^\top X_{k+1} t = 0\} \neq \emptyset, \quad Z_0(k+1) \neq Z_0(k),$$

$$k = 0, \dots, k_0 - 1;$$

$$F = \{U = \sum_{i=1}^{p(p+1)/2} \alpha_i t(i)(t(i))^\top : \alpha_i \geq 0, t(i) \in Z_0(k_0) \forall i = 1, \dots, p(p+1)/2\}. \quad (7)$$

On the other hand, let $\{X_k, k = 0, \dots, k_0\}$ be a set of matrices and $\{Z_0(k), k = 0, \dots, k_0\}$ be a set of subsets of T satisfying (6), then the set F defined by the rule (7) is a proper face of $\mathbb{CP}(p)$. This gives us a way for constructing proper faces of $\mathbb{CP}(p)$.

It is evident that if $k_0 = 0$, then the face (7) is an exposed face of $\mathbb{CP}(p)$.

Theorem 2. Let $k_0 \geq 1$ and $\{X_k, k = 0, \dots, k_0\}$ be a set of matrices satisfying (7), then the set F defined by the rule (7) is a non-exposed face of $\mathbb{CP}(p)$ iff there exists $t^* \in T_0(X_0) \setminus Z_0(k_0)$ such that for any $D \in \mathbb{COP}(p)$.

$$\text{the equalities } t^\top D t = 0 \forall t \in Z_0(k_0) \text{ imply the equality } t^{*\top} D t^* = 0. \quad (8)$$

In Theorem 2, for a given $t^* \in T$, it is not easy to test if for any $D \in \mathbb{CP}(p)$, condition (8) holds true. Let us give easily tested sufficient conditions for fulfillment of this condition.

For a given set $\{X_k, k = 0, \dots, k_0\}$ of matrices satisfying (6), consider the set F defined by the rule (7). It was shown above that the set F is a proper face of $\mathbb{CP}(p)$. One can show that the set $Z_0(k_0)$ is a union of a finite number of polytopes.

Knowing the set of matrices $\{X_k, k = 0, \dots, k_0\}$ and consistently using the algorithms from [12], we can find the sets $J(s)$, $s \in S$, and vectors $\xi(j) \in T$, $j \in J(s)$, $s \in S$, such that the set $Z_0(k_0)$ admits the representation

$$Z_0(k_0) = \bigcup_{s \in S} Z(s), \quad Z(s) = \text{conv}\{\xi(j), j \in J(s)\}. \quad (9)$$

Denote $J := \bigcup_{s \in S} J(s)$, $S(j) = \{s \in S : j \in J(s)\}$ for all $j \in J$,

$$P_*(s) := \bigcup_{j \in J(s)} \text{supp}(\xi(j)) \quad \forall s \in S; \quad M_*(j) := \bigcup_{s \in S(j)} P_*(s) \quad \forall j \in J.$$

Proposition 2. Consider a set of matrices $\{X_k, k = 0, \dots, k_0\}$ and the corresponding set $\{Z_0(k), k = 0, \dots, k_0\}$ of subsets of T satisfying (6). Suppose that the representation (9) of the set $Z_0(k_0)$ is known and let $M_*(j)$, $\xi(j)$, $j \in J$, be the corresponding sets and vectors defined above. Suppose that there exists $t^* \in Z_0(0) \setminus Z_0(k_0)$ such that for any $D \in \mathbb{S}(p)$, the qualities

$$e'_k D \xi(j) = 0 \quad \forall k \in M_*(j), \quad \forall j \in J, \quad (10)$$

imply the equalities

$$e'_k D t^* = 0 \quad \forall k \in \text{supp}(t^*). \quad (11)$$

Then the set F defined in (7) is a non-exposed face of $\mathbb{CP}(p)$. Here $\{e_k, k = 1, \dots, p\}$ is the standard basis in $\mathbb{R}^p$.

In contrast to conditions of Theorem 2, the conditions of Proposition 2 can be easily verified.

In fact, for any matrix $D \in \mathbb{S}(p)$ with elements d_{ij} , $i = 1, \dots, p$, $j = i, \dots, p$, we define the vector $svec(D) \in \mathbb{R}^{p(p+1)/2}$ by the rule

$$svec(D) = (d_{11}, \sqrt{2}d_{12}, \dots, \sqrt{2}d_{1p}, d_{22}, \sqrt{2}d_{23}, \dots, \sqrt{2}d_{2p}, \dots, d_{pp})^\top.$$

Suppose that $\xi(j)$, $M_*(j)$, $j \in J$, and t^* are given. Then it is easy to construct matrices $\mathbb{B} \in \mathbb{R}^{p_a \times p(p+1)/2}$ with $p_a = \sum_{j \in J} |M_*(j)|$ and $\mathbb{W} \in \mathbb{R}^{p_b \times p(p+1)/2}$ with $p_b = |supp(t^*)|$ such that

$$B = \{D \in \mathbb{S}(p) : \mathbb{B}svec(D) = 0\},$$

$$\{D \in \mathbb{S}(p) : e'_k D t^* = 0 \ \forall k \in supp(t^*)\} = \{D \in \mathbb{S}(p) : \mathbb{W}svec(D) = 0\}.$$

Then it is evident that condition (10), (11) can be formulated as follows

$$rank \mathbb{B} = rank \begin{pmatrix} \mathbb{B} \\ \mathbb{W} \end{pmatrix}$$

and tested by any available algorithm calculating the rank of a given matrix.

Examples. We present examples of non-exposed (polyhedral and non-polyhedral) faces for $p \geq 6$. For any $p \geq 6$, we describe non-trivial families of non-exposed faces $\mathbb{CP}(p)$. By doing this, we demonstrate that the existence of non-trivial non-exposed faces is not an exclusive feature of only 5×5 completely positive cone.

References

1. Handbook on Semidefinite, Conic and Polynomial Optimization / eds.: M. F. Anjos, J. B. Lasserre. New York : Springer, 2012. (International Series in OR/MS, Vol. 166.) DOI: [10.1007/978-1-4614-0769-0](https://doi.org/10.1007/978-1-4614-0769-0).
2. Letchford A. N, Parkes A. J. A guide to conic optimisation and its applications // RAIRO - Oper. Res. 2018. Vol. 52, iss. 4–5. P. 1087–1106.
3. Bomze I. M. Copositive optimization - Recent developments and applications // European Journal of Operations Research. 2012. Vol. 216, iss. 3. P. 509–520.
4. Bomze I. M., Schachinger W., Uchida G. Think co(mpletely) positive! Matrix properties, examples and a clustered bibliography on copositive optimization // Journal of Global Optimization. 2012. Vol. 52, iss. 3. P. 423–445.

5. *Dickinson P. J.* Geometry of the copositive and completely positive cones // Journal of Mathematical Analysis and Applications. 2011. Vol. 380, iss 1. P. 377–395.
6. *Dür M.* Copositive Programming – a Survey. // Recent advances in Optimization and its applications in Engineering : The 14th Belgian-French-German Conference on Optimization / eds.: Diehl M. [et al.]. Berlin, Heidelberg : Springer Berlin Heidelberg, 2010. P. 3–20.
7. *Hoffman A. J., Pereira F.* On copositive matrices with -1, 0, 1 entries // Journal of Combinatorial Theory. Series A, 1973. Vol. 14. iss. 3. P. 302–309.
8. *Kostyukova O. I., Tchemisova T. V.* On equivalent representations and properties of faces of the cone of copositive matrices // Optimization. 2022. Vol. 71, iss. 11. P. 3211–3239.
9. *Wang F., Wolkowicz H.* Singularity degree of non-facially exposed faces. URL: <https://arxiv.org/abs/2211.00834> (date of access: 10.09.2025).
10. *Pataki G.* On the closedness of the linear image of a closed convex cone // Mathematics of Operations Research. 2007. Vol. 32, iss. 2. P. 395–412.
11. *Eichfelder G., Jahn J.* Set-Semidefinite Optimization // Journal of Convex Analysis. 2008. Vol. 15, iss. 4, P. 767–801.
12. *Kostyukova O. I., Tchemisova T. V.* Representation of zeros of a copositive matrix via maximal cliques of a graph // Linear Algebra and its Applications. 2025. Vol. 717, iss. 15. P. 40–55.
13. *Berman A., Dür M., Shaked-Monderer N.* Open problems in the theory of completely positive and copositive matrices // The Electronic Journal of Linear Algebra. 2015. Vol. 29. P. 46–58.
14. *Zhang Q.* Completely positive cones: are they facially exposed? // Linear Algebra and its Applications. 2018. Vol. 558, P. 195–204.

ОЦЕНИВАНИЕ ПАРАМЕТРОВ НЕЛИНЕЙНЫХ РЕГИСТРОВ СДВИГА С ОБРАТНОЙ СВЯЗЬЮ

В. Ю. Палуха¹⁾, А. А. Пардаев²⁾

¹⁾ НИИ прикладных проблем математики и информатики,
Белорусский государственный университет,
Минск, Беларусь, Palukha@bsu.by

²⁾ Белорусский государственный университет,
Минск, Беларусь, fpm.pardaev@bsu.by

В данной работе исследуется оценивание параметров нелинейных регистров сдвига с обратной связью (NLFSR) с использованием моделей цепей Маркова. Анализируются параметры NLFSR с помощью статистического алгоритма, основанного на цепях Маркова с частичными связями, для оценки их криптографической стойкости. Вычислительный эксперимент демонстрирует эффективность алгоритма в восстановлении функций обратной связи.

Ключевые слова: криптографические генераторы; цепи Маркова; нелинейные регистры сдвига; статистический анализ; псевдослучайные последовательности.

EVALUATION OF PARAMETERS OF NONLINEAR FEEDBACK SHIFT REGISTERS

U. Yu. Palukha^{a)}, A. A. Pardaev^{b)}

^{a)} Research Institute for Applied Problems of Mathematics and Informatics,
Belarusian State University,
Minsk, Belarus, Palukha@bsu.by

^{b)} Belarusian State University,
Minsk, Belarus, fpm.pardaev@bsu.by

This paper investigates the evaluation of parameters of nonlinear feedback shift registers (NLFSRs) using Markov chain models. We analyze the parameters of NLFSRs through a statistical algorithm based on Markov chains with partial connections, assessing their cryptographic strength. A computational experiment demonstrates the effectiveness of the algorithm in recovering feedback functions.

Keywords: cryptographic generators; Markov chains; nonlinear feedback shift registers; statistical analysis; pseudorandom sequences.

1. Введение

Современные информационные системы требуют надёжных методов защиты данных, что делает криптографические системы их неотъемлемой частью. Важнейшим элементом криптографических систем являются генераторы случайных и псевдослучайных чисел, которые используются для формирования ключей шифрования, генерации гамм в поточных шифрах и решения других критически важных задач [1]. Безопасность всей системы напрямую зависит от качества этих генераторов, поскольку предсказуемость их выходных последовательностей может сделать криптографические алгоритмы уязвимыми.

Основное требование к таким генераторам – статистическая неотличимость вырабатываемых последовательностей от истинно случайных. Большинство криптографических генераторов представляют собой рекуррентную функцию некоторого порядка, вероятностной моделью для которых служит цепь Маркова. Криптографическая стойкость генератора ставится под сомнение, если его работу можно аппроксимировать такой моделью, поскольку это открывает потенциал для статистических атак, направленных на оценивание её параметров. Успешное оценивание параметров цепи Маркова, аппроксимирующей генератор, будет однозначно свидетельствовать о его недостаточном качестве.

2. Математическая основа и анализ регистров сдвига с нелинейной обратной связью

Нелинейные функции регистра сдвига с обратной связью были в центре внимания многих исследователей. Например, в [2] был представлен список n -битных NLFSR с оптимальным периодом $2^n - 1$ для $n < 25$. В данном исследовании использовались следующие три типа функций обратной связи с алгебраической степенью два:

$$f(x_0, x_1, \dots, x_{n-1}) = x_0 \oplus x_a \oplus x_b \oplus x_c x_d, \quad (1)$$

$$f(x_0, x_1, \dots, x_{n-1}) = x_0 \oplus x_a \oplus x_b x_c \oplus x_d x_e, \quad (2)$$

$$f(x_0, x_1, \dots, x_{n-1}) = x_0 \oplus x_a \oplus x_b \oplus x_c \oplus x_d \oplus x_e x_h, \quad (3)$$

где $a, b, c, d, e, h \in \{1, 2, \dots, n-1\}$; $x_i \in \{0, 1\}$.

Параметры этих функции оценивались в [3], однако в работе [2] были перечислены не все функции обратной связи с оптимальными периодами. В [4] был предложен новый и эффективный метод построения. Авторы использовали этот метод для построения отсутствующих функций обратной

связи NLFSR трёх типов, приведённых ранее. Они представили полные списки этих типов для всех размеров $n \in \{4, 5, \dots, 19\}$.

В [5] же предлагается новый тип функции обратной связи NLFSR степени 2 с оптимальными периодами для $n \leq 23$ следующего вида:

$$f(x_0, x_1, \dots, x_{n-1}) = x_0 \oplus x_a \oplus x_b \oplus x_c \oplus x_d \oplus x_e \oplus x_h \oplus x_w x_z, \quad (4)$$

где $a, b, c, d, e, h, w, z \in \{1, 2, \dots, n-1\}$; $x_i \in \{0, 1\}$. Всего авторами было предложено 639 новых функций NLFSR с оптимальными периодами для $8 \leq n \leq 23$. В табл. 1 представлены сводки этих новых функций.

Таблица 1

Количество новых функций обратной связи NLFSR с оптимальными периодами

n	Количество функций	n	Количество функций
8	12	16	84
9	12	17	70
10	34	18	44
11	26	19	23
12	64	20	24
13	64	21	17
14	76	22	12
15	70	23	7
Всего	639		

Таким образом, исследование функций нового типа позволяет расширить возможности анализа и применения NLFSR. В рамках нашей работы мы сосредоточимся на изучении именно этих функций, поскольку они предоставляют более сложные механизмы генерации последовательностей. А использование цепей Маркова в анализе позволит глубже понять их структуру и выявить скрытые закономерности.

3. Использование цепей Маркова для оценки параметров регистра сдвига

Анализ поведения нового типа NLFSR, характеризующегося сложной структурой функций обратной связи, требует применения методов, учитывающих детерминированность генератора и наличие скрытых зависимостей между битами состояния. Модель цепи Маркова s -го порядка с r частичными связями ЦМ(s, r) позволяет аппроксимировать генератор, минимизируя количество параметров модели. Основным предположением является то, что вероятность текущего состояния зависит от подмножества r значимых состояний из истории длиной s .

В [6] был предложен алгоритм оценивания параметров ЦМ(s, r). С его помощью строится оценка матрицы вероятностей одношаговых переходов Q , которая может быть использована для построения таблицы истинности булевой функции. Матрица Q имеет размерность $2^r \times 2$. Строки матрицы $(q_{i_{m_1}, \dots, i_{m_r}, 0}, q_{i_{m_1}, \dots, i_{m_r}, 1})$ содержат вероятности генерации 0 или 1 в зависимости от предыстории $i_{m_1}, \dots, i_{m_r}$. Булева функция описывается следующим образом:

$$f(i_{m_1}, \dots, i_{m_r}) = \begin{cases} 0, & q_{i_{m_1}, \dots, i_{m_r}, 0} > q_{i_{m_1}, \dots, i_{m_r}, 1}; \\ 1, & q_{i_{m_1}, \dots, i_{m_r}, 0} \leq q_{i_{m_1}, \dots, i_{m_r}, 1}. \end{cases} \quad (5)$$

В качестве вектора значений булевой функции возьмём второй столбец матрицы Q , так как если генератор без искажений, то строки матрицы будут иметь вид (0 1) либо (1 0).

4. Алгоритм оценивания регистра сдвига с помощью цепей Маркова

Для построения алгоритма оценки регистра сдвига с нелинейной обратной связью на основе модели ЦМ(s, r) необходимо в первую очередь определить следующие параметры:

оптимальный порядок цепи Маркова s , который характеризует длину истории, необходимую для описания динамики регистра;

число значимых связей r , минимизирующее количество битов, необходимых для восстановления функции регистра;

шаблон значимых битов M_r , указывающий, какие биты истории оказывают наибольшее влияние на переходы.

Первый шаг – подготовка данных. Включает в себя генерацию последовательностей длины T с использованием функции обратной связи f_k .

Второй шаг – применение алгоритма оценки параметров ЦМ(s, r). С помощью данного алгоритма оценивается порядок $s \in [s_-, s_+]$, число значимых связей $r \in [r_-, r_+]$, находится шаблон M_r и матрица переходов Q .

Третий шаг – построение таблицы истинности функции обратной связи на основе матрицы Q . Перечисляются все возможные 2^r состояний значимых битов истории $i_{m_1}, \dots, i_{m_r}$, где $i_j \in \{0, 1\}$. Затем для каждого состояния вычисляется (5). Таким образом получаем таблицу истинности.

Четвёртый шаг – восстановление функции f в алгебраической нормальной форме (АНФ). АНФ булевой функции $f(x_0, x_1, \dots, x_r)$ выражается как

$$f(x_0, x_1, \dots, x_r) = c_0 \oplus c_1 x_1 \oplus c_2 x_2 \oplus \dots \oplus c_{12} x_1 x_2 \oplus \dots, \quad (6)$$

где $c_i \in \{0, 1\}$.

Зная таблицу истинности, с помощью метода неопределённых коэффициентов строится АНФ, которая должна совпасть с изначальной функцией регистра сдвига.

5. Компьютерный эксперимент

В эксперименте рассмотрены пять NLFSR, каждый из которых характеризуется своей функцией обратной связи f_i , размером регистра s , числом связей r . Параметры регистров приведены в табл. 2.

Таблица 2

Параметры регистров сдвига и алгоритма оценивания

F	s	r	r_-	r_+
$f_1 = x_0 \oplus x_1 \oplus x_7 \oplus x_{10} \oplus x_9 x_{15}$	17	6	4	7
$f_2 = x_0 \oplus x_1 \oplus x_8 \oplus x_9 \oplus x_{15} \oplus x_7 x_{18}$	24	7	5	8
$f_3 = x_0 \oplus x_2 \oplus x_9 \oplus x_{10} \oplus x_{12} \oplus x_{13} \oplus x_{14} \oplus x_3 x_6$	17	9	7	10
$f_4 = x_0 \oplus x_1 \oplus x_2 \oplus x_3 \oplus x_4 \oplus x_5 \oplus x_6 \oplus x_2 x_4$	8	7	5	8
$f_5 = x_0 \oplus x_1 \oplus x_7 \oplus x_{15} \oplus x_{16} \oplus x_{19} \oplus x_{21} \oplus x_3 x_{19}$	23	8	6	9

Для каждого NLFSR сгенерировано по три последовательности длиной $T = 4000$ бит. Начальные состояния регистров выбирались случайным образом, чтобы обеспечить разнообразие входных данных. Последовательности генерировались с использованием функции обратной связи, где на каждом шаге вычислялся следующий бит, а регистр сдвигался с добавлением нового бита.

Эксперимент был проведён для всех пяти NLFSR. Для каждого регистра и каждой из трёх последовательностей алгоритм корректно определил порядок r , шаблон M_r , и АНФ функции обратной связи была восстановлена верно. На рис. 1 проиллюстрирован вывод программы в ходе восстановления функции f_3 .

В предыдущих экспериментах, описанных в [3], длина последовательности $T = 1000$ бит была достаточной для восстановления функций

```

Регистр f3 (s=17, истинный r=9, истинный M=[0, 2, 3, 6, 9, 10, 12, 13, 14], r_minus=7, r_plus=10):
2025-05-16 01:15:28,818 - INFO - ► Начало анализа: регистр f3, последовательность 1
2025-05-16 01:15:28,818 - INFO - ►► A3: регистр f3, последовательность 1
2025-05-16 01:15:28,821 - INFO - Оценка 11440 шаблонов для r=10
2025-05-16 01:17:32,860 - INFO - A1 завершен: r=10, I=0.6926, шаблон: (0, 2, 3, 6, 9, 10, 12, 13, 14, 15), время: 124.04с
2025-05-16 01:17:32,870 - INFO - Оценка 9 уменьшенных шаблонов для r=9
2025-05-16 01:17:32,951 - INFO - Оценка 8 уменьшенных шаблонов для r=8
2025-05-16 01:17:33,013 - INFO - Оценка 7 уменьшенных шаблонов для r=7
2025-05-16 01:17:33,060 - INFO - A3 завершен: лучший r=9, шаблон: (0, 2, 3, 6, 9, 10, 12, 13, 14), время: 124.24с
Последовательность 1: ☒ Успех
Оценённый r: 9, Оценённый M: (0, 2, 3, 6, 9, 10, 12, 13, 14)
Восстановленная АНФ: x0 ^ x10 ^ x12 ^ x13 ^ x14 ^ x2 ^ x9 ^ (x3 & x6)

```

Рис. 1. Восстановление функции реализованной программой

обратной связи более простых регистров типов, описанных в [2]. Однако для успешного восстановления NLFSR нового типа, описанных в разделе 2, потребовалась генерация последовательности большей длины. Это связано с экспоненциальным ростом числа шаблонов (2^r), для которых необходимо оценить вероятности переходов. Число наблюдений на шаблон можно приближённо оценить как $n = \frac{T-s}{2^r}$. Для точной оценки вероятностей требуется, чтобы это число было достаточно большим. В статистике стандартная ошибка для биномиальной вероятности p составляет

$$SE = \sqrt{\frac{p(1-p)}{n}}, \quad (7)$$

где n – число наблюдений. Для $p = 0.5$ ошибка равна $\frac{0.5}{\sqrt{n}}$, и тогда для ошибки менее 0.1 необходимо, чтобы выполнялось $n > 25$.

6. Заключение

Проведённый вычислительный эксперимент продемонстрировал высокую эффективность алгоритма для восстановления функций обратной связи NLFSR при использовании последовательностей длиной 4000 бит. Успешное восстановление во всех 15 случаях подтверждает, что выбранная длина последовательности обеспечивает достаточное количество данных для оценки вероятностей переходов даже для NLFSR с высоким порядком. Увеличение длины последовательности с 1000 до 4000 бит было необходимо из-за роста числа шаблонов, характерных для рассмотренных NLFSR. В будущем планируется подробнее исследовать минимальную длину последовательности для различных конфигураций s и r , чтобы оптимизировать вычислительные ресурсы при сохранении точности.

Библиографические ссылки

1. Криптология: учебник / Ю. С. Харин [и др.]. Минск : БГУ, 2013.
2. Dubrova E. A List of Maximum Period NLFSRs. URL: <https://eprint.iacr.org/2012/166> (date of access: 31.03.2025).
3. Палуха В. Ю., Харин Ю. С. Оценка качества криптографических генераторов на основе цепей Маркова высокого порядка // Вопросы защиты информации. 2015. №1. С. 12–14.
4. NLFSR Functions with Optimal Periods / S. Almuhammadi [et al.] // Computational Science and Its Applications – ICCSA 2018 : 18th International Conference, Melbourne, VIC, Australia, July 2–5, 2018, Proceedings, Part I / eds.: O. Gervasi [et al.]. Springer Cham, 2018. P. 67–79. (Lecture Notes in Computer Science, Vol. 10960.)
5. Al-Hejri I., Al-Kharobi T. A New Type of NLFSR Functions with Maximum Periods // Transactions on Networks and Communications. 2019. Vol. 7, № 2. P. 20–30.
6. Харин Ю. С., Петлицкий А. И. Цепь Маркова s -го порядка с r частичными связями и статистические выводы о ее параметрах // Дискретная математика. 2007. Т. 19. № 2. С. 109–130.

РАЗДЕЛ V

**ПАРАЛЛЕЛЬНАЯ И РАСПРЕДЕЛЁННАЯ
ОБРАБОТКА ДАННЫХ**

СРАВНЕНИЕ МНОГОПОТОЧНЫХ МОДЕЛЕЙ В C++, GO И RUST В ЗАДАЧАХ ПАРАЛЛЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ

Е. А. Азаров, А. Н. Марков, С. Н. Барсукевич

*Белорусский государственный университет информатики и радиоэлектроники,
Минск, Беларусь, georga399@gmail.com*

Данное исследование сравнивает эффективность применения многопоточных моделей в C++, Go и Rust в задачах параллельной обработки данных. Анализируются производительность, безопасность и удобство разработки. Наибольшая производительность достигается в C++, Go продемонстрировал лучшую масштабируемость, а Rust показал лучшую безопасность. Результаты подкреплены бенчмарками производительности.

Ключевые слова: многопоточность; C++; Go; Rust; параллельная обработка; производительность; безопасность.

COMPARATIVE ANALYSIS OF MULTITREADING MODELS IN C++, GO AND RUST FOR PARALLEL DATA PROCESSING TASKS

E. A. Azarov, A. N. Markov, S. N. Barsukevich

*Belarusian State University of Informatics and Radioelectronics,
Minsk, Belarus, georga399@gmail.com*

This study compares the effectiveness of using multithreaded models in C++, Go and Rust in parallel data processing tasks. It analyzes performance, safety, and development convenience. The highest performance is achieved in C++, Go demonstrated the best scalability and Rust showed the best safety. The results are supported by performance benchmarks.

Keywords: multithreading; C++; Go; Rust; parallel processing; performance; safety.

1. Введение

Многопоточность является ключевым инструментом для параллельной обработки данных, позволяя эффективно использовать многоядерные процессоры в задачах с большими объемами данных. Такие языки программирования, как C++, Go и Rust, предлагают различные механизмы реализации многопоточности, отличающиеся по производительности, безопасности и удобству разработки. Данное исследование анализирует их

возможности в контексте задач обработки данных, таких как вычисления и операции ввода-вывода, и предоставляет результаты бенчмарков для оценки производительности, а также определяет оптимальные сценарии использования [1].

2. Многопоточность в C++, Go, Rust

C++ предоставляет низкоуровневые механизмы многопоточности через стандартную библиотеку, начиная с C++11. Основные инструменты включают потоки *std::thread*, *std::atomic*. Эти примитивы обеспечивают точный контроль над потоками, но требуют внимательного управления для предотвращения гонок данных и взаимных блокировок. Внешние библиотеки, такие как *OpenMP* и *Intel TBB*, упрощают реализацию параллельных алгоритмов, особенно для задач с неоднородной нагрузкой [2].

Go использует модель конкурентности, основанную на горутинах – легковесных потоках, управляемых средой выполнения. Планировщик Go эффективно распределяет горуты по системным потокам, что делает его подходящим для задач с высокой конкуренцией, таких как обработка сетевых запросов. Каналы *chan* обеспечивают безопасную передачу данных между горутами, но сборка мусора может вызывать паузы в вычислительно интенсивных задачах [3].

Rust применяет уникальную систему владения и заимствования, которая предотвращает гонки данных на этапе компиляции. Типы *Send* и *Sync* обеспечивают безопасную передачу данных между потоками. Библиотеки, такие как *Rayon* и *Tokio*, расширяют возможности для параллельных и асинхронных вычислений, делая Rust подходящим для критически важных систем, где безопасность имеет приоритет [4].

3. Разработка тестовых сценариев

Для сравнительного анализа были составлены следующие тестовые сценарии: параллельная обработка независимых задач (*Coro-Prime-Sieve*), синхронизированный доступ к общему ресурсу, конвейерная обработка с передачей данных.

Тесты проводились в операционной системе *Linux Ubuntu 22.04* на вычислительной машине с процессором *Intel Core i5-1135G7* (8 ядер) с использованием компиляторов *g++ 12.2 (-O3)*, *Go 1.21* и *Rust 1.68*.

4. Сравнительный анализ параллельной обработки независимых задач

Задача *Coro-Prime-Sieve* разделяет вычисление простых чисел на подзадачи. C++ использует *std::thread*, Go – горуты, Rust – *std::thread*. C++ показывает лучшее время выполнения (450 мс) для конкретной

конфигурации (8 потоков), однако *Go* показал лучшую масштабируемость, при увеличении числа потоков, благодаря легковесным горутинам.

Сравнение производительности в задаче вычисления простых чисел до заданного предела (вычисления простых чисел меньших 4000) с использованием решета Эратосфена [5], разделенного на независимые подзадачи, представлено на рис. 1.

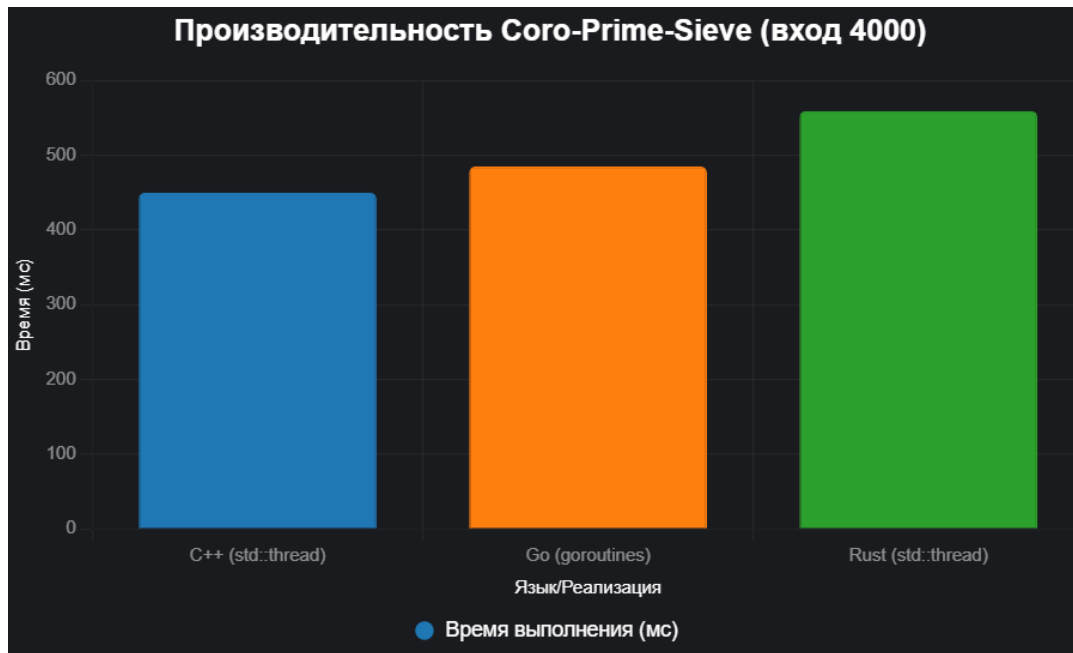


Рис. 1. График производительности в задаче *Coro-Prime-Sieve*

C++ использует явное создание потоков (*std::thread*) со статическим распределением диапазонов чисел между потоками. Каждый поток обрабатывает заранее выделенный сегмент решета Эратосфена независимо. Синхронизация между потоками отсутствует, так как подзадачи полностью независимы.

Go применяет динамическое распределение подзадач (проверка чисел на простоту) между горутинами через пул воркеров (набор горутин-обработчиков, каждая из которых последовательно выполняет задачи из общей очереди) или каналы. Горутины, завершившие свою задачу, немедленно получают новую из общего канала. Лучшая масштабируемость обеспечивается за счёт низкой стоимости создания горутин (0,2–0,5 мкс), эффективной работой планировщика *Go* (*M:N* модель), динамически балансирующего нагрузку на системные потоки, отсутствием перегрузки при создании большого количества горутин.

Rust реализует подход, аналогичный *C++*, со статическим распределением диапазонов по потокам (*std::thread*). Ключевая особенность – гарантии

безопасности на этапе компиляции. Система владения *Rust* проверяет отсутствие гонок данных при передаче сегментов данных в потоки (тип *Send*), исключая необходимость в *runtime*-проверках или неявных копиях.

5. Сравнительный анализ задачи с синхронизированным доступом к общему ресурсу

Задача заключается в инкременте общего счетчика (10^6 итераций) несколькими потоками с использованием средств синхронизации [6].

График сравнения производительности в задаче с синхронизированным доступом изображен на рис. 2.

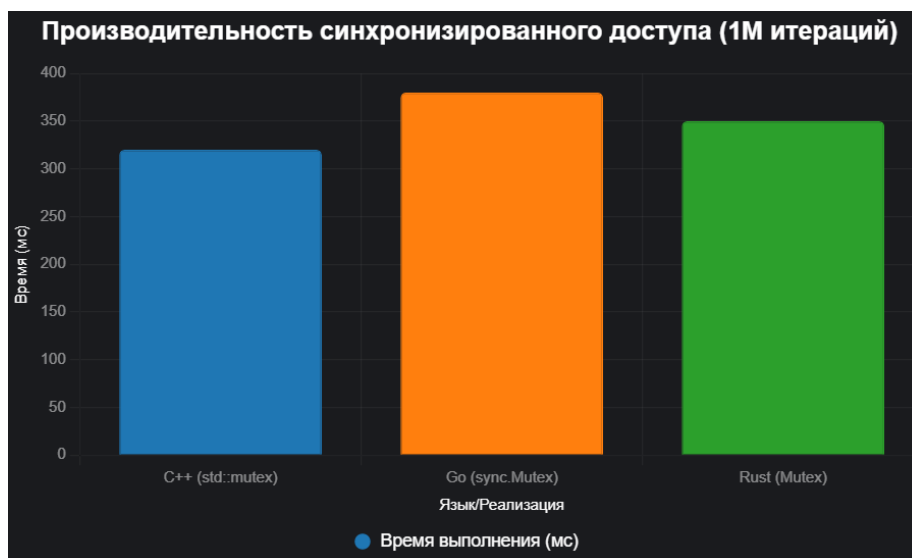


Рис. 2. Сравнение производительности в задаче с синхронизированным доступом

C++ использует *std::mutex* с явным блокированием (*lock()/unlock()*), *Go* – *sync.Mutex*, *Rust* – *Mutex*. *C++* показал лучшее время (320 мс) благодаря низким накладным расходам. В *Rust* потенциальные ситуации, когда два или более потока бесконечно ожидают освобождения ресурсов, заблокированных друг другом (*deadlock'u*), обнаруживаются только на этапе выполнения (как и в *C++/Go*), но гарантируется отсутствие гонок данных (*data races*) на этапе компиляции.

6. Сравнительный анализ задачи конвейерной обработки с передачей данных

Задача заключается в реализации конвейера, который обрабатывает 10^3 элементов в три этапа (вычисление квадрата, фильтрация, суммирование) [7]. *Go* показал лучшее время (260 мс) благодаря каналам.

График сравнения производительности конвейерной обработки показан на рис. 3.



Рис. 3. Сравнение производительности конвейерной обработки

Программа тестового сценария на C++ реализуется с использованием очередей (`std::queue` или `std::deque`) для хранения элементов между этапами, мьютексов (`std::mutex`) и условных переменных (`std::condition_variable`) для синхронизации доступа к очередям и уведомления потоков о появлении новых данных или свободного места, явных потоков (`std::thread`) для каждого этапа конвейера.

Go использует каналы (`chan`) как основной механизм передачи данных между горутинами-этапами. Лучшая производительность (260 мс) достигается за счет семантики *CSP* (*Communicating Sequential Processes*): каналов – первоклассной конструкции языка, глубоко интегрированной в планировщик и легковесности горутин. Передача данных через канал часто эффективнее, чем через разделяемую память с мьютексом.

Rust использует каналы `std::sync::mpsc`, а именно *Multi-Producer, Single-Consumer* каналы (похожи на каналы Go, но с одним получателем). Передача данных обычно требует перемещения (*move* семантика). Гарантии владения обеспечивают безопасную передачу. Производительность высока (близка к Go) благодаря эффективной реализации.

7. Сравнение накладных расходов

Накладные расходы при реализации многопоточности представляют собой существенный фактор, влияющий на общую производительность параллельных систем. Величина этих расходов детерминирована как

архитектурными особенностями языка программирования, так и выбранной моделью параллелизма.

Ниже (табл. 1) приведено сравнение накладных расходов многопоточных механизмов в *C++*, *Go*, *Rust* [7].

Таблица 1

Сравнение накладных расходов многопоточных механизмов в *C++*, *Go*, *Rust*

Операция	<i>C++</i>	<i>Go</i>	<i>Rust</i>
Создание потока	10–100 мкс	0,2–0,5 мкс	10–100 мкс
Переключение	1–3 мкс	0,1–0,3 мкс	1–3 мкс
Блокировка мьютекса	20–50 нс	30–70 нс	40–80 нс
Атомарная операция	10–30 нс	–	15–40 нс

Примечание. Для языка *Go* показатели атомарных операций не включены ввиду преобладания канальной коммуникации как основного метода синхронизации в идиоматических паттернах. Замеры проводились для стандартных примитивов синхронизации (*std::mutex*, *sync.Mutex*, *std::sync::Mutex*).

По данным таблицы можно сделать вывод, что *Go* лидирует в переключении контекста. *C++* минимален по накладным расходам, *Rust* обеспечивает безопасность с умеренными накладными расходами.

8. Заключение

Сравнительное тестирование показало, что *C++*, *Go* и *Rust* демонстрируют разную производительность и накладные расходы в параллельных вычислениях, что определяет их применимость для различных классов задач. *C++* обеспечивает максимальную производительность в задачах с интенсивными вычислениями и синхронизацией, что делает его предпочтительным для высокопроизводительных систем, таких как научные вычисления или игровые движки. *Go* упрощает разработку высококонкурентных приложений, таких как веб-серверы или системы обработки потоковых данных, благодаря легковесным горутинам и каналам. *Rust* гарантирует безопасность памяти, что критически важно для надежных систем, например, в телекоммуникациях или блокчейн-приложениях.

Результаты бенчмарков в проведенном исследовании помогут разработчикам выбрать язык в зависимости от требований проекта.

Библиографические ссылки

1. Benchmarking low-level I/O: C, C++, Rust, Golang, Java, Python. URL: <https://medium.com/star-gazers/benchmarking-low-level-i-o-c-c-rust-golang-java-python-9a0d505f85f7> (date of access: 05.06.2025).
2. C++ Reference: Concurrency support library. URL: <https://en.cppreference.com/w/cpp/thread> (date of access: 05.06.2025).
3. The Go Programming Language. URL: <https://go.dev/doc/> (date of access: 05.06.2025).
4. The Rust Programming Language. URL: <https://doc.rust-lang.org/book/> (date of access: 05.06.2025).
5. Prime number generator using Sieve of Eratosthenes. URL: <https://felspar.com/coro/examples/primes-1-generator.cpp> (date of access: 05.06.2025).
6. *Herlihy M., Shavit N.* The Art of Multiprocessor Programming. Morgan Kaufmann, 2012.
7. *Mattson T. G., Sanders B. A., Massingill B. L.* Patterns for Parallel Programming. Addison-Wesley, 2004.

ИСПОЛЬЗОВАНИЕ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ ИИ ДЛЯ МАСКИРОВКИ АНОМАЛЬНОЙ СЕТЕВОЙ АКТИВНОСТИ В РАСПРЕДЕЛЁННЫХ СИСТЕМАХ

М. А. Баранчик, Е. Г. Крупенич, С. А. Мигалевич

*Белорусский государственный университет информатики и радиоэлектроники,
Минск, Беларусь, runolero@gmail.com*

Рассматривается новая потенциальная угроза кибербезопасности, связанная с использованием генеративных моделей искусственного интеллекта (GAN, VAE и др.) для маскировки аномальной сетевой активности в распределённых системах. Теоретически обосновывается возможность использования моделей, обученных на нормальном пользовательском поведении, для сокрытия вредоносных действий. Предлагаются возможные меры противодействия.

Ключевые слова: генеративные модели; информационная безопасность; аномальная сетевая активность; распределённые системы.

USING GENERATIVE AI MODELS TO MASK ANOMALOUS NETWORK ACTIVITIES IN DISTRIBUTED SYSTEMS

M. A. Baranchik, E. G. Krupenich, S. A. Migalevich

*Belarusian State University of Informatics and Radioelectronics,
Minsk, Belarus, runolero@gmail.com*

A new potential cybersecurity threat is considered, associated with the use of generative artificial intelligence models (GAN, VAE, etc.) to mask anomalous network activity in distributed systems. The possibility of using models trained on normal user behavior to hide malicious actions is theoretically substantiated. Possible countermeasures are proposed.

Keywords: generative models; information security; anomalous network activity; distributed systems.

1. Введение

Современные распределённые информационные системы характеризуются высокой степенью масштабируемости, неоднородностью компонентов и динамичной структурой взаимодействия между узлами [1]. Компоненты распределённой инфраструктуры обмениваются значительными объёмами данных, часто в реальном времени, что создаёт множество точек

потенциального внедрения злоумышленника. Угрозы, использующие такие уязвимости, проявляются в виде статистически или поведенчески отклоняющегося от нормы трафика.

Одним из основных инструментов мониторинга и защиты сетевого периметра являются системы обнаружения вторжений (Intrusion Detection Systems, IDS), действующие на основе анализа сетевого трафика [2]. IDS выполняют автоматизированный сбор и обработку информации о проходящих через сеть пакетах данных, сопоставляя полученные метаданные с заранее известными сигнатурами атак или оценивая статистические и поведенческие аномалии (Heuristics-Based Anomaly Detection).

Среди систем обнаружения вторжений (IDS/IPS) выделяют два типа.

- Синтаксические (сигнатурно-ориентированные) выявляют атаки путём сравнения сетевого трафика с базой известных шаблонов (сигнатур), где каждая сигнатура описывает последовательность байтов или специфические фрагменты пакетов, характерные для конкретного вида угрозы. Такие системы эффективны против ранее задокументированных атак, но оказываются бесполезными при появлении новых, неизвестных видов вредоносного ПО.

- Поведенческие (анализ аномалий) строят модель «нормального» трафика на основе статистических, временных или семантических характеристик, а затем ищут отклонения от этой базы. Ключевой недостаток данного подхода заключается в высокой вероятности ложных срабатываний при изменении легитимных сценариев работы системы и возможном адаптивном поведении злоумышленника.

С развитием генеративных моделей ИИ появляется новая угроза – возможность маскировки вредоносной активности под легитимное поведение, обучая модели на «норме» и встраивая отклонения в рамки ожидаемых параметров.

2. Аномальная активность как форма вредоносного воздействия

В настоящее время наиболее распространены следующие формы вредоносного воздействия.

1. Сетевое сканирование и разведка. Атаки начинаются с непрерывного отправления запросов к большому количеству портов и служб с целью выявить активные сервисы и определить уязвимости на удалённых узлах. Часто такие сканирования выполняются распределённо (распределённые сканирующие боты), что позволяет снизить уровень подозрительности с точки зрения традиционных IDS.

2. Эксплуатация уязвимостей приложений и служб. После разведки злоумышленник использует конкретные уязвимости (например, RCE,

SQL-инъекции, ошибки в протоколах) для внедрения вредоносного кода или получения несанкционированного доступа. Данная активность проявляется в виде аномальных последовательностей запросов, отклоняющихся от шаблонов легитимного трафика.

3. DDoS-атаки. Распространённая форма вывода сервиса из строя путём создания огромного объёма трафика с большого числа скомпрометированных устройств. Такое поведение характеризуется резким ростом частоты и объёмов пакетов, и традиционные IDS/IPS хорошо детектируют подобные всплески, однако при «мягких» DDoS-сценариях (slowloris, HTTP-флуд с малой скоростью) аномалия становится менее заметной.

4. Эксфильтрация данных. Скрытая передача конфиденциальной информации из внутренней сети на внешние ресурсы. Злоумышленники часто используют зашифрованные каналы (HTTPS, DNS-туннели), пытаясь замаскировать объёмы передаваемых данных и их частоту под обычный пользовательский трафик.

5. Brute-force атаки (перебор аутентификационных данных). Повторные попытки входа в систему с разными учётными данными генерируют аномально большое количество запросов к механизму аутентификации. Хотя отдельные точки остаются незаметными, совокупность таких попыток создаёт характерный паттерн, легко обнаруживаемый поведенческими IDS [3].

3. Механизм маскировки аномальной активности с использованием генеративных моделей

Современные генеративные модели искусственного интеллекта, включая генеративные состязательные сети (GAN), вариационные автокодировщики (VAE) и диффузионные модели, обладают способностью воссоздавать сложные распределения данных, имитируя поведение, неотличимое от нормального [4]. В контексте сетевого трафика это означает возможность синтезировать потоки, статистически соответствующие легитимной активности, при этом скрывая действия, обладающие вредоносной направленностью. Такой подход создаёт угрозу подмены поведения на уровне сетевого взаимодействия, при которой зловредные операции маскируются под фоновую активность, признанную системой мониторинга допустимой.

Злоумышленник, обладая доступом к параметрам нормального сетевого поведения (частоте обращений к определённым портам, типам протоколов, средним и медианным задержкам в ответах, объёмам передаваемых данных и т.п.), может обучить генеративную модель, которая будет имитировать эти характеристики [5]. Затем она используется для внедрения вредоносного поведения в поток, сохраняющий признаки

легитимности. Таким образом, активность, которая по своей сущности нарушает политику безопасности, остаётся невидимой для средств обнаружения, работающих на основе отклонения от нормы.

С практической точки зрения маскировка может реализовываться в форме постобработки вредоносного трафика (перекрытие статистических характеристик) или путём генерации целого массива пакетов, среди которых зловердные команды распределяются во времени и потоке данных так, чтобы не нарушать допустимые пороги, заданные системой обнаружения. Например, эксфилльтрация данных может осуществляться малыми порциями, встроенными в имитируемую службу мониторинга или обновления, при этом сигнатура атаки отсутствует, а поведение в целом соответствует допустимому шаблону.

4. Потенциальная уязвимость информационной инфраструктуры

Сигнатурные системы обнаружения полагаются на известные шаблоны атак, включающие специфические последовательности пакетов, характерные байтовые структуры и определённые порты или адреса. При этом любая модификация структуры трафика, не затрагивающая ключевые сигнатуры, остаётся незамеченной. В случае использования генеративной модели злоумышленник имеет возможность формировать уникальные варианты атаки, не подпадающие под существующие шаблоны. Аналогично, детекторы аномалий оценивают отклонения от предварительно зафиксированных норм. Однако генеративный трафик, обладающий той же статистикой, что и реальный, оказывается вне зоны их чувствительности.

Поведенческий анализ основан на предположении, что отклонение от устоявшихся шаблонов действий – признак потенциальной угрозы. При использовании генеративных моделей возникает феномен «синтетической правдоподобности», при котором вредоносное поведение копирует паттерны легитимных процессов [6]. Это подрывает саму основу эвристического анализа, поскольку создаётся впечатление непрерывности и согласованности действий, в действительности нарушающих политику безопасности. В результате формируется ложное чувство защищённости, при котором поведенческие средства мониторинга не только не фиксируют угрозу, но и способствуют её незаметному распространению.

5. Контрмеры и направления развития систем обнаружения вторжений

Одним из перспективных направлений в борьбе с маскирующим вредоносным трафиком является выявление признаков синтетичности сетевых потоков. Несмотря на высокую степень достоверности, достигаемую

современными генеративными моделями, сформированные ими данные могут содержать скрытые следы искусственного происхождения. В частности, при использовании GAN или VAE в сгенерированном трафике возможно присутствие микроскопических регулярностей, отличающихся от естественных флуктуаций живого сетевого взаимодействия. К ним могут относиться, например, неестественно низкая энтропия временных интервалов между пакетами, одинаковые значения редко изменяющихся полей, а также слабо выраженные корреляции между полями различных уровней сетевого стека [7].

Помимо этого, синтетический трафик может демонстрировать недостаточную реакцию при изменении контекста – например, к текущей нагрузке, неожиданным отказам узлов, изменению политик маршрутизации. Таким образом, сочетание анализа скрытых структур данных и мониторинга контекстной адаптивности взаимодействия позволяет выявлять подозрительные потоки, даже если они формально соответствуют статистике нормального поведения.

Как показала практика использования генеративных моделей, наиболее эффективно синтетичность удаётся определять при наличии обученного «дискриминатора» – нейросетевой модели, специализированной на различении реальных и сгенерированных образцов [8]. Интеграция таких моделей в архитектуру системы обнаружения вторжений позволяет не просто анализировать признаки аномальности, а проводить полноценную верификацию достоверности источников трафика.

Подобные дискриминаторы могут быть обучены на специально размеченных выборках, содержащих как настоящие, так и синтетические потоки, сформированные с использованием известных моделей (например, GAN, VAE, diffusion-based). При достаточном разнообразии обучающих данных дискриминатор способен выявлять даже адаптированные к обходу IDS модели, определяя искусственность их поведения не по явным признакам, а по совокупности вторичных характеристик.

6. Заключение

Использование генеративных моделей искусственного интеллекта (GAN, VAE и др.) представляет собой принципиально новый вектор развития угроз кибербезопасности в распределённых системах. За счёт способности имитировать легитимный сетевой трафик с высокой степенью достоверности, такие модели позволяют злоумышленникам эффективно скрывать вредоносную активность, не выходя за рамки поведенческой или статистической нормы, принятой в системах мониторинга. Это подрывает надёжность как сигнатурных, так и поведенческих систем обнаружения вторжений, в корне меняя принципы определения аномального поведения.

Предложенные в работе методы противодействия – в частности, разработка IDS нового поколения с интеграцией дискриминаторов, обученных на различение синтетического и реального трафика – открывают перспективное направление повышения устойчивости информационной инфраструктуры к новым типам угроз. Анализ латентных признаков, поведенческой пластичности и структурных особенностей трафика, дополненный механизмами верификации достоверности источников, позволяет значительно повысить чувствительность к атакам, маскирующимся под норму.

Современные вызовы кибербезопасности требуют переосмысления традиционных подходов к обнаружению угроз и перехода к более гибким, адаптивным системам защиты, способным учитывать появление синтетических форм цифровой активности. В этом контексте критически важно не только развитие технологий обнаружения, но и формирование междисциплинарной исследовательской базы, объединяющей знания в области ИИ, сетевой безопасности и анализа поведения.

Библиографические ссылки

1. Распределенные информационные системы: особенности применения и построения. URL: <https://apni.ru/article/6996-raspredelennye-informatsionnye-sistemi-osoben> (дата обращения: 05.06.2025).
2. Как обнаруживать вредоносный сетевой трафик. URL: <https://labex.io/ru/tutorials/wireshark-how-to-detect-malicious-network-traffic-419254> (дата обращения: 05.06.2025).
3. Advanced Persistent Threat (APT) Таргетированные или целевые кибератаки “Развитая устойчивая угроза”. URL: https://www.tadviser.ru/index.php/Статья:APT_-Таргетированные_или_целевые_атаки (дата обращения: 05.06.2025).
4. Что такое Generative AI. URL: <https://www.sap.com/central-asia-caucasus/products/artificial-intelligence/what-is-generative-ai.html> (дата обращения: 05.06.2025).
5. Сетевые аномалии и их обнаружение: как современные системы анализируют трафик. URL: <https://www.mobilis.ru/news/setevye-anomalii-i-ikh-obnaruzhenie-kak-sovremennye-sistemy-analiziruyut-trafik-opisanie-tekhnologiy/> (дата обращения: 05.06.2025).
6. Чжоу К., Фримэн Д. Машинное обучение и безопасность: защита систем с помощью данных и алгоритмов (пер. с англ. А. В. Снастина). М. : ДМК Пресс, 2020.
7. Корреляция между временными рядами: что может быть проще? URL: <https://habr.com/ru/articles/542638/> (дата обращения: 05.06.2025).
8. Принципы работы генеративных моделей для создания текста и изображений. URL: https://habr.com/ru/companies/fix_price/articles/836206/ (дата обращения: 05.06.2025).
9. Цифровая безопасность как стратегический приоритет: новые вызовы и решения. URL: <https://roscongress.org/materials/tsifrovaya-bezopasnost-kak-strategicheskii-prioritet-novye-vyzovy-i-resheniya/> (дата обращения 05.06.2025).

РАСПРЕДЕЛЕННЫЙ БЛОЧНЫЙ АЛГОРИТМ ВЫЧИСЛЕНИЯ ПОПЕРЕЧНОЙ И ПРОДОЛЬНОЙ СКОРОСТЕЙ ТЕЧЕНИЯ КРОВИ В СОСУДЕ

С. В. Баханович¹⁾, Н. А. Лиходед²⁾

¹⁾ *Институт математики НАН Беларуси,
Минск, Беларусь, bsv@im.bas-net.by*

²⁾ *Белорусский государственный университет,
Минск, Беларусь, likhoded@bsu.by*

Работа посвящена разработке для реализации на многопроцессорных системах с распределенной памятью алгоритма решения блочно-трехдиагональных систем линейных алгебраических уравнений с вложенной структурой блоков с использованием варианта блочного метода Гаусса-Зейделя в неявной форме, обладающего естественным параллелизмом. Такие системы возникают при одновременном получении поперечной и продольной компонент скорости движения крови в кровеносных сосудах.

Ключевые слова: параллельные вычисления; многопроцессорная система с распределенной памятью; блочный метод Гаусса-Зейделя; параллельные вычислительные процессы.

DISTRIBUTED BLOCK ALGORITHM FOR CALCULATING TRANSVERSE AND LONGITUDINAL BLOOD FLOW VELOCITIES IN A VESSEL

S. V. Bakhanovich^{a)}, N. A. Likhoded^{b)}

^{a)} *Institute of Mathematics, National Academy of Science,
Minsk, Belarus, bsv@im.bas-net.by*

^{b)} *Belarusian State University,
Minsk, Belarus, likhoded@bsu.by*

The paper is devoted to the development of algorithms for implementing on distributed memory parallel computers for solving block-tridiagonal systems of linear algebraic equations with a nested structure of blocks using a variant of the block Gauss-Seidel method in an implicit form, which provides natural parallelism. Such systems arise when the transverse and longitudinal components of the blood velocity in blood vessels are simultaneously obtained.

Keywords: parallel computing; distributed memory parallel computer; Gauss-Seidel block method; parallel computing processes.

1. Введение

Пусть блочно-трёхдиагональная система линейных алгебраических уравнений (СЛАУ) с трёхдиагональными диагональными матрицами-блоками имеет вложенную структуру блоков: блоки некоторого порядка $M + 1$ содержат на месте каждого элемента блок порядка 2. Такие СЛАУ возникают на каждом временном слое при численном решении двумерных параболических уравнений гемодинамики для одновременного получения поперечной и продольной компонент скорости движения крови в кровеносных сосудах. Явные блочные алгоритмы Якоби и Гаусса-Зейделя включают обращение диагональных матриц-блоков порядка $M + 1$ и на каждом шаге итерационного процесса включают сначала умножения и суммирование матриц-блоков порядка $M + 1$ на вектор, затем умножения плотных обратных матриц на вектор. Компоненты векторов, на которые умножаются матрицы-блоки, получены на предыдущем шаге или, если они уже уточнены, на этом шаге.

Рассмотрим для решения такой системы блочные алгоритмы Якоби или Гаусса-Зейделя в неявной форме. В этом случае можно, по аналогии с работой [1], после умножения и суммирования матриц-блоков на вектор решить методом матричной прогонки (матрицы порядка 2) трёхдиагональную СЛАУ (каждый коэффициент – блок порядка 2) и обойтись без обращения диагональных матриц-блоков и умножения (на каждом итерационном шаге) плотных обратных матриц на вектор. Алгоритм обладает естественным параллелизмом.

В работе [2] разработан алгоритм блочного метода Гаусса-Зейделя решения на каждом временном слое блочно-трёхдиагональных СЛАУ для получения поперечной компоненты скорости; аналогично можно получить и алгоритм для случая продольной компоненты. В этой работе построен алгоритм решения блочно-трёхдиагональных СЛАУ указанного вида для одновременного получения поперечной и продольной компонент скорости. Алгоритм ориентирован на многопроцессорные системы с распределенной памятью.

2. Блочно-трёхдиагональная СЛАУ с вложенными блоками порядка 2

Пусть N, M – некоторые натуральные числа, $C_0, C_1, \dots, C_N, A_1, \dots, A_N, B_0, B_1, \dots, B_{N-1}$ – вещественные матрицы порядка $2(M + 1)$, $Y_0, Y_1, \dots, Y_N, F_0, F_1, \dots, F_N$ – $2(M + 1)$ -мерные векторы. Рассмотрим блочно-трёхдиагональную систему линейных алгебраических уравнений вида

$$\begin{aligned}
C_0 Y_0 + B_0 Y_1 &= F_0, \\
A_i Y_{i-1} + C_i Y_i + B_i Y_{i+1} &= F_i, \quad i=1, \dots, N-1, \\
A_N Y_{N-1} + C_N Y_N &= F_N.
\end{aligned} \tag{1}$$

Будем рассматривать важный для практического использования в геодинамике случай трёхдиагональных матриц-блоков C_i с вложенной структурой блоков:

$$C_i = \begin{pmatrix} c_0^{C_i} & b_0^{C_i} & 0 & 0 & \dots \\ a_1^{C_i} & c_1^{C_i} & b_1^{C_i} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & a_{M-1}^{C_i} & c_{M-1}^{C_i} & b_{M-1}^{C_i} \\ \dots & 0 & 0 & a_M^{C_i} & c_M^{C_i} \end{pmatrix}, \quad i=0, \dots, N,$$

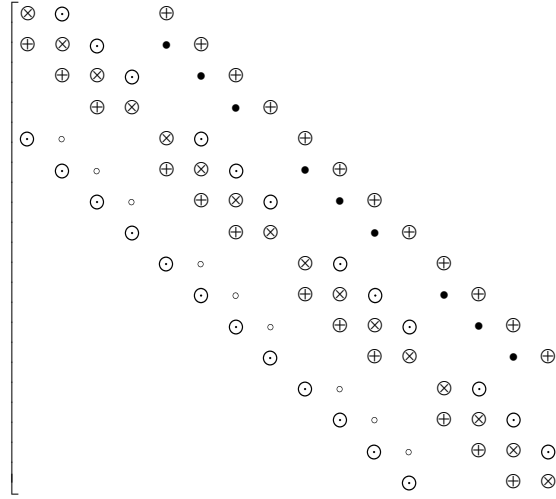
и двухдиагональных матриц-блоков A_i , $i=1, \dots, N$, B_i , $i=0, \dots, N-1$, с вложенной структурой блоков:

$$A_i = \begin{pmatrix} c_0^{A_i} & b_0^{A_i} & 0 & 0 & \dots \\ 0 & c_1^{A_i} & b_1^{A_i} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & 0 & c_{M-1}^{A_i} & b_{M-1}^{A_i} \\ \dots & 0 & 0 & 0 & c_M^{A_i} \end{pmatrix}, \quad B_i = \begin{pmatrix} c_0^{B_i} & 0 & 0 & 0 & \dots \\ a_1^{B_i} & c_1^{B_i} & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & a_{M-1}^{B_i} & c_{M-1}^{B_i} & 0 \\ \dots & 0 & 0 & a_M^{B_i} & c_M^{B_i} \end{pmatrix}.$$

Здесь

$$\begin{aligned}
c_k^{C_i} &= \begin{pmatrix} c_{k,1}^{C_i,u} & c_{k,1}^{C_i,w} \\ c_{k,2}^{C_i,u} & c_{k,2}^{C_i,w} \end{pmatrix}, \quad k=0,1, \dots, M, \\
a_k^{C_i} &= \begin{pmatrix} a_{k,1}^{C_i,u} & a_{k,1}^{C_i,w} \\ 0 & a_{k,2}^{C_i,w} \end{pmatrix}, \quad k=1, \dots, M, \quad b_k^{C_i} = \begin{pmatrix} b_{k,1}^{C_i,u} & 0 \\ b_{k,2}^{C_i,u} & b_{k,2}^{C_i,w} \end{pmatrix}, \quad k=0, \dots, M-1, \\
c_k^{A_i} &= \begin{pmatrix} c_{k,1}^{A_i,u} & 0 \\ c_{k,2}^{A_i,u} & c_{k,2}^{A_i,w} \end{pmatrix}, \quad k=0, \dots, M, \quad b_k^{A_i} = \begin{pmatrix} 0 & 0 \\ b_{k,2}^{A_i,u} & 0 \end{pmatrix}, \quad k=0, \dots, M-1, \\
c_k^{B_i} &= \begin{pmatrix} c_{k,1}^{B_i,u} & c_{k,1}^{B_i,w} \\ 0 & c_{k,2}^{B_i,w} \end{pmatrix}, \quad k=0, \dots, M, \quad a_k^{B_i} = \begin{pmatrix} 0 & a_{k,1}^{B_i,w} \\ 0 & 0 \end{pmatrix}, \quad k=1, \dots, M.
\end{aligned}$$

На рисунке изображён портрет матрицы, согласованной с рассматриваемой моделью гемодинамики, для случая $N = 3, M = 3$.



Структура матрицы системы (1), согласованной с гемодинамикой;

- $\otimes$ – плотные матрицы порядка 2, $\oplus$ – матрицы порядка 2 с нулевым левым нижним элементом, $\ominus$ – матрицы порядка 2 с нулевым правым верхним элементом,
- $\circ$ – матрицы порядка 2 с ненулевым левым нижним элементом, $\bullet$ – матрицы порядка 2 с ненулевым правым верхним элементом

Матрица системы (1) имеет порядок $2(N+1)(M+1)$. Матрица имеет блочно-трёхдиагональную вложенную структуру. На главной диагонали – $N+1$ трёхдиагональных матриц-блоков порядка $2(M+1)$, на поддиагонали и на наддиагонали – N двухдиагональных матриц-блоков порядка $2(M+1)$.

3. Уточнение СЛАУ, согласованной с гемодинамикой

Обозначим компоненты вектора неизвестных:

$$(Y_0, Y_1, \dots, Y_N) = (y_{0,0}, \dots, y_{0,M}, y_{1,0}, \dots, y_{1,M}, \dots, y_{N,0}, \dots, y_{N,M}) = \\ (y_{0,0}^u, y_{0,0}^w, \dots, y_{0,M}^u, y_{0,M}^w, y_{1,0}^u, y_{1,0}^w, \dots, y_{1,M}^u, y_{1,M}^w, \dots, y_{N,0}^u, y_{N,0}^w, \dots, y_{N,M}^u, y_{N,M}^w).$$

Вектор правой части:

$$(F_0, F_1, \dots, F_N) = (f_{0,0}, \dots, f_{0,M}, f_{1,0}, \dots, f_{1,M}, \dots, f_{N,0}, \dots, f_{N,M}) = \\ (f_{0,0}^u, f_{0,0}^w, \dots, f_{0,M}^u, f_{0,M}^w, f_{1,0}^u, f_{1,0}^w, \dots, f_{1,M}^u, f_{1,M}^w, \dots, f_{N,0}^u, f_{N,0}^w, \dots, f_{N,M}^u, f_{N,M}^w).$$

В подробной записи системы (1) имеется $2(N+1)(M+1) = 2NM + 2N + 2M + 2$ уравнений и столько же неизвестных. При фиксированных i и k неизвестные $y_{i,k}^u, y_{i,k}^w$ располагаются при диагональных коэффициентах матрицы $c_k^{C_i}$ двух соседних уравнений. Эта система будет использоваться на практике для одновременного получения приближений $u_{i,k}$, $i = 0, 1, \dots, N$, $k = 1, \dots, M$, поперечной компоненты скорости u и приближений $w_{i,k}$, $i = 1, \dots, N$, $k = 0, 1, \dots, M$, продольной компоненты скорости w . Суммарно это $(N+1)M + N(M+1) = 2NM + N + M$ неизвестных. Для их получения известно и столько же уравнений. Таким образом, в системе (1) есть $N + M + 2$ «лишних» (фиктивных) уравнений и неизвестных.

Неизвестные $y_{0,1}^u, \dots, y_{0,M}^u, y_{1,1}^u, \dots, y_{1,M}^u, \dots, y_{N,1}^u, \dots, y_{N,M}^u$ системы (1) соответствуют $u_{i,k}$, $i = 0, 1, \dots, N$, $k = 1, \dots, M$; неизвестные $y_{0,0}^u, y_{1,0}^u, \dots, y_{N,0}^u$ (это $N+1$ неизвестных, формально здесь $k = 0$) – «лишние», фиктивные. Неизвестные $y_{1,0}^w, \dots, y_{1,M}^w, \dots, y_{N,0}^w, \dots, y_{N,M}^w$ системы (1) соответствуют $w_{i,k}$, $i = 1, \dots, N$, $k = 0, 1, \dots, M$; неизвестные $y_{0,0}^w, \dots, y_{0,M}^w$ (это $M+1$ неизвестных, формально здесь $i = 0$) – «лишние», фиктивные.

Рассмотрим подробную запись системы (1), т. е. рассмотрим запись, в которой $2(N+1)(M+1)$ уравнений и неизвестных. Каждое первое уравнение в каждой блочной строке ($N+1$ уравнений, формально $k = 0$) – «лишнее», фиктивное. Остальные строки с нечётными номерами формируются на основе аппроксимаций уравнения движения для $u_{i,k}$, $i = 0, 1, \dots, N$, $k = 1, \dots, M$. Каждое второе уравнение в блочной строке с номером 0 (это $M+1$ неизвестных, формально $i = 0$) – «лишнее», фиктивное. Остальные строки с чётными номерами формируются на основе аппроксимаций уравнения движения для $w_{i,k}$, $i = 1, \dots, N$, $k = 0, 1, \dots, M$.

Рассмотрим случаи, соответствующие фиктивным («лишним») уравнениям и неизвестным: $i = 0, 1, \dots, N$, $k = 0$, и $i = 0$, $k = 0, 1, \dots, M$.

Рассмотрим первые уравнения в каждой блочной строке подробно записанной системы (1); это $N+1$ уравнений, $i = 0, 1, \dots, N$, $k = 0$. Положим в каждом таком уравнении коэффициент на главной диагонали матрицы системы, т. е. коэффициент при неизвестном $y_{i,0}^u$, равным 1: $c_{0,1}^{C_i,u} = 1$. Остальные коэффициенты и число в правой части этого уравнения положим равными 0 (возможно, этот коэффициент является нулевым и из других соображений). Тогда фиктивные неизвестные являются заведомо нулевыми: $y_{i,0}^u = 0$. В других уравнениях системы коэффициенты при

неизвестных $y_{i,0}^u$, $i = 0, 1, \dots, N$, положим равными 0, если только они не определяются по другим формулам. Поэтому имеем

$$i = 0, 1, \dots, N, k = 0: c_0^{C_i} = \begin{pmatrix} 1 & 0 \\ 0 & c_{0,2}^{C_i,w} \end{pmatrix}, b_0^{C_i} = \begin{pmatrix} 0 & 0 \\ b_{0,2}^{C_i,u} & b_{0,2}^{C_i,w} \end{pmatrix}, f_{i,0} = \begin{pmatrix} 0 \\ f_{i,0}^u \end{pmatrix}.$$

Рассмотрим каждое второе уравнение в блочной строке с номером 0 подробно записанной системы (1); это $M + 1$ уравнений, $i = 0, k = 0, 1, \dots, M$. Положим в каждом таком уравнении коэффициент на главной диагонали матрицы системы, т. е. коэффициент при неизвестном $y_{0,k}^w$, равным 1: $c_{k,2}^{C_0,w} = 1$. Остальные коэффициенты и число в правой части этого уравнения положим равным 0 (возможно, этот коэффициент является нулевым и из других соображений). Тогда заведомо $y_{0,k}^w = 0$. В других уравнениях системы коэффициенты при неизвестных $y_{0,k}^w$, $k = 0, 1, \dots, M$, положим равными 0, если только они не определяются по другим формулам. Поэтому имеем (рассмотренный случай $i = 0, 1, \dots, N, k = 0$ здесь не учтён)

$$i = 0, k = 0, 1, \dots, M:$$

$$c_k^{C_0} = \begin{pmatrix} c_{k,1}^{C_0,u} & 0 \\ 0 & 1 \end{pmatrix}, a_k^{C_0} = \begin{pmatrix} a_{k,1}^{C_0,u} & 0 \\ 0 & 0 \end{pmatrix} (k \neq 0), b_k^{C_0} = \begin{pmatrix} b_{k,1}^{C_0,u} & 0 \\ 0 & 0 \end{pmatrix} (k \neq M),$$

$$c_k^{B_0} = \begin{pmatrix} c_{k,1}^{B_0,u} & 0 \\ 0 & 0 \end{pmatrix}, a_k^{B_0} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} (k \neq 0), f_{0,k} = \begin{pmatrix} f_{0,k}^u \\ 0 \end{pmatrix}.$$

Таким образом, можно не принимать во внимание первые координаты $y_{0,0}^u, y_{1,0}^u, \dots, y_{N,0}^u$ векторов $y_{0,0}, y_{1,0}, \dots, y_{N,0}$ и вторые координаты $y_{0,0}^w, \dots, y_{0,M}^w$ векторов $y_{0,0}, \dots, y_{0,M}$, а также уравнения, в которых эти $y_{i,0}^u$, $i = 0, 1, \dots, N$, и $y_{0,k}^w$, $k = 0, 1, \dots, M$, стоят при коэффициентах на главной диагонали матрицы системы (всего – $N + M + 2$ уравнений и неизвестных). Действительно, эти «лишние» уравнения заведомо являются равенствами, «лишние» неизвестные не влияют на другие уравнения (полученные при моделировании гемодинамики).

Заметим ещё, что из краевых условий следует $u_{0,k} = 0$, $k = 1, \dots, M$. В этом ($i = 0$) случае первые координаты $y_{0,0}^u, y_{0,1}^u, \dots, y_{0,M}^u$ векторов $y_{0,0}, y_{0,1}, \dots, y_{0,M}$ заведомо нулевые. Аналогично, было использовано, что

$u_{N,k} = 0, k = 1, \dots, M$. В этом ($i = N$) случае первые координаты $y_{N,0}''$, $y_{N,1}''$, ..., $y_{N,M}''$ векторов $y_{N,0}$, $y_{N,1}$, ..., $y_{N,M}$ заведомо нулевые. Далее Y_0 вычислять не будем, воспользуемся известным $Y_0 = 0$; Y_N будем вычислять и использовать для контроля вычислений.

4. Блочный алгоритм Гаусса-Зейделя в неявной форме для систем специального вида

Рассмотрим блочный неявный метод Гаусса-Зейделя решения систем вида (1):

$$\begin{aligned} C_0 Y_0^{k+1} &= F_0 - B_0 Y_1^k, \\ C_i Y_i^{k+1} &= F_i - A_i Y_{i-1}^{k+1} - B_i Y_{i+1}^k, \quad i = 1, \dots, N-1, \\ C_N Y_N^{k+1} &= F_N - A_N Y_{N-1}^{k+1}, \\ k &= 0, 1, 2, \dots \end{aligned}$$

$Y_0^0, Y_1^0, \dots, Y_N^0$ – векторы начального приближения.

Запишем псевдокод для решения системы (1); используется вариант метода Гаусса-Зейделя, предложенный для случая без малых блоков в работе [2]:

```
do i = 0, N
   $Y_i = Y_i^0 = \dots$  // выбирается начальное приближение
enddo
 $Y_0 = F_0 = 0$ 
do l = 0, L - 1 // L – заданное или предельное число итераций
  do i = 1, N
     $\Phi = F_i - A_i Y_{i-1}$  //  $\Phi$  – временный массив
    if  $i \neq N$  then  $\Phi = \Phi - B_i Y_{i+1}$ 
    Решить СЛАУ  $C_i Y_i = \Phi$ 
  enddo
enddo(k)
```

(2)

На каждом шаге l требуются умножения блочно-двухдиагональных матриц с вложенными блоками порядка 2 (блоков на главной диагонали – $M + 1$) на вектор. Возникающие на каждом шаге l системы линейных алгебраических уравнений с блочно-трёхдиагональными матрицами (блоки порядка 2, блоков на главной диагонали – $M + 1$) можно решить методом матричной (матрицы порядка 2) прогонки.

5. Параллельный алгоритм для реализации на компьютерах с распределённой памятью

Алгоритм (2), можно использовать для получения реализаций на компьютерах с распределённой памятью. Для MPI+OpenMP-реализации алгоритма требуется организовать параллельные вычислительные процессы, а в каждом процессе – параллельные потоки вычислений.

Пусть P – количество вычислительных процессов, Q^{proc} – количество потоков вычислений в одном процессе (предполагается, что это число одинаковое для всех процессов). Целые числа P и Q^{proc} – параметры MPI+OpenMP-реализации алгоритма. Будем считать, что число $I^{thred} = \frac{N}{P \cdot Q^{proc}}$ является целым; I^{thred} – число итераций цикла с параметром i в одном потоке. Число итераций цикла с параметром i в одном процессе равно $I^{proc} = I^{thred} Q^{proc}$. При фиксированном параметре l внешнего цикла независимо друг от друга можно выполнять итерации цикла с параметром i , определяющего организацию как процессов, так и потоков вычислений. Обмены данными, необходимыми вычислительным процессам, требуются на каждой итерации l .

На практике необходимо произвести так называемое выделение массивов, используемых процессами. В данном случае необходимо выделить векторы Y_i , F_i , матрицы C_i , A_i , B_i и в каждом процессе начать нумерацию по i с нуля.

В p -м процессе вычисляются I^{proc} векторов Y_i , $pI^{proc} + 1 \leq i \leq pI^{proc} + I^{proc}$, при вычислениях используются Y_{i-1} и, если $i \neq N$, Y_{i+1} . В каждом p -м процессе, кроме последнего, $(P-1)$ -го, будем использовать векторы YP_{ip} такие, что $YP_{ip} = Y_i$, $ip = i - pI^{proc}$, $pI^{proc} \leq i \leq pI^{proc} + I^{proc} + 1$ (это $I^{proc} + 2$ векторов). В $(P-1)$ -м процессе будем использовать $I^{proc} + 1$ векторов YP_{ip} : $YP_{ip} = Y_i$, $ip = i - (P-1)I^{proc}$, $N - I^{proc} \leq i \leq N$ (последнее двойное неравенство, с учётом $PI^{proc} = N$, означает $(P-1)I^{proc} \leq i \leq (P-1)I^{proc} + I^{proc}$).

В новых обозначениях p -й процесс при $p < P-1$ отправляет процессу с номером $p+1$ не Y_i , $i = pI^{proc} + I^{proc}$, а YP_{ip} , $ip = I^{proc}$. Соответственно, p -й процесс при $p > 0$ получает от процесса с номером $p-1$ массив, который записывается на место вектора YP_0 . При $p > 0$ процесс с номером p отправляет процессу с номером $p-1$ не Y_i , $i = pI^{proc} + 1$, а YP_1 . Соответственно, p -й процесс при $p < P-1$ получает от процесса с номером $p+1$ массив, который записывается на место вектора YP_{ip} , $ip = I^{proc} + 1$.

При выделении остальных массивов учтём, что в каждом процессе имеется столько же векторов F_i , матриц C_i , A_i , B_i (в $(P-1)$ -м процессе B_N отсутствует), сколько векторов Y_i вычисляются. В p -м процессе, кроме $(P-1)$ -го, при вычислениях используются по I^{proc} векторов F_i , матриц C_i , A_i , B_i , $pI^{proc} + 1 \leq i \leq pI^{proc} + I^{proc}$. В каждом p -м процессе, кроме $(P-1)$ -го, будем теперь использовать векторы FP_{ip} и матрицы CP_{ip} , AP_{ip} , BP_{ip} такие, что $FP_{ip} = F_i$, $CP_{ip} = C_i$, $AP_{ip} = A_i$, $BP_{ip} = B_i$, $ip1 = i - pI^{proc} - 1$, где $pI^{proc} + 1 \leq i \leq pI^{proc} + I^{proc}$. Это по I^{proc} векторов FP_{ip} и матриц CP_{ip} , AP_{ip} , BP_{ip} ; здесь $ip1$ на 1 меньше, чем ip для Y_i . В $(P-1)$ -м процессе будем использовать $I^{proc} - 1$ матриц BP_{ip} : $BP_{ip} = B_i$, $ip1 = i - N + I^{proc} - 1$ ($ip1 = i - (P-1)I^{proc} - 1$ с учётом $P \cdot I^{proc} = N$), где $N - I^{proc} + 1 \leq i \leq N - 1$ (последнее двойное неравенство означает $(P-1)I^{proc} + 1 \leq i \leq (P-1)I^{proc} + I^{proc} - 1$).

Пусть все векторы начального приближения равны нулю и распределены между процессами. Будем ещё предполагать, что векторы F_i , матрицы C_i , A_i , B_i (все распределения указаны в комментариях к псевдокоду) до начала вычислений распределены между процессами.

Для каждого процесса ранга p , $0 \leq p \leq P-1$ (P и Q^{proc} – параметры):

if $p = 0$ then $YP_0 = 0$ // нулевая координата решения равна нулю

$$I^{thred} = \frac{N}{P \cdot Q^{proc}}$$

$$I^{proc} = I^{thred} Q^{proc}$$

do $i^{proc} = 0$, $I^{proc} - 1$

$$ip = i^{proc} + 1$$

$YP_{ip} = 0$ // в процесс ранга p распределены массивы

$$Y_i, \quad pI^{proc} + 1 \leq i \leq pI^{proc} + I^{proc} : YP_{ip} = Y_i, \quad ip = i - pI^{proc}$$

enddo(i^{proc})

if $p < P-1$ send($p+1$; $YP_{I^{proc}}$; $2M$)

if $p > 0$ send($p-1$; YP_1 ; $2M$)

do $l = 0, L-1$

if $p > 0$ receive($p-1$; YP_0 ; $2M$) // полученное записывается как YP_0

if $p < P-1$ receive($p+1$; $YP_{I^{proc}+1}$; $2M$) // записывается как $YP_{I^{proc}+1}$

```

dopar  $q = 0, Q^{proc} - 1$  // независимое выполнение  $Q^{proc}$  потоков
do  $i^{thred} = 0, I^{thred} - 1$ 
 $ip = qI^{thred} + i^{thred} + 1$  //  $ip$ , при которых вычисляются  $YQ_{ip}$ ;
наименьшее  $ip$  в процессе равно 1, наибольшее –
 $(Q^{proc} - 1)I^{thred} + (I^{thred} - 1) + 1 = I^{thred} Q^{proc} = I^{proc}$ 
 $\Phi = FP_{ip-1} - AP_{ip-1}YP_{ip-1}$  // в процесс ранга  $p$  распределены массивы
 $F_i, C_i, A_i, B_i, pI^{proc} + 1 \leq i \leq pI^{proc} + I^{proc}$ :
 $FP_{ip1} = F_i, CP_{ip1} = C_i, AP_{ip1} = A_i, BP_{ip1} = B_i, ip1 = i - pI^{proc} - 1$ 
( $i < (P - 1)I^{proc} + I^{proc}$  для  $B_i$ , так как  $B_N$  не существует)
if  $pI^{proc} + qI^{thred} + i^{thred} + 1 \neq N$  then  $\Phi = \Phi - BP_{ip-1}YP_{ip+1}$ 
Решить СЛАУ  $CP_{ip-1}YP_{ip} = \Phi$ 
enddo( $i^{thred}$ )
enddopar( $q$ )
if  $p < P - 1$  send( $p + 1; YP_{I^{proc}}; 2M$ )
if  $p > 0$  send( $p - 1; YP_1; 2M$ )
enddo( $l$ )
if  $p > 0$  send( $0; YP_{ip}; 1 \leq ip \leq I^{proc}; I^{proc} \times 2M$ ) // передача процессу
ранга 0 вычисленных векторов  $Y_i$ 
if  $p = 0$  receive( $\xi; Y_i, \xi I^{proc} + 1 \leq i \leq \xi I^{proc} + I^{proc}; I^{proc} \times 2M$ ),  $1 \leq \xi \leq P - 1$ 

```

Разработанный алгоритм может быть использован как задел для построения алгоритмов и программ для реализаций двумерных и трёхмерных параболических уравнений гемодинамики. Результаты исследований позволят построить и численно реализовать более адекватную, по сравнению с известной [3], модель кровотока в условиях наличия дефектов в полости сосуда, например тромбоза сосудов.

Библиографические ссылки

1. Yang W., Li K., Li K. A parallel solving method for block-tridiagonal equations on CPU–GPU heterogeneous computing systems // Supercomputing. 2017. Vol. 73, No 5. P. 1760–1781.
2. Параллельная реализация алгоритмов численного решения двумерных параболических уравнений гемодинамики / С. В. Баханович [и др.] // Параллельные вычислительные технологии (ПаВТ'2024): Труды международной научной конференции, Челябинск, 2–4 апреля 2024 г. / Челябинск: Издательский центр ЮУрГУ, 2024. С. 52–64.
3. Конюхов В. М., Храмченков М. Г., Конюхов И. В. Фильтрационная модель и параллельные вычисления характеристик кровотока в системе кровообращения при наличии закупоривающего кольцевого тромба // Программные продукты и системы. 2022. Т. 35. № 3. С. 340–347.

АСИНХРОННОСТЬ И МАСШТАБИРУЕМОСТЬ: ОБРАБОТКА ДАННЫХ В ANDROID-ПРИЛОЖЕНИЯХ

И. Н. Блинов, Т. А. Лядинский

*Белорусский государственный университет,
Минск, Беларусь,
blinov@gmail.com, timofeylyadinsky2@gmail.com*

В работе рассмотрены особенности асинхронности и параллелизма в Android-разработке. Показано различие между конкурентностью и параллельным выполнением задач, а также ограничения однопоточной модели Android. Проанализированы эволюция подходов от потоков и колбэков до RxJava и современного использования Kotlin Coroutines и Flow. Подчеркивается важность асинхронности для обеспечения отзывчивости интерфейса, корректной работы с сетью и базами данных, а также роль WorkManager в долгоживущих процессах. Сделан вывод о том, что асинхронность является фундаментом для построения надежных, масштабируемых и высокопроизводительных Android-приложений.

Ключевые слова: асинхронность; параллелизм; конкурентность; Android; многопоточность; потоки данных.

ASYNCHRONY AND SCALABILITY: DATA PROCESSING IN ANDROID APPLICATIONS

I. N. Blinov, T. A. Lyadinsky

*Belarusian State University,
Minsk, Belarus,
blinov@gmail.com, timofeylyadinsky2@gmail.com*

The article examines the features of asynchrony and parallelism in Android development. It highlights the difference between concurrency and parallel task execution, as well as the limitations of Android's single-threaded model. The evolution of approaches is analyzed, from threads and callbacks to RxJava and the modern use of Kotlin Coroutines and Flow. The importance of asynchrony is emphasized for ensuring UI responsiveness, reliable network and database operations, as well as the role of WorkManager in long-running processes. It is concluded that asynchrony serves as the foundation for building reliable, scalable, and high-performance Android applications.

Keywords: asynchrony; parallelism; concurrency; Android; multithreading; data flows.

1. Введение

В рамках разработки Android приложений и их проектирования явно стоит различать конкурентность и параллелизм. Конкурентность — это архитектурное свойство системы, позволяющее ей обрабатывать несколько задач одновременно, но не обязательно в один и тот же момент времени [1, 2]. В отличие от этого, параллелизм — это способность системы выполнять несколько задач одновременно, в один и тот же момент времени. Для этого требуется наличие нескольких вычислительных ресурсов, например, нескольких ядер процессора, что позволяет физически выполнять несколько операций одновременно. Параллельные операции требуют выделения отдельных потоков, в то время как конкурентные операции могут быть выполнены в одном потоке без его блокировки.

Архитектура Android построена на основе однопоточной модели [1, 3]. В ней весь пользовательский интерфейс управляется и обновляется в главном потоке, Main Thread, который является UI-потоком. Это решение позволяет предотвратить сложные состояния гонки, которые могли бы возникнуть при одновременном доступе нескольких потоков к UI-компонентам. Также это упрощает управление состоянием интерфейса, делая его предсказуемым и надежным. Это в свою очередь накладывает ограничения, так как любая длительная или «блокирующая» операция, такая как сетевой запрос, чтение из базы данных или интенсивные вычисления, выполненная в главном потоке, приведет к его «замораживанию». В результате приложение перестанет реагировать на пользовательский ввод, что в конечном итоге может привести к ошибке «Приложение не отвечает» (ANR - Application Not Responding). Чтобы избежать ANR и обеспечить отзывчивость приложения, все длительные операции должны быть вынесены в отдельные фоновые, или «рабочие» (worker), потоки. Таким образом, однопоточная модель заложила фундаментальную архитектурную потребность в асинхронной и параллельной обработке данных.

2. Теоретические основы

Исторически для асинхронности в Android-разработке были разные подходы. В начале основным инструментом были низкоуровневые Java-потоки и колбеки (AsyncTask, Handler, Looper), но эти подходы сложны в управлении и склонны к утечкам памяти, связанным с жизненными циклами приложения, что явно сказывалось на производительности и устойчивости приложений [1].

Эпоха RxJava, вдохновленная функциональным реактивным программированием, предложила новую парадигму. Она представила

концепцию асинхронных «потоков» данных, которые можно было обрабатывать с помощью декларативных цепочек операторов. Это позволило значительно упростить код по сравнению с колбэками, но ввело свою сложность: большое количество специализированных типов (Observable, Flowable, Single, Maybe, Completable), крутая кривая обучения и непредсказуемые стектрейсы, затрудняющие отладку [1].

На данный момент в разработке применяются Kotlin Coroutines.

Они позволяют разработчикам писать асинхронный код в простом, императивном стиле, который выглядит и читается как обычный синхронный код, но при этом остается неблокирующим и высокоэффективным.

Одним из наиболее значимых преимуществ Coroutines является их «легковесность». Они потребляют значительно меньше ресурсов, чем традиционные потоки JVM [2]. В результате код, который при использовании потоков мог бы легко исчерпать доступную память, может быть выражен с помощью тысяч Coroutines без каких-либо проблем с производительностью или ресурсными ограничениями. Эта эффективность позволяет создавать сложные, высококонкурентные приложения, сохраняя при этом отзывчивость и производительность. Coroutines придерживаются принципа структурированной конкурентности, который является ключевым для создания надежного и поддерживаемого кода.

CoroutineScope [4] — это обязательная область видимости, в которой может быть запущена новая Coroutine. Она определяет жизненный цикл Coroutine, обеспечивая, что фоновая работа привязана к определенному компоненту или области приложения. Job [4] — это дескриптор, который отслеживает жизненный цикл Coroutine и является частью ее контекста. Он позволяет явно управлять Coroutine, например, ожидать ее завершения или отменять. Структурированная конкурентность гарантирует, что родительская область видимости не может завершиться, пока не завершатся все ее дочерние Coroutines. Эта модель предотвращает утечки ресурсов и потерю исключений. Диспетчеры играют ключевую роль в Coroutines, определяя, на каком потоке или пуле потоков будет выполняться код. Это позволяет разработчику явно контролировать, где будет выполняться каждая часть работы, обеспечивая, что длительные операции не блокируют главный поток. Dispatchers.Main: Предназначен для взаимодействия с пользовательским интерфейсом. Он привязан к главному потоку Android, что позволяет безопасно обновлять UI-компоненты. Dispatchers.IO: Оптимизирован для блокирующих операций ввода/вывода (I/O), таких как работа с сетью, чтение или запись файлов, а также взаимодействие с базами данных. Dispatchers.Default: используется для ресурсоемких вычислительных задач, которые требуют больших ресурсов процессора. Пул потоков этого диспетчера ограничен количеством ядер CPU [3, 4].

Ключевое слово `suspend` является краеугольным камнем `Coroutines` [2, 4]. Оно помечает функцию, которая может «приостановить» выполнение `Coroutine`, не блокируя при этом поток операционной системы. Это позволяет выполнять длительные операции, такие как вызов сетевого API или задержка, без блокировок потоков. `withContext` является одним из наиболее мощных и часто используемых инструментов в `Coroutines`. Он позволяет сменить контекст выполнения `Coroutine` на другой диспетчер. Например, можно находиться в `Dispatchers.Main`, затем переключиться на `Dispatchers.IO` для выполнения сетевого запроса, а после его завершения автоматически вернуться в `Dispatchers.Main` для обновления UI. Это делает код невероятно читаемым, так как асинхронные операции выглядят как последовательные вызовы.

Современные приложения Android должны следовать принципам архитектуры, которые обеспечивают масштабируемость, надежность и простоту тестирования. Ключевым элементом такой архитектуры является реактивная модель, основанная на однонаправленном потоке данных (`Unidirectional Data Flow`, `UDF`) [3]. В этой модели данные перемещаются от источника истины (`Single Source of Truth`, `SSOT`) к слою UI, в то время как события, инициированные пользователем, поднимаются вверх по иерархии, защищая целостность данных и упрощая управление состоянием.

`Kotlin Flow` — это мощный инструмент для реализации этой модели. Он предоставляет структурированный и `coroutine`-ориентированный способ обработки асинхронных потоков данных. `Flow` является «холодным потоком» (`cold stream`), что означает, что он начинает отдавать значения только тогда, когда у него появляется коллектор. Это позволяет эффективно преобразовывать и обрабатывать данные с помощью богатого набора операторов. Такой подход особенно полезен для запросов в базу данных или сетевых вызовов: данные не тратят ресурсы впустую, пока их никто не слушает.

Для управления состоянием UI и событиями в реактивной архитектуре используются `StateFlow` и `SharedFlow` [4]. `StateFlow` — это `Flow`, который хранит состояние. Он идеально подходит для представления состояния UI (`Loading`, `Success`, `Error`) в `ViewModel`, так как он всегда имеет начальное значение и испускает последние обновления состояния своим коллекторам.

`SharedFlow` — это более гибкий «горячий поток» (`hot flow`), который используется для передачи событий множеству коллекторов. Разделение состояния и событий на эти два типа потоков является фундаментальным архитектурным решением, которое приводит к созданию более предсказуемого и надежного UI-слоя, что, в свою очередь, является ключом к созданию масштабируемых и поддерживаемых приложений.

Благодаря этому сетевые операции, операции с базами данных и фоновые операции могут эффективно применяться в рамках мобильного приложения и операционной системы.

Благодаря coroutines такие сетевые вызовы (например, через сетевую библиотеку Retrofit) могут выполняться асинхронно и безопасно для главного потока. Использование suspend-функций позволяет описывать сетевую логику в том же стиле, что и синхронный код, но при этом запросы автоматически переключаются на оптимизированный пул потоков (Dispatchers.IO) [3].

Ключевая особенность — возможность отмены запросов при изменении жизненного цикла. Если пользователь ушёл с экрана или приложение было закрыто, coroutine прерывается, и операция завершает работу. Это критично для мобильных приложений, где важно сохранять ресурсы и не держать «висящие» вызовы.

Мобильное устройство ограничено по количеству оперативной памяти, вычислительной мощности и времени работы от батареи. Поэтому масштабируемость в Android неразрывно связана с грамотным управлением ресурсами. Coroutines предоставляют разработчику возможность контролировать область видимости задач и их жизненный цикл, что предотвращает утечки памяти и снижает энергопотребление. В сочетании с библиотекой Lifecycle достигается автоматическая отмена coroutines при уничтожении компонента, что особенно важно в условиях активного переключения экранов и сложной навигации [3].

Современные Android-приложения часто зависят от большого числа сетевых вызовов. Масштабируемость здесь достигается благодаря применению таких инструментов, как OkHttp и Retrofit, которые в связке с coroutines и Flow позволяют создавать эффективные конвейеры обработки запросов. Параллельное выполнение нескольких запросов осуществляется через coroutineScope и оператор async/await, а также через комбинирование Flow. При этом важно учитывать оптимизацию под разные сети (Wi-Fi, LTE, 5G) и сценарии с ограниченным доступом к интернету [3].

Базы данных в Android, в частности SQLite через Room, изначально ограничены: любой доступ к ним в главном потоке приводит к блокировке UI. Coroutines решают эту проблему: операции с базой выполняются через suspend-функции или в реактивной форме с помощью Flow. В рамках масштабируемости это играет важную роль. Например, изменение в таблице автоматически испускается как новое значение потока и мгновенно отражается в UI. Такой подход исключает необходимость опроса базы и повышает отзывчивость интерфейса. Для сложных сценариев применяется withTransaction, который гарантирует атомарность набора операций. Это позволяет сохранять целостность данных даже при высоком уровне конкурентности [3, 4].

Асинхронность в мобильных приложениях не сводится только к работе с экраном или сетевыми запросами. Часто возникают задачи, которые должны выполняться независимо от жизненного цикла интерфейса: синхронизация данных, периодические обновления или обработка больших файлов [5].

Для таких сценариев используется WorkManager. В отличие от coroutines, ограниченных областью действия определённого Scope, WorkManager гарантирует выполнение фоновых задач даже при закрытии приложения или перезапуске устройства. При этом он интегрируется с coroutines, что позволяет описывать фоновые операции в привычной форме suspend-функций [5].

WorkManager играет ключевую роль в масштабируемости долгосрочных процессов, таких как синхронизация данных или периодическая загрузка обновлений. Он автоматически учитывает состояние устройства (заряд батареи, доступ к сети, режим энергосбережения) и оптимизирует выполнение задач. Таким образом, приложение сохраняет предсказуемость поведения даже при больших объемах фоновой работы [3, 5].

В результате разработчик получает единый подход к асинхронности: coroutines обеспечивают удобную работу внутри UI, а WorkManager – надежное выполнение долгоживущих процессов в фоновом слое приложения.

3. Заключение

Таким образом, асинхронность выступает одним из ключевых инструментов современной разработки Android-приложений. Она позволяет эффективно распределять нагрузку между потоками выполнения, исключая блокировки основного пользовательского интерфейса и обеспечивая плавность работы приложения даже при выполнении ресурсоемких операций. Благодаря лёгкости и простоте API, разработчик получает возможность проектировать архитектуру приложений быстрее и гибче, опираясь на принципы реактивного программирования.

Реактивная архитектура в данном контексте подразумевает построение систем, где данные обрабатываются в режиме реального времени, а изменения в одном компоненте мгновенно отражаются на других. Это не только повышает удобство для конечного пользователя, но и облегчает сопровождение и масштабирование приложений. В сочетании с возможностями параллельной и распределенной обработки данных асинхронность становится фундаментом для создания высокопроизводительных, надежных и устойчивых Android-приложений, готовых к интеграции с современными облачными экосистемами.

Библиографические ссылки

1. *Babić F., Kordić L., Srivastava N.* Kotlin Coroutines by Tutorials. Razeware LLC, 2022.
2. Kotlin coroutines: design and implementation / R. Elizarov [et al.] // Onward! 2021: Proceedings of the 2021 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software, Chicago, October 20–22, 2021 / eds.: W. De Meuter, E. Baniassad. New York : Association for Computing Machinery, 2021. P. 68–84. URL: https://www.researchgate.net/publication/355361443_Kotlin_coroutines_design_and_implementation (date of access: 01.09.2025).
3. Android Mobile App Developer Tools – Android Developers. URL: <https://developer.android.com> (date of access: 01.09.2025).
4. Kotlin docs. URL: <https://kotlinlang.org/docs/coroutines-basics.html> (date of access: 01.09.2025).
5. Work Manager – Android. URL: <https://medium.com/@appdevinsights/work-manager-android-6ea8daad56ee> (date of access: 01.09.2025).

АНАЛИЗ ЭФФЕКТИВНОСТИ ЦВЕТОВЫХ ПРОСТРАНСТВ CIELAB И RGB ДЛЯ ЗАДАЧ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ ВОДОРΟΣЛЕЙ

И. Д. Климович, О. В. Недзьведь

Белорусский государственный университет,
Минск, Беларусь, posket.klimovich@gmail.com

Разработана и апробирована методика автоматизированного сравнительного анализа эффективности цветковых пространств для задач анализа изображений водорослей. Методика применена для оценки качества сегментации и морфометрии изображений микроводоросли *Chlorella*. Выполнено сравнение алгоритмов на основе пространств CIELAB и RGB.

Ключевые слова: *Chlorella*; Lab; RGB; цветковые пространства; сегментация.

PERFORMANCE ANALYSIS OF CIELAB AND RGB COLOR SPACES FOR ALGAE IMAGE SEGMENTATION TASKS

I. D. Klimovich, O. V. Nedzvedz

Belarusian State University,
Minsk, Belarus, posket.klimovich@gmail.com

A method for automated comparative analysis of the efficiency of color spaces for the tasks of analyzing algae images has been developed and tested. The method has been applied to assess the quality of segmentation and morphometry of images of the microalga *Chlorella*. A comparison of algorithms based on CIELAB and RGB spaces has been performed.

Keywords: *Chlorella*; Lab; RGB; color spaces; segmentation.

1. Введение

Массовое культивирование водорослей рода *Chlorella* для получения биотоплива, пищевых добавок и фармацевтических препаратов требует точного и оперативного контроля параметров культуры. Одним из ключевых аспектов такого контроля является морфометрический анализ, включающий подсчет клеток и определение их размеров. Традиционный ручной анализ этих параметров является весьма трудоемким и субъективным, что стимулирует развитие автоматизированных систем компьютерного анализа изображений [1].

Ключевым этапом автоматизированного анализа является сегментация. Эффективность сегментации напрямую определяет точность всех последующих морфометрических измерений. Одним из фундаментальных выборов при построении алгоритма сегментации является цветовое пространство, в котором проводятся вычисления. Наиболее распространенным и простым в обработке является RGB-пространство. Однако его серьезным недостатком является сильная зависимость от условий освещения [2]. Альтернативой выступает перцептивное цветовое пространство CIELAB (далее – Lab), в котором яркость (канал L) отделена от цветовой информации (каналы a и b), что обеспечивает лучшую устойчивость к изменчивости условий съемки [3, 4].

Несмотря на очевидные теоретические преимущества, вопрос о том, насколько выбор между Lab и RGB влияет на точность и эффективность конкретных алгоритмов сегментации и морфометрии при работе с изображениями микроводорослей, остается открытым. В литературе отсутствуют прямые сравнительные исследования, которые бы количественно оценили это влияние на репрезентативных выборках данных.

Целью данной работы является разработка формализованной методики для сравнительного анализа эффективности цветковых пространств Lab и RGB в пайплайне обработки изображений *Chlorella*.

2. Материалы и методы исследования

Для сравнительной оценки эффективности алгоритмов сегментации на основе цветковых пространств Lab и RGB обрабатывалось 50 изображений микроводоросли *Chlorella* из открытого датасета [1]. Ручные измерения были проведены с помощью веб-софта ImageJ.JS на том же наборе данных и далее использовались в качестве эталона. Для автоматизированного анализа был разработан скрипт на языке Python с использованием библиотек OpenCV, scikit-image, scipy, numpy и pandas. Обработка данных и расчет метрик производились в среде Jupyter Notebook.

Разработанная методика представляет собой формализованный пайплайн обработки изображений, состоящий из следующих универсальных этапов: предобработка, сегментация и постобработка изображений. Предобработка осуществлялась для обоих пространств улучшением контраста с применением метода CLAHE (clipLimit=3,0) и Гауссова размытия (kernel=7×7) для уменьшения шума. Пороговые значения для идентификации клеток были установлены как: минимальная площадь клетки – 5 px² и максимальная – 300 px².

Далее была проведена сегментация, которая в пространстве Lab (рис. 1) включала выделение канала a и бинаризацию методом Оцу, а в

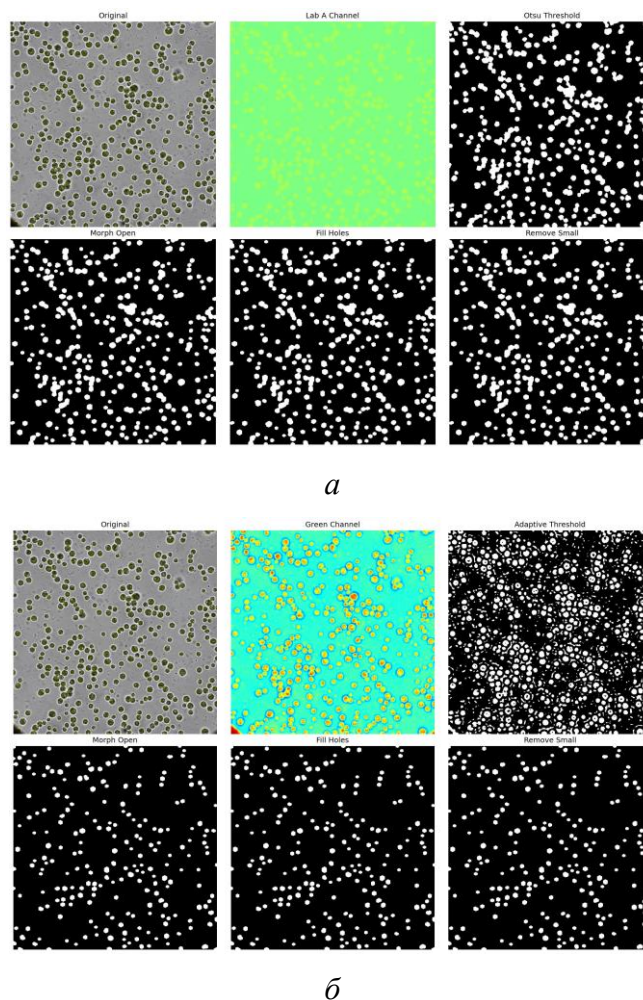


Рис. 1. Визуализация этапов сегментации и постобработки изображений *Chlorella*:
 a – в пространстве Lab; b – в пространстве RGB

RGB – выделение G-канала и применение адаптивной бинаризации с использованием скользящего окна для учета локальных изменений освещенности.

Постобработка (см. рис. 1) для обоих методов осуществлялась за счет морфологической коррекции (к бинарной маске применяются операции открытия и закрытия для удаления шума и заполнения мелких разрывов в контурах клеток) и анализа связанных компонентов, который выполняет поиск всех связанных областей (клеток) с фильтрацией по размеру.

Для каждого изображения рассчитывались морфометрические параметры: среднее количество клеток, средняя площадь клеток и средний диаметр с соответствующим указанием стандартного отклонения.

Для оценки эффективности методов использовались сравнительные метрики (для сегментации – полнота, для морфометрии – средняя абсолютная ошибка (MAPE)) [5]. Полноту сегментации определяли как отношение количества определенных программой клеток к числу клеток, определенное ручным подсчетом, выраженное в процентах.

Суммарная методика представлена схематично на рис. 2.

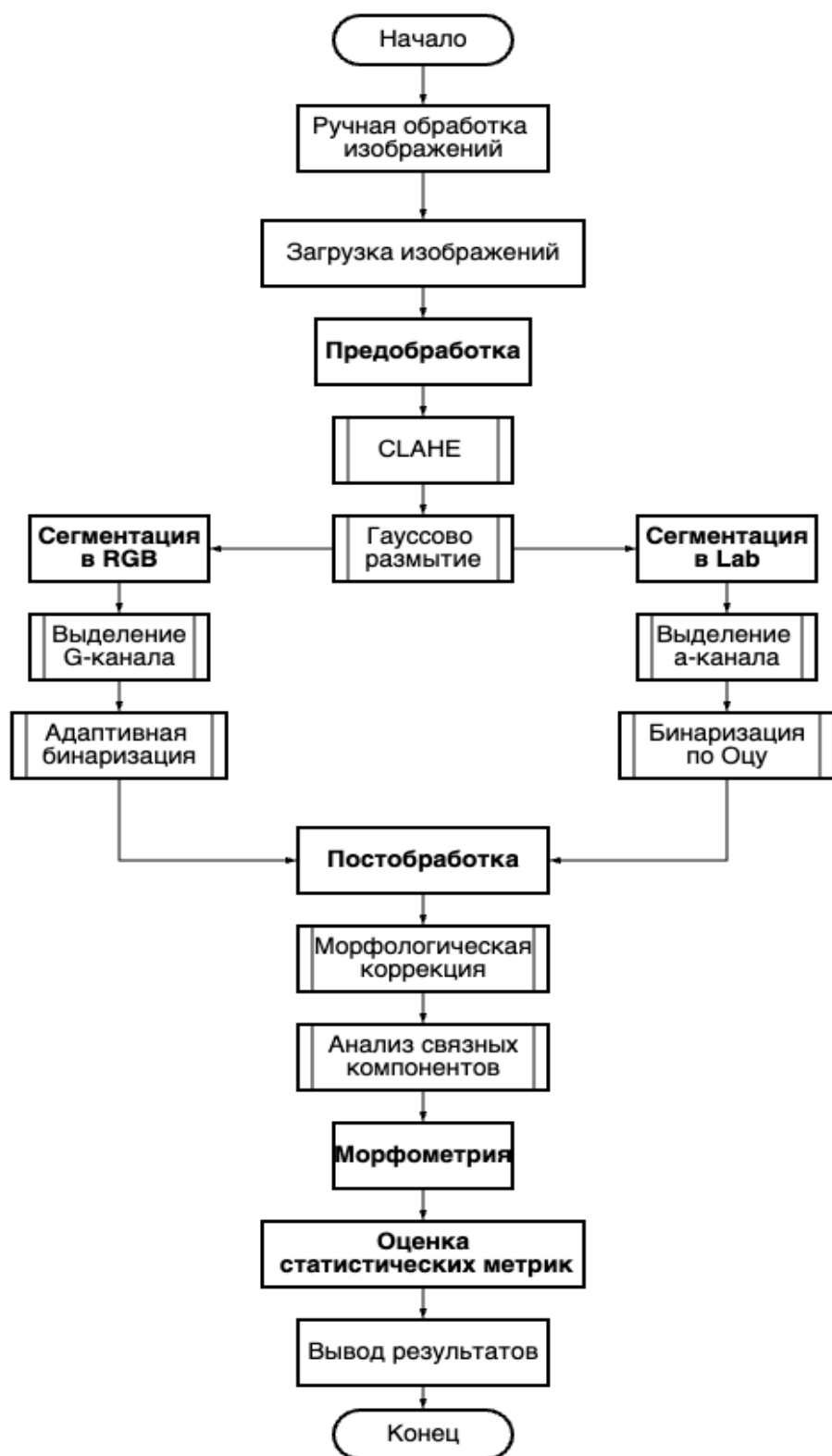


Рис. 2. Основные этапы сравнительной методики методов анализа изображений *Chlorella* в RGB- и Lab-пространствах

3. Результаты и их обсуждение

Проведенное исследование выявило существенные различия в эффективности алгоритмов сегментации и морфометрии в обоих пространствах (табл. 1, 2).

Таблица 1

Результаты морфометрического анализа изображений *Chlorella*

Характеристика	Ручной подсчет	Lab	RGB
Среднее количество клеток	372,1±24,2	289,4±56,8	251,0±51,6
Средняя площадь клеток, рх	111,2±12,3	107,8±6,0	76,5±4,2
Средний диаметр клеток, рх	19,5±0,71	11,4±0,3	9,7±0,2

Таблица 2

Метрики эффективности сегментации и морфометрии

Метрика	Lab	RGB
Сегментация		
Полнота, %	77,8	67,5
Морфометрия		
MAPE по площади, %	2,9	31,1
MAPE по диаметру, %	41,5	50,2

Наиболее значимое различие наблюдается в эффективности обнаружения клеток. Алгоритм на основе Lab-пространства продемонстрировал явное превосходство над RGB, обеспечив полноту сегментации на 10,3% выше. Это может объясняться особенностями цветовой модели Lab: разделение информации о яркости и цвете обеспечивает оптимальную контрастность для выделения зеленых клеток *Chlorella* на фоне. Однако высокие значения стандартных отклонений указывают на нестабильность работы обоих алгоритмов.

Для площади клеток алгоритм на основе Lab показал исключительно высокую точность со значением MAPE = 2,9%. Это свидетельствует о том, что данный метод не только корректно обнаруживает клетки, но и с высокой точностью определяет их границы. В отличие от этого, RGB-метод показал на порядок большую ошибку (MAPE = 31,1%).

Диаметр клеток оба алгоритма определили некорректно, значительно занижив показатели относительно ручного измерения. Но Lab-метод

продемонстрировал относительное преимущество (MAPE = 41,5%). Данное систематическое занижение линейных размеров возможно связано с фундаментальным ограничением использованных алгоритмов сегментации, которые «съедают» периферийные области клеток. Это указывает на необходимость разработки и применения постобработки (например, морфологических операций расширения) или использования более совершенных алгоритмов (например, на основе U-Net) для коррекции границ объектов.

4. Заключение

В работе разработана и успешно апробирована универсальная методика для сравнительного анализа эффективности цветковых пространств в задачах сегментации и морфометрического анализа изображений. Методика формализует процесс обработки данных от предобработки до расчета объективных метрик качества.

Применение методики к задаче анализа изображений микроводоросли *Chlorella* показало статистически значимое преимущество цветкового пространства CIELAB над RGB по ключевым метрикам: полноте обнаружения объектов и точности определения площади. Алгоритм на основе пространства CIELAB показал превосходство над RGB, продемонстрировав на 10,3% более высокую полноту обнаружения объектов и ошибку определения площади всего в 2,9%.

Основным результатом использования методики является не только вывод о предпочтительности пространства Lab для данного типа данных, но и выявление систематической ошибки алгоритмов в определении линейных размеров. Это демонстрирует диагностическую ценность методики для выявления слабых мест в пайплайнах обработки изображений.

Библиографические ссылки

1. Chong J. W. R. Artificial intelligence-driven microalgae autotrophic batch cultivation: A comparative study of machine and deep learning-based image classification models // *Algal Research*. 2024. Vol. 79. Article no. 103400.
2. Zeileis A., Hornik K., Murrell P. Escaping RGBland: Selecting colors for statistical graphics // *Computational Statistics & Data Analysis*. 2009. Vol. 53, iss. 9. P. 3259–3270.
3. Sharma G., Wu W., Dalal E. N. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations // *Color Research & Application*. 2005. Vol. 30, iss. 1. P. 21–30.
4. Gonzalez R. C., Woods R. E. *Digital Image Processing*. 4th ed. Pearson, 2018.
5. Haslwanter T. *An introduction to statistics with Python. With applications in the life sciences*. Switzerland : Springer International Publishing, 2016.

OPENMP-РЕАЛИЗАЦИЯ БЛОЧНОГО АЛГОРИТМА ВЫЧИСЛЕНИЯ ПОПЕРЕЧНОЙ И ПРОДОЛЬНОЙ СКОРОСТЕЙ ТЕЧЕНИЯ КРОВИ В СОСУДЕ

Н. А. Лиходед¹⁾, Э. Е. Малиев²⁾

¹⁾ *Белорусский государственный университет,
Минск, Беларусь, likhoded@bsu.by*

²⁾ *Институт математики НАН Беларуси,
Минск, Беларусь, maliev@im.bas-net.by*

Работа посвящена разработке и реализации на многоядерных системах с общей памятью алгоритма решения блочно-трехдиагональных систем линейных алгебраических уравнений с вложенной структурой блоков. Такие системы возникают при одновременном получении поперечной и продольной компонент скорости движения крови в кровеносных сосудах. Используется красно-черное упорядочение модифицированного блочного метода Гаусса-Зейделя в неявной форме, обладающего естественным параллелизмом.

Ключевые слова: параллельные вычисления; многопроцессорная система с общей памятью; блочный метод Гаусса-Зейделя; красно-черное упорядочение; параллельные потоки вычислений.

OpenMP IMPLEMENTATION OF A BLOCK ALGORITHM FOR CALCULATING TRANSVERSE AND LONGITUDINAL BLOOD FLOW VELOCITIES IN A VESSEL

N. A. Likhoded^{a)}, E. E. Maliev^{b)}

^{a)} *Belarusian State University,
Minsk, Belarus, likhoded@bsu.by*

^{b)} *Institute of Mathematics, National Academy of Science,
Minsk, Belarus, maliev@im.bas-net.by*

The paper is devoted to the development and implementation on multi-core shared-memory systems of an algorithm for solving block-tridiagonal systems of linear algebraic equations with a nested block structure. Such systems arise from the simultaneous determination of the transverse and longitudinal components of blood flow velocity in blood vessels. The red-black ordering of the modified Gauss-Seidel block method is used in an implicit form, which exhibits natural parallelism.

Keywords: parallel computing; shared-memory parallel computer; Gauss-Seidel block method; red-black ordering; parallel threads.

1. Введение

Настоящая работа во многом опирается на материал «Распределенный блочный алгоритм вычисления поперечной и продольной скоростей течения крови в сосуде» этого международного научного конгресса (см. с. 399–408). Здесь рассматриваются такие же блочно-трёхдиагональная система линейных алгебраических уравнений (СЛАУ) с трёхдиагональными диагональными матрицами-блоками вложенной структуры, метод решения, вспомогательные алгоритмы, направления дальнейших исследований, ссылки на литературу. Поэтому обозначения и всё перечисленное не будем дублировать, их надо смотреть в указанном материале. Новым в этой работе являются: алгоритмы решения СЛАУ, их ориентированность на многоядерные CPU, OpenMP-реализации алгоритмов, вычислительные эксперименты.

2. Блочный алгоритм Гаусса-Зейделя в неявной форме для систем специального вида и с использованием красно-черного упорядочения

Красно-черное упорядочение сводит алгоритм зейделевского типа к двухфазной процедуре типа Якоби с сохранением точности и числа операций исходного алгоритма. Красно-черное упорядочение позволяет сохранить зейделевский тип вычислений при переходе к параллельной версии алгоритма. Без применения красно-черного упорядочения параллельные версии алгоритма несколько теряют точность вычислений, так как часть вычислений зейделевского типа заменяется вычислениями типа Якоби – особенность, возникающая при организации параллельных потоков вычислений.

Рассмотрим алгоритм, основанный на методе Гаусса-Зейделя в неявной форме. Если в этом алгоритме использовать красно-черное упорядочение, то получим

```
do  $i = 0, N$ 
   $Y_i = Y_i^0 = \dots$  // выбирается начальное приближение,  $Y_0 = F_0 = 0$ 
enddo
do  $l = 0, L - 1$  //  $L$  – заданное или предельное число итераций
  do  $i = 2, N, \text{step}=2$ 
     $\Phi = F_i - A_i Y_{i-1}$  // используется  $Y_i$  с нечётным  $i$ 
    if  $i \neq N$  then  $\Phi = \Phi - B_i Y_{i+1}$  // используется  $Y_i$  с нечётным  $i$ 
```

```

    Решить СЛАУ  $C_i Y_i = \Phi$ 
enddo
do  $i = 1, N$ , step=2
     $\Phi = F_i - A_i Y_{i-1}$  // используется  $Y_i$  с чётным  $i$ 
    if  $i \neq N$  then  $\Phi = \Phi - B_i Y_{i+1}$  // используется  $Y_i$  с чётным  $i$ 
    Решить СЛАУ  $C_i Y_i = \Phi$ 
enddo
enddo(k)

```

На каждом шаге l требуются умножения блочно-двухдиагональных матриц (с вложенными блоками второго порядка) на вектор. Возникающие системы линейных алгебраических уравнений с блочно-трёхдиагональными матрицами (блоки второго порядка) можно решить методом матричной (матрицы второго порядка) прогонки.

3. Псевдокод параллельного алгоритма для реализации на многоядерных компьютерах с общей памятью

Пусть $Ax = f$ – СЛАУ порядка n . Для достижения критерия остановки итерационного процесса требуется на l -м итерационном шаге проанализировать норму вектора $r^{l+1} = f - Ax^{l+1}$ невязки системы: $\|r^{j+1}\|_\infty = \max_{1 \leq i \leq n} |r_i^{j+1}|$ (используется максимум-норма). Потоки вычислений будем задавать циклом с параметром i .

Алгоритм. Блочный неявный метод Гаусса-Зейделя решения СЛАУ, ориентированной на гемодинамику.

Вход:

N, M – параметры СЛАУ;

ε – предельное значение нормы невязка, характеризует точность алгоритма;

двумерный массив, каждый элемент которого есть вектор размерности 2 – векторы начального приближения $Y_0^0, Y_1^0, \dots, Y_N^0$:

$$y(i, k), 0 \leq i \leq N, 0 \leq k \leq M // y(0, k) = y_{0,k} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, y(i, k) = y_{i,k} = \begin{pmatrix} y_{i,k}^u \\ y_{i,k}^w \end{pmatrix};$$

двумерные массивы, элементы которых есть матрицы порядка 2:

$$cC(i,k) = \begin{pmatrix} cC1u(i,k) & cC1w(i,k) \\ cC2u(i,k) & cC2w(i,k) \end{pmatrix}, \quad 1 \leq i \leq N, 0 \leq k \leq M$$

$$// \quad cC(i,k) = c_k^{C_i} = \begin{pmatrix} c_{k,1}^{C_i,u} & c_{k,1}^{C_i,w} \\ c_{k,2}^{C_i,u} & c_{k,2}^{C_i,w} \end{pmatrix}$$

$$aC(i,k) = \begin{pmatrix} aC1u(i,k) & aC1w(i,k) \\ 0 & aC2w(i,k) \end{pmatrix}, \quad 1 \leq i \leq N, 1 \leq k \leq M$$

$$// \quad aC(i,k) = a_k^{C_i} = \begin{pmatrix} a_{k,1}^{C_i,u} & a_{k,1}^{C_i,w} \\ 0 & a_{k,2}^{C_i,w} \end{pmatrix}$$

$$bC(i,k) = \begin{pmatrix} bC1u(i,k) & 0 \\ bC2u(i,k) & bC2w(i,k) \end{pmatrix}, \quad 1 \leq i \leq N, 0 \leq k \leq M-1$$

$$// \quad bC(i,k) = b_k^{C_i} = \begin{pmatrix} b_{k,1}^{C_i,u} & 0 \\ b_{k,2}^{C_i,u} & b_{k,2}^{C_i,w} \end{pmatrix}$$

$$cA(i,k) = \begin{pmatrix} cA1u(i,k) & 0 \\ cA2u(i,k) & cA2w(i,k) \end{pmatrix}, \quad 1 \leq i \leq N, 0 \leq k \leq M$$

$$// \quad cA(i,k) = c_k^{A_i} = \begin{pmatrix} c_{k,1}^{A_i,u} & 0 \\ c_{k,2}^{A_i,u} & c_{k,2}^{A_i,w} \end{pmatrix}$$

$$bA(i,k) = \begin{pmatrix} 0 & bA1w(i,k) \\ 0 & 0 \end{pmatrix}, \quad 1 \leq i \leq N, 0 \leq k \leq M-1$$

$$// \quad bA(i,k) = b_k^{A_i} = \begin{pmatrix} 0 & b_{k,1}^{A_i,w} \\ 0 & 0 \end{pmatrix}$$

$$cB(i,k) = \begin{pmatrix} cB1u(i,k) & cB1w(i,k) \\ 0 & cB2w(i,k) \end{pmatrix}, \quad 1 \leq i \leq N-1, 0 \leq k \leq M$$

$$// \quad cB(i,k) = c_k^{B_i} = \begin{pmatrix} c_{k,1}^{B_i,u} & c_{k,1}^{B_i,w} \\ 0 & c_{k,2}^{B_i,w} \end{pmatrix}$$

$$aB(i,k) = \begin{pmatrix} 0 & aB1w(i,k) \\ 0 & 0 \end{pmatrix}, \quad 1 \leq i \leq N-1, 1 \leq k \leq M$$

$$// \quad aB(i,k) = a_k^{B_i} = \begin{pmatrix} 0 & a_{k,1}^{B_i,w} \\ 0 & 0 \end{pmatrix};$$

двумерный массив, элементы которого есть векторы размерности 2:

$$fslae(i,k) = \begin{pmatrix} fu(i,k) \\ fw(i,k) \end{pmatrix}, \quad 1 \leq i \leq N, \quad 0 \leq k \leq M \quad // \quad fslae(0,k) = f_{0,k} = \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

подпрограмма умножения матрицы порядка 2 на двумерный вектор;
подпрограмма матричной прогонки для решения системы, все коэффициенты матрицы которой – матрицы-блоки порядка 2, а каждый элемент вектора неизвестных и вектора правой части – двумерный подвектор.

Выход:

двумерный массив, элементы которого есть векторы размерности 2:

$$y(i, k), \quad 0 \leq i \leq N, \quad 0 \leq k \leq M \quad // \quad y(i, k) = y_{i,k} = \begin{pmatrix} y_{i,k}^u \\ y_{i,k}^w \end{pmatrix}.$$

do $i = 0, N$

$Y_i = Y_i^0 = \dots$ // выбирается начальное приближение, $Y_0 = F_0 = 0$

enddo

do $l = 0, 1, 2, \dots$ (пока не достигнут критерий остановки $\|r^{j+1}\| < \varepsilon$):

// итерации следующего цикла распределяются между потоками

dopar $i = 2, N$, step=2

// начало получения вектора $\Phi = F_i - A_i Y_{i-1} - B_i Y_{i+1}$, если $i \neq N$ или
вектора $\Phi = F_i - A_i Y_{i-1}$, если $i = N$; используется Y_i с нечётным i

do $k = 0, M - 1$

$phi(k) = fslae(i, k) - cA(i, k)y(i-1, k) - bA(i, k)y(i-1, k+1)$

// ϕ_k – k -й элемент Φ ; умножения – матрично-векторные

enddo(k)

$phi(M) = fslae(i, M) - cA(i, M)y(i-1, M)$

if $i \neq N$ do

$phi(0) = phi(0) - cB(i, 0)y(i+1, 0)$

do $k = 1, M$

$phi(k) = phi(k) - cB(i, k)y(i+1, k) - aB(i, k)y(i+1, k-1)$

enddo(k)

endif($i \neq N$)

// конец получения вектора $\Phi = F_i - A_i Y_{i-1} - B_i Y_{i+1}$, если $i \neq N$

или вектора $\Phi = F_i - A_i Y_{i-1}$, если $i = N$

Решение СЛАУ $C_i Y_i = \Phi$ методом матричной прогонки

enddopar($i = 2, N$, step=2)

```

// итерации следующего цикла распределяются между потоками
dopar i = 1, N, step=2
операторы тела цикла dopar i = 2, N, step=2
// используется  $Y_i$  с чётным  $i$ 

enddopar(i = 1, N, step=2)
enddo(l)

```

4. Вычислительные эксперименты

Для вычислительных экспериментов использовался компьютер с общей памятью с 12 физическими ядрами (логических потоков – 20). Вычисления проводились на модельном примере при $N = 1000$, $M = 3000$ (рис. 1) и $N = 3000$, $M = 3000$ (рис. 2), $\varepsilon = 10^{-6}$. Сравнивалось время выполнения четырёх алгоритмов. На рисунках синий цвет графика соответствует последовательному алгоритму (SeqGS), оранжевый – последовательному алгоритму с использованием красно-черного упорядочения (SeqRBGS),

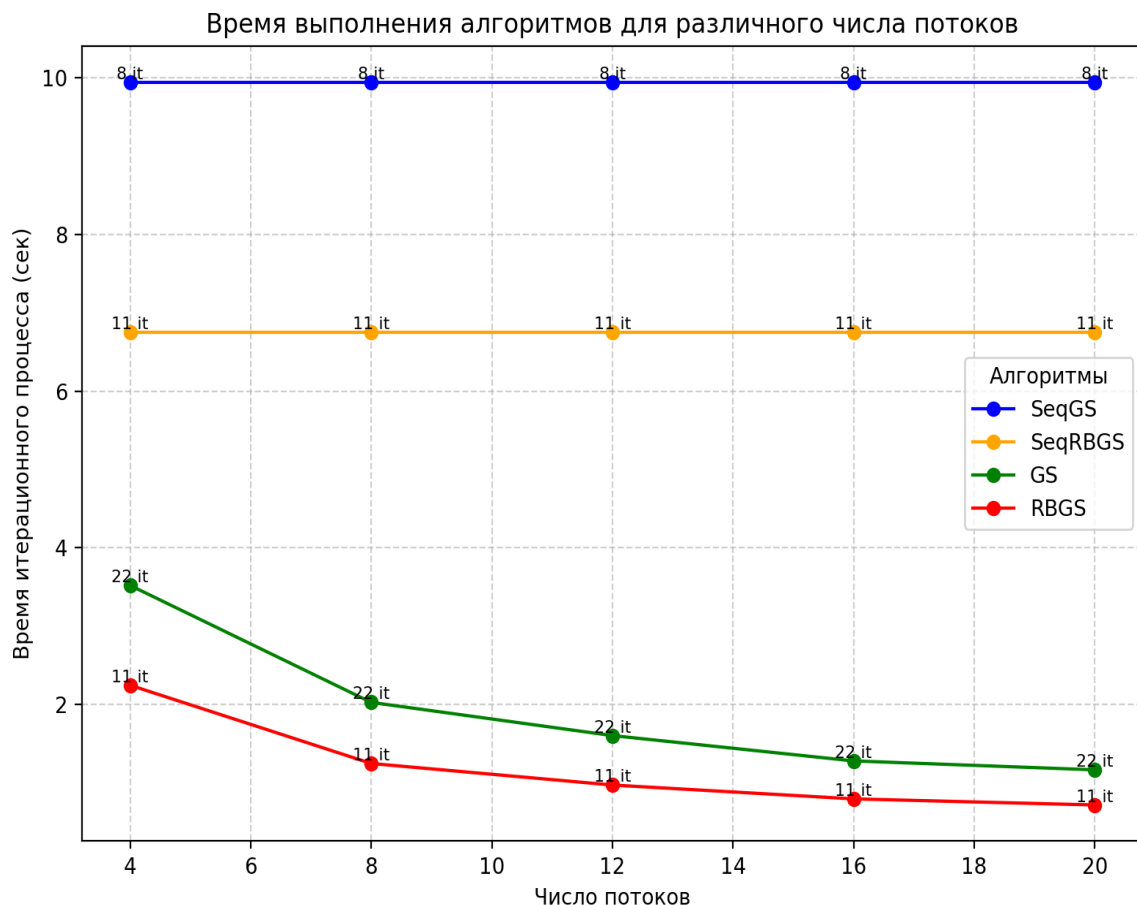


Рис. 1. Зависимость времени работы алгоритмов от количества потоков,
 $N = 1000$, $M = 3000$

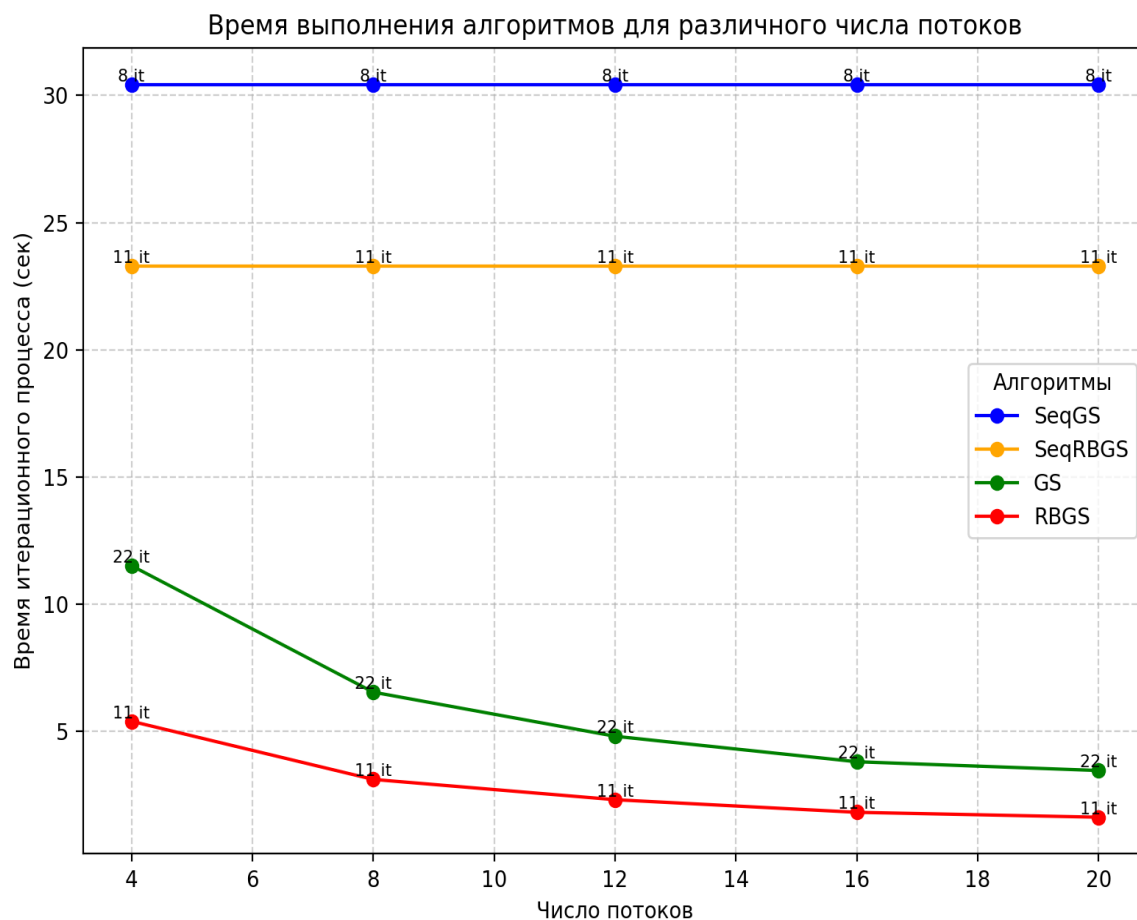


Рис. 2. Зависимость времени работы алгоритмов от количества потоков, $N = 3000, M = 3000$

зеленый – параллельному алгоритму (GS), красный – параллельному алгоритму с использованием красно-черного упорядочения (RBGS). Последовательные алгоритмы реализованы на языке программирования C++, параллельные – на языке C++ с использованием OpenMP. На графиках также указано количество итераций, необходимых для достижения требуемой точности.

Меньшее время выполнения алгоритма SeqRBGS, по сравнению с алгоритмом SeqGS, несмотря на большее количество выполненных итераций, можно объяснить лучшим использованием памяти с быстрым доступом. Из графиков видно, что наименьшее время для достижения требуемой точности достигается для блочного параллельного алгоритма Гаусса-Зейделя с использованием красно-черного упорядочения, он значительно превосходит другой параллельный алгоритм.

СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ ГОЛОВЫ ЧЕЛОВЕКА НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

М. М. Лукашевич^{1), 2)}, В. В. Венгеренко²⁾, А. А. Воронов²⁾

¹⁾ Белорусский государственный университет,
Минск, Беларусь, lukashevichmm@bsu.by

²⁾ Объединённый институт проблем информатики НАН Беларуси,
Минск, Беларусь, lukashevichmm@bsu.by, vengerenko@lsi.bas-net.by,
voronov@lsi.bas-net.by

В статье рассматривает сегментации головы человека на изображениях с целью последующего построения 3D-модели для биометрической идентификации и сравнения с данными из видеонаблюдения. Представлены и протестированы современные архитектуры нейронных сетей: U-Net, PortraitNet, Segmenter и Segment Anything Model (SAM). Эксперименты проведены на наборе данных из пар «изображение–маска» с различными ракурсами.

Ключевые слова: сегментация изображений; удаление фона; нейронные сети; U-Net; Segmenter; предобработка видеоданных; алгоритмы компьютерного зрения.

HUMAN HEAD IMAGE SEGMENTATION BASED ON NEURAL NETWORKS

М. М. Lukashevich^{a), b)}, V. V. Vengerenko^{b)}, A. A. Voronov^{b)}

^{a)} Belarusian State University,

Minsk, Belarus, lukashevichmm@bsu.by

^{b)} The United Institute of Informatics Problems,

Minsk, Belarus, lukashevichmm@bsu.by, vengerenko@lsi.bas-net.by,
voronov@lsi.bas-net.by

The article addresses human head segmentation in images for subsequent 3D model reconstruction, aimed at biometric identification and comparison with surveillance data. Modern neural network architectures (U-Net, PortraitNet, Segmenter, and Segment Anything Model (SAM)) are presented and tested. Experiments are conducted on a dataset of image-mask pairs with various viewing angles.

Keywords: image segmentation; background removal; neural networks; U-Net; Segmenter; video data preprocessing; computer vision algorithms.

1. Введение

Сравнение фотографий человека, полученных в различных условиях, с существующими фотографиями, хранящимися в базах данных, является распространенной практикой для принятия решений. Такие виды

сравнения используются в основном для идентификации личностей преступников, подозреваемых, жертв и неопознанных лиц, а также в других аналогичных областях. В ходе таких сравнений используются изображения человека в реальной сцене и изображения, сделанные в контролируемой среде. Однако углы обзора объекта интереса (человека) в подавляющем большинстве случаев не совпадают, т.к. изображения реальных сцен регистрируются случайным образом, а изображения человека в контролируемой среде делаются под прямым углом. Очевидно, что произвести идентификацию личности достаточно сложно. В этом случае требуется другой ракурс анализа изображений, а он может быть доступен. Одним из возможных решений является технология получения 3D моделей головы человека. На ее основе возможно получение изображения с заданным углом поворота головы [1].

В общепринятой методологии построения 3D модели головы человека условно выделяют несколько этапов:

- 1) установка параметров положения и поворота устройства ввода (камеры) является тривиальной задачей и может быть оптимизирована с помощью обычных инженерных методов;
- 2) получение изображений из видео и выделение силуэта человека (удаление фона) на этих изображениях;
- 3) сегментация изображения головы человека на изображении для лучшего выравнивания и измерения;
- 4) удаление неинформативных артефактов и определением очертаний объектов;
- 5) построение трехмерной сетки и вписывание в нее полученных изображений;
- 6) текстурирование.

Как указано, технология генерации компьютерной 3D модели головы человека для сравнения с изображениями, полученными с камеры, включает этап сегментации головы человека, отделение ее от фона. Указанная задача и является целью данной работы, а ее результаты станут элементом разрабатываемой технологии построения 3D модели головы.

2. Алгоритмы удаления фона на изображениях головы человека

Удаление фона на изображениях головы человека представляет собой важное направление в области компьютерного зрения и цифровой обработки изображений. Эта задача находит применение в телемедицине, системах дополненной реальности, видеоконференцсвязи, биометрической идентификации, а также в инструментах цифрового редактирования.

Наиболее заметные успехи в данной области связаны с использованием глубоких сверточных нейронных сетей. Архитектуры, такие как U-Net, DeepLabV3+ и Mask R-CNN, показывают высокое качество

сегментации благодаря своей способности моделировать сложные зависимости между пикселями. Эти модели обеспечивают детализированное выделение контуров лица и хорошо справляются с разнообразием фонов и условий освещения. Недостатком является необходимость использования значительных вычислительных мощностей и наличие большого объема размеченных данных для обучения.

Для работы на мобильных устройствах и в реальном времени предложены оптимизированные архитектуры, такие как LiteUNet, MODNet и PortraitNet. Они сочетают в себе приемлемое качество сегментации и высокую скорость обработки, что делает их подходящими для внедрения в потребительские устройства и облачные сервисы с ограничениями по производительности.

Одним из наиболее перспективных направлений является использование Transformer-архитектур, таких как Segmenter, TransUNet и Swin-Unet, которые способны эффективно захватывать контекст и пространственные зависимости, особенно в сложных сценариях. Также набирает популярность комбинация CNN и Transformer для создания гибридных моделей, сочетающих преимущества обоих подходов.

В условиях недостатка размеченных данных всё большее значение приобретают методы самообучения и few-shot learning. Модели, такие как DINO, MAE и Segment Anything Model (SAM), открывают новые возможности для применения таких нейронных сетей без необходимости полной аннотации данных.

Исходя из результатов анализа определены следующие перспективные архитектуры для сегментации и удаления фона на изображениях головы человека: U-Net, PortraitNet, Segmenter и SAM [2–7].

3. Экспериментальные исследования

Выполнено прототипирование обучения и тестирования нейронных сетей на основе языка программирования Python, версия 3.12.11 с использованием GPU T4. Используемые при написании кода ключевые библиотеки, а также их назначение приведены в табл. 1.

Таблица 1

Ключевые библиотеки для прототипирования алгоритмов удаления фона на основе нейронных сетей

Библиотека	Назначение
torch	Создание архитектуры нейронной сети, работа с GPU, работа с тензорами, реализация алгоритма обучения, а именно вычисление функции потерь и оптимизация весов нейронной сети; загрузка набора данных, батчей, итерации по обучающим и тестовым данным. Загрузка MobileNetV2 как backbone для архитектуры PortraitNet, использование предобученных весов (IMAGENET1K_V1) для переноса обучения.

Библиотека	Назначение
torchvision	Изменение размера изображения, конвертация в тензор, нормализация изображения. Нормализация по ImageNet-статистикам.
Pillow	Работа с изображениями в Python, загрузка изображений, перевод изображений из RGB в полутон.
scikit-learn	Разбиение на обучающую и тестовую выборки набора изображений, вычисление метрик качества.
transformers	Загрузка модели, загрузка препроцессора, дообучение модели SegFormer B0 на пользовательских данных.
segment-anything	Загрузка архитектуры SAM, создание предиктора, сегментация по точкам
cv2	Загрузка и обработка изображений

Основой для создания моделей нейронных сетей являются наборы данных. Для обучения и тестирования моделей использовался набор данных, содержащий изображений и маски сегментации в формате .png размером 512×512 пикселей. Пример изображения и соответствующей ему маски приведены на рис. 1.

Всего набор данных содержит 730 изображений и соответствующих им масок. Из них 307 изображений головы со спины, 428 изображений – вид спереди головы человека, а также 211 изображений – вид сбоку. 584 изображения использовались для обучения нейронных сетей, а 146 изображений – для тестирования.



Рис. 1. Изображение головы человека и соответствующая ему маска

Сегментация головы человека фактически сводится к задаче бинарной сегментации, где присутствует фон и один класс. В этом случае

необходимо использовать следующие метрики для оценки качества сегментации [8–10].

IoU (Intersection over Union) или Jaccard Index – отношение пересечения (предсказанной и истинной масок) к их объединению:

$$IoU = \frac{TP}{TP+FP+FN}, \quad (1)$$

где *TP* (True Positive) – пиксели, правильно предсказанные как объект; *FP* (False Positive) – пиксели, ошибочно предсказанные как объект (ложные срабатывания); *FN* (False Negative) – пиксели, ошибочно предсказанные как фон (пропущенные объекты).

Dice коэффициент:

$$Dice = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}, \quad (2)$$

где *TN* (True Negative) – пиксели, правильно предсказанные как фон.

Ассурасу (точность) – доля правильно классифицированных пикселей:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}. \quad (3)$$

Каждая из трёх архитектур нейронных сетей (U-net, PortraitNet, Segmenter) была обучена на обучающем наборе данных и протестирована на тестовом наборе данных. Модель SAM (sam_vit_b_01ec64.pth) не до-обучалась, а тестировалась на тех же тестовых данных. Были определены координаты референсных точек для сегментации. В табл. 2 приведены результаты тестирования обученных моделей и модели SAM. Результаты показывают, что максимальные значения трёх метрик для оценки качества сегментации показывает модель Segmenter.

Таблица 2

Результаты тестирования моделей

Архитектура	IoU	Dice	Accuracy
U-net	0,9307	0,9641	0,9927
PortraitNet	0,9159	0,9560	0,9914
Segmenter	0,9739	0,9867	0,9973
SAM	0,6634	0,7931	0,9597

На рис. 2–4 представлены графики обучения (изменение функции потерь) и тестирование моделей на тестовых данных.

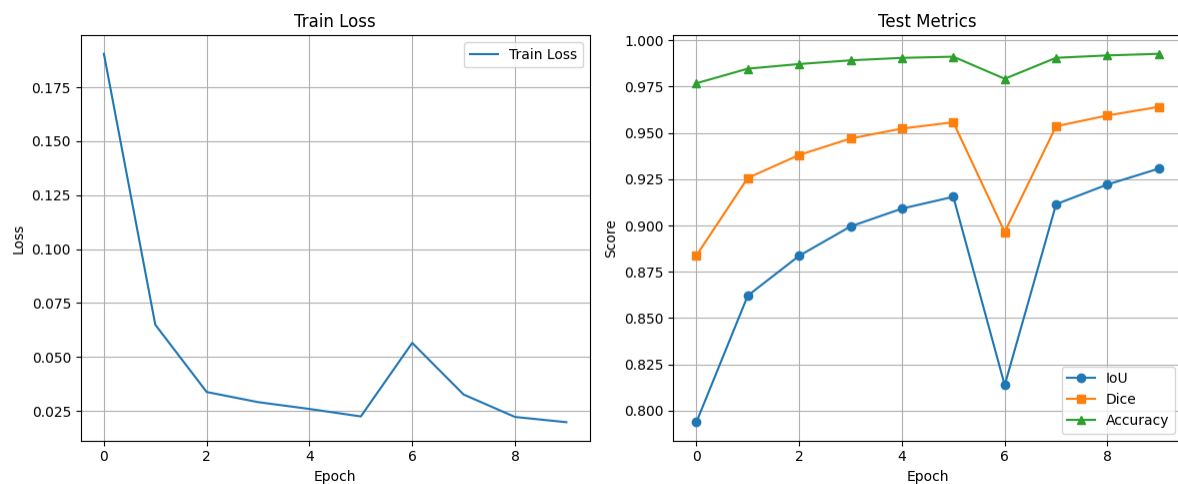


Рис. 2. Графики обучения (изменение функции потерь) U-net и тестирования модели U-net

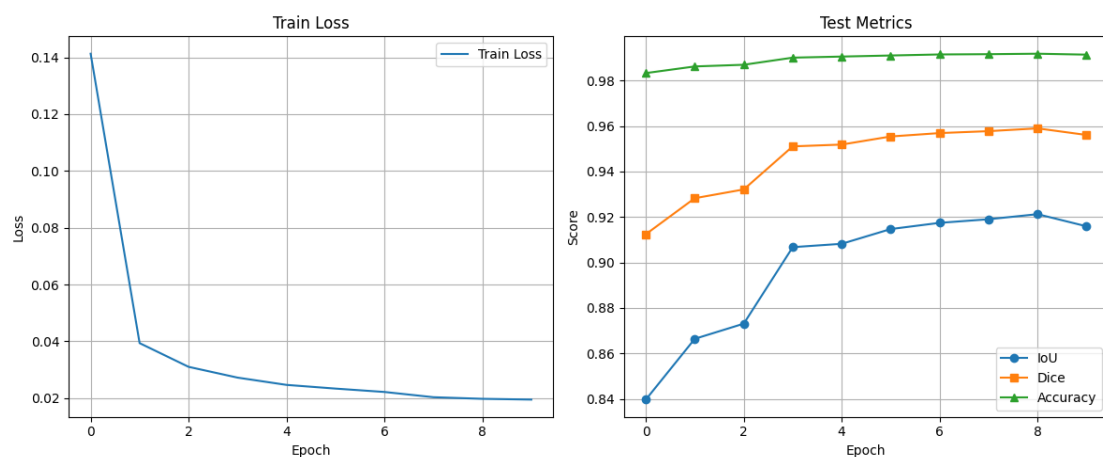


Рис. 3. Графики обучения (изменение функции потерь) и тестирования модели PortraitNet

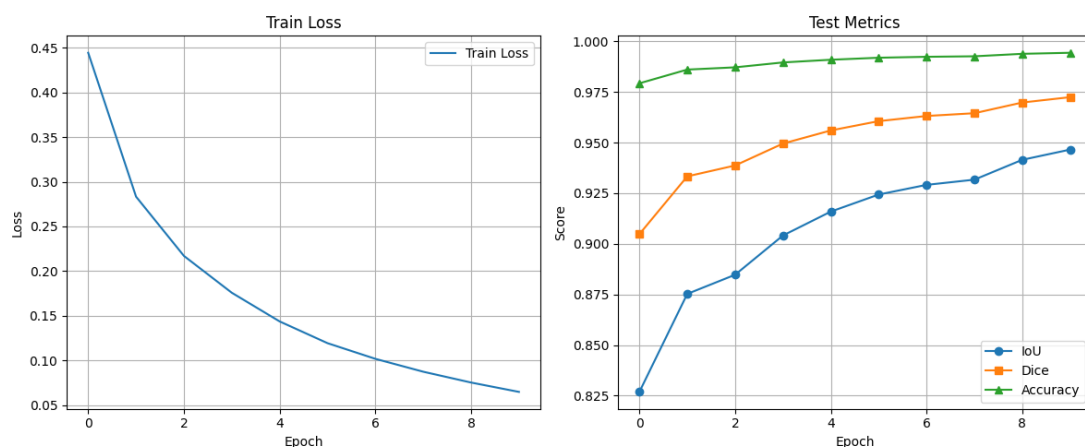


Рис. 4. Графики обучения (изменение функции потерь) и тестирования модели Segmenter

Проведенное в работе исследование выполнено в рамках совместного белорусско-турецкого проекта № Ф25ТУРГ-001.

4. Заключение

Для решения задачи удаления фона на изображениях головы человека были выбраны и протестированы четыре современные архитектуры нейронных сетей: U-Net, PortraitNet, Segmenter и SAM. Модели U-Net, PortraitNet и Segmenter были обучены на наборе данных, представленных исходными изображениями головы человека в разных ракурсах и эталонными масками сегментации. Модель SAM использовалась без дообучения с предварительно расстановкой опорных точек для сегментации.

Результаты тестирования показали, что наилучшую точность демонстрирует модель Segmenter ($\text{IoU} = 0,9739$, $\text{Dice} = 0,9867$, $\text{Accuracy} = 0,9973$), что делает её наиболее подходящей для задачи сегментации головы. Модели U-Net и PortraitNet также показали высокие результаты, тогда как SAM, несмотря на гибкость, показал значительно более низкое качество ($\text{IoU} = 0,6634$) из-за сложности автоматического определения опорных точек. Полученные результаты станут элементом разрабатываемой технологии построения 3D модели головы человека.

Библиографические ссылки

1. Towards fast, accurate and stable 3D dense face alignment / J. Guo [et al.] // Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX / Cham : Springer International Publishing, 2020. P. 152–168.
2. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Freiburg : University of Freiburg, 2015.
3. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation / L.-C. Chen [et al.]. Mountain View : Google Inc., 2018.
4. Convolutional Neural Networks for Large-Scale Remote Sensing Image Classification / E. Maggiori [et al.]. Cambridge : HAL, 2017.
5. Shelhamer E., Long J., Darrell T. Fully Convolutional Networks for Semantic Segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016. Vol. 39, № 4. P. 640–651.
6. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. Vol. 39, № 12. P. 2481–2495.
7. Pyramid Scene Parsing Network / H. Zhao [et al.] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. № 1. P. 6230–6239.
8. Szeliski R. Computer Vision: Algorithms and Applications. 2nd ed. Cham : Springer, 2022.
9. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge : MIT Press, 2016.
10. Prince S. J. D. Computer Vision: Models, Learning, and Inference. Cambridge : Cambridge University Press, 2012.

ВЕБ-ПРИЛОЖЕНИЕ ДЛЯ РЕКОМЕНДАЦИИ МЕТОДОВ БАЛАНСИРОВКИ ДАННЫХ

М. М. Лукашевич¹⁾, Е. Клицунова²⁾

¹⁾ Белорусский государственный университет,
Минск, Беларусь, LukashevichMM@bsu.by

²⁾ Белорусский государственный университет,
Минск, Беларусь, kateryna.klitsunova@gmail.com

Статья посвящена разработке программного средства для решения проблемы выбора методов обработки несбалансированных данных. Представлено веб-приложение, интегрирующее модель машинного обучения, которая рекомендует алгоритмы балансировки на основе характеристик входного набора данных. Описаны процесс формирования обучающих данных, проектирование модели и архитектура веб-приложения, реализованного на фреймворке Streamlit.

Ключевые слова: несбалансированные данные; машинное обучение; балансировка данных; рекомендательная система; веб-приложение.

WEB APPLICATION FOR RECOMMENDING DATA BALANCING METHODS

M. M. Lukashevich^{a)}, K. Klitsunova^{b)}

^{a)} Belarusian State University,
Minsk, Belarus, LukashevichMM@bsu.by

^{b)} The United Institute of Informatics Problems,
Minsk, Belarus, kateryna.klitsunova@gmail.com

The article is devoted to the development of a software tool for solving the problem of choosing methods for processing imbalanced data. A web application is presented that integrates a machine learning model that recommends balancing algorithms based on the characteristics of the input dataset. The process of generating training data, designing the model, and the architecture of the web application implemented on the Streamlit framework are described.

Keywords: imbalanced data; machine learning; data balancing; recommender system; web application.

1. Введение

Широкое распространение задач бинарной и мультиклассовой классификации в таких критически важных областях, как медицина, финансовая безопасность и инженерия, часто сопряжено с проблемой

несбалансированности данных – существенного преобладания объектов одного или нескольких классов.

Для решения данной проблемы предложен широкий спектр методов [1–4], которые условно можно разделить на следующие категории: методы на уровне данных (балансировка путем синтетического увеличения меньшего класса или уменьшения большего), методы на уровне алгоритмов (например, обучение с учетом издержек классификации) и ансамблевые методы. Однако отсутствие универсального подхода, эффективного для всех типов задач, и многообразие доступных техник балансировки создают сложность выбора подходящей техники на практике.

В связи с этим актуальной задачей является разработка инструмента, способного автоматизировать и оптимизировать процесс выбора метода предобработки данных на основе их объективных характеристик. Целью данной работы является описание разработки программного средства – интерактивного веб-приложения, которое интегрирует в себя модель машинного обучения для рекомендации наиболее эффективных методов и их комбинаций для балансировки конкретного набора данных, основанную на результатах предварительного комплексного экспериментального исследования.

2. Формирование обучающих данных для модели

Основой для построения модели рекомендаций послужили результаты экспериментального исследования, подробно описанного в предыдущей работе [5]. Целью эксперимента была всесторонняя оценка влияния различных методов балансировки, а также их комбинаций, на качество моделей классификации.

Было отобрано пять наборов данных с платформы Kaggle объемом от ~2 тыс. до ~254 тыс. записей и с разной степенью дисбаланса (от 2:100 до 29:100), их описание приводится в табл. 1.

Для каждого набора данных был выполнена стандартная предобработка: обработка пропущенных значений и выбросов, кодирование категориальных признаков, нормализация числовых признаков и стратифицированное разделение на обучающую и тестовую выборки в соотношении 70 к 30.

Таблица 1

Описание наборов данных

Название	Размер набора данных	Тип классификации	Соотношение классов
Классификация здоровья плода	~2 тысячи записей	мультиклассовая	8:100 и 13:100
Прогнозирование церебрального инсульта	~43 тысячи записей	бинарная	2:100

Название	Размер набора данных	Тип классификации	Соотношение классов
Кредитный риск	~32 тысячи записей	бинарная	22:100
Классификация кредитного рейтинга	100 тысяч записей	мультиклассовая	17:100 и 29:100
Показатели здоровья при диабете	~254 тысячи записей	мультиклассовая	2:100 и 15:100

На каждом из подготовленных наборов данных проводилось обучение и оценка качества множества моделей. Тестировались как классические алгоритмы (решающее дерево, k-ближайших соседей, метод опорных векторов, наивный байесовский классификатор), так и ансамблевые (случайный лес, градиентный бустинг, XGBoost, LightGBM, AdaBoost). Для каждой модели фиксировалось значение сбалансированной точности (Balanced Accuracy). Балансировка данных проводилась изолированно следующими методами.

Увеличение меньшего класса: Random Oversampling, SMOTE, Borderline-SMOTE, Borderline-SMOTE SVM, ADASYN.

Уменьшение большего класса: Random Undersampling, NearMiss, Tomek Links, Condensed Nearest Neighbors, Edited Nearest Neighbors, One-Sided Selection, Neighborhood Cleaning Rule.

Далее исследовалось комбинированное применение наиболее эффективных методов из обеих групп. Для каждого набора данных, каждой модели и каждого метода балансировки (как изолированного, так и комбинированного) сохранялось итоговое значение сбалансированной точности.

Для создания набора данных для обучения модели рекомендаций из результатов эксперимента были извлечены следующие признаки для каждого наблюдения:

- 1) Size: объем набора данных в тысячах записей;
- 2) Imbalance Ratio: степень дисбаланса, выраженная как отношение количества объектов в наименьшем классе к количеству объектов в наибольшем классе;
- 3) Type: тип классификации, бинарная или мультиклассовая;
- 4) Oversampling: метод увеличения меньшего класса (если применялся);
- 5) Undersampling: метод уменьшения большего класса (если применялся);
- 6) Balanced Accuracy: значение сбалансированной точности, достигнутое после применения данной комбинации методов.

Поскольку абсолютные значения метрики сильно варьировались от набора к набору, для нивелирования этого эффекта была проведена

минимаксная нормализация значения сбалансированной точности в пределах каждого исходного набора данных. Это позволило перейти от абсолютных значений точности к относительным рангам эффективности методов внутри каждого конкретного кейса.

Категориальные признаки (Type, Oversampling, Undersampling) были преобразованы в числовой формат с использованием кодирования меток (Label Encoding). Итоговый обучающий набор представлял собой таблицу, где каждой строке соответствовала уникальная комбинация характеристик набора данных и примененных методов балансировки, а целевой переменной выступало нормализованное значение эффективности этой комбинации.

3. Проектирование и обучение модели ранжирования

Задачей модели было ранжирование методов балансировки по предполагаемой эффективности для новых данных.

В качестве базовых алгоритмов были выбраны регрессоры, показавшие высокую эффективность в задачах ранжирования [6, 7]: Gradient Boosting Regressor, Extra Trees Regressor, Random Forest Regressor, XGBoost Regressor и LightGBM Regressor. Для настройки их гиперпараметров и объективного сравнения был применен фреймворк Optuna, который проводил поиск по заданному пространству гиперпараметров для каждого алгоритма.

Оценка качества моделей в процессе кросс-валидации проводилась с использованием метрики R^2 (коэффициент детерминации), так как она хорошо интерпретирует долю дисперсии, объясненную моделью. Для обеспечения репрезентативности оценки использовалась 5-кратная кросс-валидация с перемешиванием (Stratified K-Fold).

По итогам оптимизации были отобраны три алгоритма, показавшие наилучшие и наиболее стабильные результаты: XGBoost Regressor, Gradient Boosting Regressor и Extra Trees Regressor.

Для повышения обобщающей способности и устойчивости предсказаний была протестирована идея объединения лучших моделей в ансамбль. Были опробованы следующие подходы: бэггинг, стекинг и голосящий ансамбль.

Наилучшие результаты показал ансамбль на основе голосования (Voting Regressor), объединивший три лучших алгоритма, который и был выбран в качестве финальной модели.

Для количественной оценки были рассчитаны стандартные метрики регрессии: средняя абсолютная ошибка (MAE), среднеквадратическая ошибка (MSE) и коэффициент детерминации (R^2) для каждого набора кросс-валидации (табл. 2).

Таблица 2

Обобщение результатов экспериментальных исследований

Номер блока	MSE	MAE	Коэффициент детерминации
1	0,01135477	0,07120671	0,69812307
2	0,02629933	0,10227519	0,60890527
3	0,00695144	0,04180257	0,65440132
4	0,01836801	0,05555495	0,59469433
5	0,01021363	0,0535413	0,79880918

Полученные значения подтвердили адекватную предсказательную способность модели и ее применимость для задачи ранжирования.

4. Разработка веб-приложения

Для практического применения модели было разработано веб-приложение. Для его реализации был выбран язык программирования Python, что обеспечило бесшовную интеграцию с уже разработанной моделью машинного обучения и библиотеками для работы с данными (Pandas, NumPy, Scikit-learn, Imbalanced-learn).

Для реализации веб-интерфейса был выбран фреймворк Streamlit. Его основное преимущество – возможность создания интерактивных веб-приложений с богатыми визуальными возможностями практически целиком на языке Python, без необходимости написания клиент-серверного кода. Это значительно ускорило процесс разработки и позволило сосредоточиться на логике приложения.

Приложение развернуто в облачной среде Streamlit Community Cloud, не требует установки и доступно по публичной ссылке <https://resample.streamlit.app/>. Интерфейс выполнен на английском языке для универсальности и ориентирован на специалистов в области машинного обучения (рис. 1).

Работа пользователя с приложением следует линейной последовательности.

1. Загрузка данных. Пользователь выбирает файл или демонстрационный набор данных.

2. Предобработка. Приложение определяет наличие пропусков и предлагает варианты их обработки.

3. Рекомендация. Пользователь получает список лучших комбинаций методов наилучшей предполагаемой эффективностью (рис. 2).

4. Балансировка. Пользователь выбирает желаемый метод и параметры балансировки из выпадающих списков.

5. Визуализация и экспорт. Приложение показывает диаграммы соотношения классов до и после балансировки и предоставляет кнопку для скачивания сбалансированного набора данных.

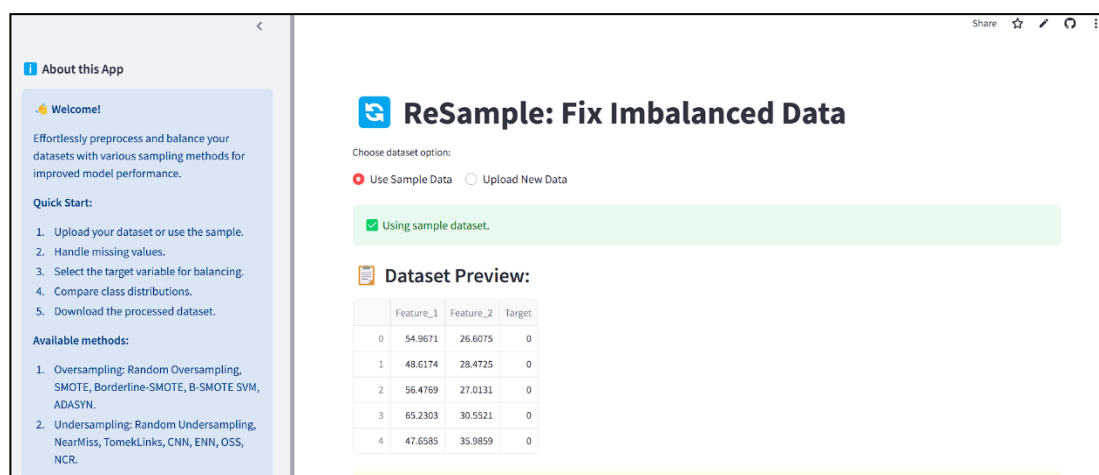


Рис. 1. Интерфейс веб-приложения

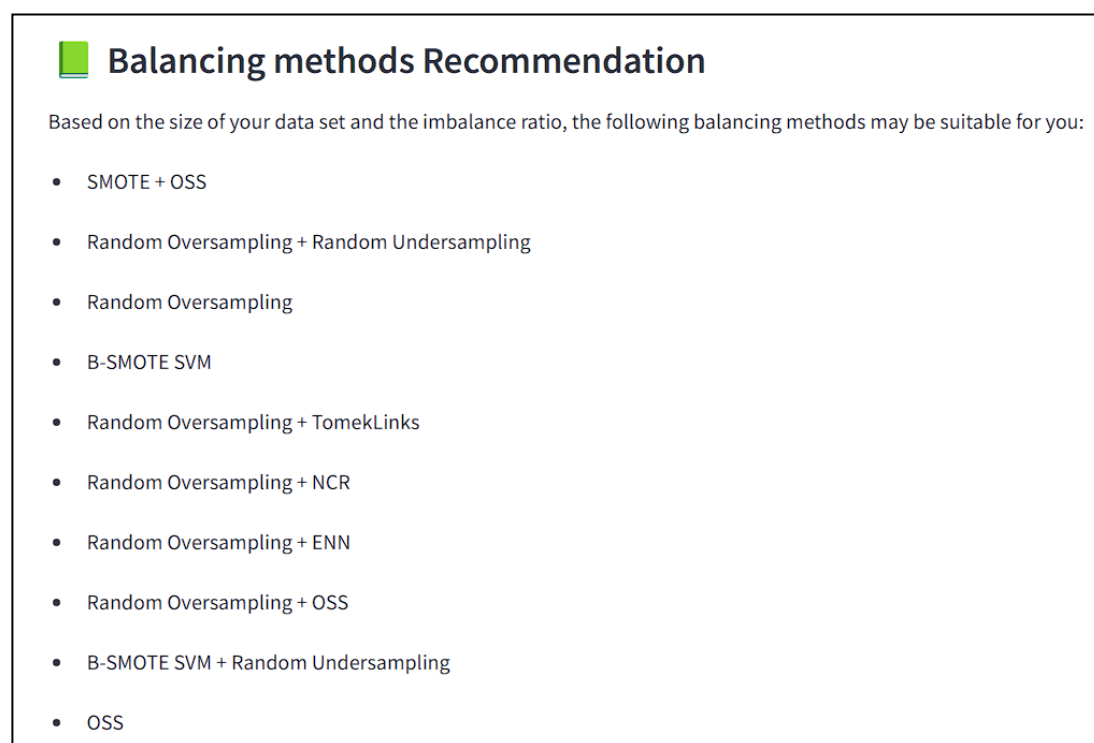


Рис. 2. Рекомендация методов балансировки в веб-приложении

Приложение развернуто в облачной среде и доступно онлайн, что обеспечивает его кроссплатформенность и удобство использования.

5. Заключение

В данной статье представлена разработка программного средства, предназначенного для решения актуальной проблемы выбора методов обработки несбалансированных данных в машинном обучении. Основное преимущество предлагаемого решения заключается в интеграции модели

машинного обучения, обученной на результатах комплексного экспериментального исследования, в удобный и доступный веб-интерфейс.

Разработанное программное средство может быть использовано специалистами по анализу данных и машинному обучению в различных предметных областях для быстрого старта работы с несбалансированными наборами данных и повышения качества построенных классификационных моделей.

Библиографические ссылки

1. Classification of imbalanced data: review of methods and applications / P. Kumar [et al.] // IOP conference series: materials science and engineering. IOP Publishing, 2021. Vol. 1099, №. 1. Article no. 012077.
2. *Krawczyk B.* Learning from imbalanced data: open challenges and future directions // Progress in artificial intelligence. 2016. Vol. 5, №. 4. P. 221–232.
3. *Branco P., Torgo L., Ribeiro R.* A survey of predictive modeling on imbalanced domains // ACM computing surveys (CSUR). 2016. Vol. 49, №. 2. P. 1–50.
4. *Sun Y., Wong A. K. C., Kamel M. S.* Classification of imbalanced data: A review // International journal of pattern recognition and artificial intelligence. 2009. Vol. 23, №. 4. P. 687–719.
5. *Клицунова Е., Лукашевич М. М.* Сравнительный анализ методов балансировки данных для задач машинного обучения // BIG DATA и анализ высокого уровня : сборник научных статей XI Международной научно-практической конференции / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: В. А. Богуш [и др.]. Минск : БГУИР, 2025. С. 74–83.
6. *Li H.* A short introduction to learning to rank // IEICE TRANSACTIONS on Information and Systems. 2011. Vol. 94, №. 10. P. 1854–1862.
7. Learning to rank for information retrieval / T. Y. Liu [et al.] // Foundations and Trends in Information Retrieval. 2009. Vol. 3, №. 3. P. 225–331.

АРХИТЕКТУРА ВЕРОЯТНОСТНОГО ТРАНСФОРМЕРА ДЛЯ МОДЕЛИРОВАНИЯ ЗАДЕРЖЕК В МНОГОЭТАПНЫХ ПРОЦЕССАХ

Д. А. Сьянов

*Белорусский государственный университет,
Минск, Беларусь, syanov@bsu.by*

В статье представлена адаптивная архитектура Adaptive Probabilistic Sequence-to-Vector Transformer (APS2V-T), предназначенная для совместного прогнозирования задержек ключевых этапов университетского документооборота, связанного с организацией конференции студентов и аспирантов БГУ. Предлагаемый подход обеспечивает интерпретируемость, калибруемость прогнозов и снижает вычислительную сложность при фиксированном наборе целевых этапов.

Ключевые слова: трансформеры; прогнозирование задержек; документооборот; вероятностное моделирование; process mining.

A PROBABILISTIC TRANSFORMER ARCHITECTURE FOR MODELING DELAYS IN MULTI-STAGE PROCESSES

D. A. Syanov

*Belarusian State University,
Minsk, Belarus, syanov@bsu.by*

The article presents an adaptive architecture, the Adaptive Probabilistic Sequence-to-Vector Transformer (APS2V-T), designed for the joint prediction of delays in key stages of university document workflows related to organizing a conference for students and postgraduate students at BSU. The proposed approach ensures interpretability, forecast calibration, and reduces computational complexity for a fixed set of target stages.

Keywords: transformers; delay prediction; document workflow; probabilistic modeling; process mining.

1. Введение

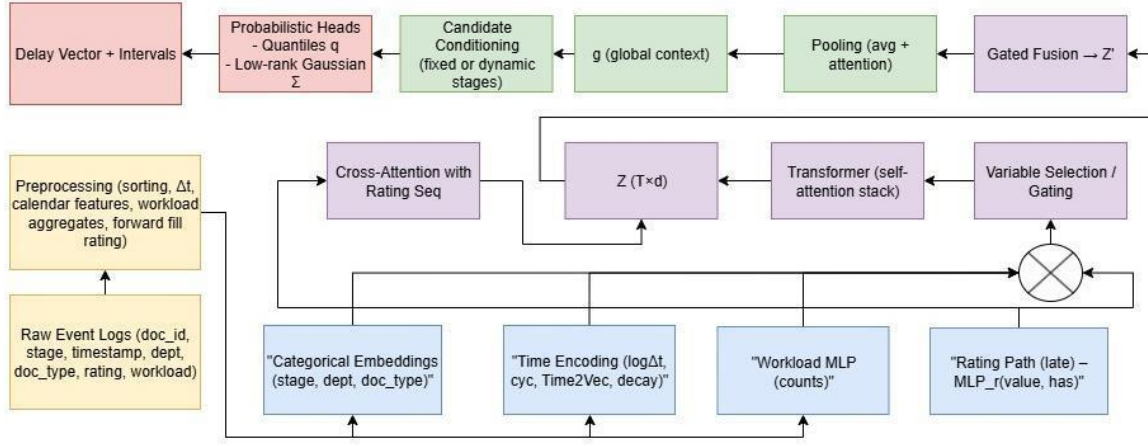
Прогнозирование задержек на этапах документооборота, связанного с организацией студенческих научных конференций в университетах, таких как приём тезисов, рецензирование или согласование программы, является критически важным для эффективного управления сроками и ресурсами, а также является ключевым элементом проактивного управления

SLA и ресурсами в электронном документообороте [1]. Задержки в этих процессах могут нарушить жёсткие дедлайны, снизить удовлетворённость участников и увеличить административные расходы. Современные подходы в области predictive business process monitoring (PBPM) преимущественно фокусируются на предсказании времени завершения всего процесса или следующего события [2, 3].

Существующие решения обладают рядом существенных ограничений. Прежде всего, отсутствие возможности одновременного прогнозирования полного вектора задержек приводит к фрагментарным и несогласованным результатам. Дополнительной проблемой является недостаточный учет глобального контекста, из-за чего ранние события, способные повлиять на весь процесс, зачастую остаются вне внимания. Ситуацию осложняет и то, что большинство подходов не предполагают гибкой обработки динамически изменяющегося множества этапов, характерного для ветвящихся процессов. Традиционные методы, такие как независимые регрессоры для каждого этапа или LSTM-модели [2], либо игнорируют корреляции между этапами, либо ограничены локальным контекстом. Например, независимые модели могут предсказывать противоречивые сроки для рецензирования и публикации сборника, а LSTM-подходы [2] не учитывают ранние события в длительных процессах. Даже более продвинутые архитектуры, такие как Temporal Fusion Transformer [5], оптимизированы прежде всего под многошаговый прогноз по фиксированным временным горизонтам, а не под семантически маркированные стадии, и не предлагают специализированного механизма для поздно появляющихся признаков (рейтинга документа, формирующегося только после его верификации).

Предлагаемая архитектура (рисунок) решает эту несогласованность, переформулируя задачу как вероятностный sequence-to-vector (S2V) прогноз: вся наблюдаемая префиксная история документа преобразуется одним трансформер-энкодером в глобальное контекстное представление, из которого за один проход выводится многомерный вектор задержек по ключевым этапам, дополненный квантилинтервальными или гауссовскими (low-rank) оценками неопределенности. Такая архитектура устраняет каскад накопления ошибок, повышает латентную согласованность между компонентами (поскольку оптимизация ведется по совместной функции потерь), а также снижает вычислительную задержку инференса при росте числа параллельно отслеживаемых документов.

С точки зрения интерпретируемости архитектура предоставляет прозрачные точки наблюдения: self-attention [6] выявляет исторические события с наибольшим вкладом, cross-attention – зависимость от рейтинга, gating – относительную значимость групп признаков. Это упрощает диагностику деградаций качества и формирование объяснений для доменных



APS2V-T Архитектура

экспертов без глубоких ML-компетенций. Встраивание архитектуры в визуальную среду моделирования цветных сетей Петри (CPN IDE) [7] представляется дальнейшим шагом к интерактивному управлению: прогнозы и их неопределенность могут быть непосредственно отображены на переходах формальной модели, а what-if сценарии (перераспределение ресурсов, изменение рейтинга, выбор альтернативной ветви) – мгновенно переоценены. Таким образом формируется замкнутый цикл: предсказание → визуальная диагностика узких мест → интервенция → обновленные логи → адаптация модели.

2. Архитектура модели

Пусть $\sigma_{\leq T} = \langle e_1, \dots, e_T \rangle$ – префикс трассы одного документа. Каждое событие описывается шестёркой признаков:

$$e_i = (a_i, t_i, d_i, u_i, r_i, w_i), \quad (1)$$

где a_i – идентификатор этапа; $t_i \in R_+$ – временная метка; d_i – департамент; u_i – тип документа; r_i – пара рейтинговых признаков (значение и наличие); w_i – агрегаты нагрузки (например, число активных заявок). Множества категориальных признаков $\mathcal{A}$, $\mathcal{D}$, $\mathcal{U}$ предполагаются конечными. Для временной динамики вводится приращение (для $i=1$ берётся специальный токен): $\Delta t_i = t_i - t_{i-1}$, $i \geq 2$. Целевой объект прогнозирования – вектор задержек:

$$\mathbf{D} = (D_1, \dots, D_m), \quad m = |\mathcal{S}_*|, \quad (2)$$

где $\mathcal{S}_* = \{s_1, \dots, s_m\}$ – фиксированное множество ключевых этапов (например, получение, рецензирование, верификация, доработка, утверждение); D_k – оставшаяся задержка до этапа s_k при текущем префиксе ($D_k = \mathbb{E}[T_{s_k}^{(start)} - t_T | \sigma_{\leq T}]$ для непройденных этапов; $D_k = 0$ или маскируется для завершённых). Предполагается стационарность отображения $f : \sigma_{\leq T} \mapsto \mathbf{D}$ в пределах обучающего периода с допустимым концептуальным дрейфом.

Таким образом, задача заключается в построении параметрического отображения:

$$\hat{f}_{\Theta} : \sigma_{\leq T} \mapsto (\boldsymbol{\mu}, \hat{\Sigma}) \text{ или } (\{\hat{q}_{k,\alpha}\}_{\alpha \in \mathcal{Q}}), \quad (3)$$

где $\boldsymbol{\mu} \in R^m$ – точечные прогнозы; $\hat{\Sigma} \in R^{m \times m}$ – низкоранговая ковариационная матрица, либо квантили для вероятностной интерпретации.

Категориальные признаки кодируются с помощью обучаемых эмбеддингов, задаваемых отображениями, что позволяет каждому элементу множества категорий сопоставить вектор фиксированной размерности:

$$E^{(a)} : \mathcal{A} \rightarrow R^{d_a}, \quad E^{(d)} : \mathcal{D} \rightarrow R^{d_d}, \quad E^{(u)} : \mathcal{U} \rightarrow R^{d_u}. \quad (4)$$

Для временного аспекта формируется гибридная репрезентация, объединяющая несколько компонент: $u_i^{(time)}$. Во-первых, используется логарифмированный интервал между событиями $\psi_{\Delta t} = \text{MLP}_{\Delta}(\log(1 + \Delta t_i / \tau))$, что позволяет стабилизировать распределения с тяжёлыми хвостами, как это принято в моделировании нерегулярных временных процессов. Во-вторых, добавляются циклические признаки, $\sin(2\pi \text{hour}(t_i) / 24), \cos(\cdot); \sin(2\pi \text{dow}(t_i) / 7), \cos(\cdot)$, отражающие периодичность часов и дней недели, что согласуется с подходами позиционного кодирования [8]. Третьим компонентом служит Time2Vec [9], который позволяет моделировать как линейные, так и нелинейные временные зависимости, и описывается выражением $\text{t2v}(t_i) = [\omega_0 t_i + \phi_0, \sin(\omega_1 t_i + \phi_1), \dots, \sin(\omega_K t_i + \phi_K)]$.

Дополнительно вводится экспоненциальное затухание $\gamma(\Delta t_i) = [\exp(-\lambda_1 \Delta t_i), \exp(-\lambda_2 \Delta t_i), \dots, \exp(-\lambda_p \Delta t_i)]$, отражающее убывающую значимость событий по мере увеличения временного интервала.

Отдельным признаком выступает рейтинговый путь, кодируемый через $r_i^{(emb)} = \text{MLP}_r([\text{rating\_value}_i, \text{has\_rating}_i])$, что позволяет учитывать как наличие, так и значение оценки на этапе verification. Наконец, важную роль играет показатель загрузки ресурсов, задаваемый вектором:

$$w_i^{(emb)} = \text{MLP}_w(w_i), \quad (5)$$

где w_i содержит агрегированные характеристики: количество активных документов в департаменте за последние W часов, число незавершённых review и другие показатели, отражающие идеи ресурсных очередей в process mining [1].

Все выделенные компоненты объединяются в итоговую конкатенацию, которая служит входным вектором признаков перед проекцией в латентное пространство модели:

$$\tilde{h}_i = [E^{(a)}(a_i) \parallel E^{(d)}(d_i) \parallel E^{(u)}(u_i) \parallel u_i^{(time)} \parallel w_i^{(emb)}].$$

Для управления вкладом различных групп признаков (categorical, time, workload) используется механизм легковесного gating. Каждая группа подаётся в малую MLP, которая формирует вектор весов:

$$g^{(g)} = \sigma(W^{(g)}\bar{h}^{(g)} + b^{(g)}), \quad g^{(g)} \in (0,1)^{d_g}, \quad (6)$$

где $\bar{h}^{(g)}$ представляет собой агрегированное представление внутри группы, а σ – сигмоидальная функция активации. Эти веса определяют степень релевантности признаков конкретной группы для текущего прогноза. Итоговое смешивание осуществляется по формуле $h_i^{(base)} = \sum_{g \in \mathcal{G}} g^{(g)} \odot h_i^{(g)}$, что позволяет адаптивно подавлять шум от слабо информативных групп и усиливать значимые компоненты.

Такой подход перекликается с variable selection networks [5] и реализует принцип мягкого отбора признаков, аналогичный используемому в Temporal Fusion Transformer (TFT).

Для учёта порядка событий в последовательности в модель добавляется позиционный компонент p_i , реализованный в виде синусоидального кодирования [6]. Дополнительно внутри механизма внимания используется относительный сдвиг $B_{rel}(i, j) = w_{\text{clip}(|i-j|, K)}$, предложенный Peter Shaw et al. [10], что позволяет корректно учитывать вариативную длину

префиксов и усиливать устойчивость к различным масштабам последовательностей. Итоговое входное представление строится как $x_i^{(0)} = \text{LayerNorm}(h_i^{(base)} + p_i)$.

Для кодирования префикса используется стек слоёв самовнимания. На каждом уровне $l=1, \dots, L$ вычисляется многоголовое внимание $\text{MHA}^{(l)}(X) = \text{Concat}(\text{head}_1, \dots, \text{head}_H)W^O$, где каждая голова определяется как

$$\text{head}_h = \text{Softmax}\left(\frac{Q_h K_h^\top + B_{rel} + A_{mask}}{\sqrt{d_h}}\right)V_h \quad (7)$$

с матрицами проекций $Q_h = XW_h^Q$, $K_h = XW_h^K$, $V_h = XW_h^V$. Маска A_{mask} используется для зануления паддинговых позиций ($M_i = 0$), исключая их влияние на вычисления.

После блока внимания применяется нормализация и остаточные связи:

$$\tilde{X}^{(l)} = \text{LayerNorm}(X^{(l-1)} + \text{Dropout}(\text{MHA}^{(l)})), \quad (8)$$

$$X^{(l)} = \text{LayerNorm}(\tilde{X}^{(l)} + \text{Dropout}(\text{FFN}(\tilde{X}^{(l)}))). \quad (9)$$

Таким образом, выходной тензор кодировщика задаётся как $Z = X^{(L)} = [z_1, \dots, z_T]$, где векторы, соответствующие позициям паддинга, принимают тривиальные значения.

На завершающем этапе кодирования вводится специальный механизм для интеграции информации о рейтингах, которые появляются не сразу, а только начиная с определённых стадий процесса. Для этого формируется отдельная последовательность рейтинговых эмбеддингов, отражающая моменты появления и сохранения признака: $R = [r_1^{(emb)}, \dots, r_T^{(emb)}]$.

Слияние осуществляется через cross-attention: матрицы запросов Q строятся на основе скрытых представлений основной последовательности Z , а ключи K и значения V – из последовательности R . Полученный контекст C далее адаптивно встраивается в Z с помощью gated-fusion, где коэффициент α , вычисляемый через усреднённый вектор рейтингов $\bar{r}$, управляет степенью включения дополнительного сигнала. Итоговое представление Z' позволяет «условить» прогноз на поздно наблюдаемые признаки, не искажая ранние этапы последовательности. Такая стратегия согласуется с практиками работы с частично наблюдаемыми признаками во временных моделях [11].

После интеграции всех источников информации требуется переход от последовательного представления к глобальному вектору признаков. Для этого используется комбинированная агрегация: усреднение по непустым позициям (g_{avg}) и attention-пулинг (g_{att}) с обучаемым запросом q , что позволяет выделять наиболее значимые фрагменты префикса. Весовые коэффициенты α_i корректируются маской паддинга через слагаемое $\beta_i = \log M_i$.

Итоговое глобальное представление объединяет равномерную и фокусированную агрегацию $g = [g_{avg} \parallel g_{att}]$, обеспечивая баланс между полной историей и локальными акцентами. Заключительный слой моделирует предсказания для фиксированного множества ключевых этапов $F = s_1, \dots, s_m$, например, received, review, verification, revision, approved. Для каждого этапа используется эмбединг $c_j = E^{(a)}(s_j)$, который далее конкатенируется с глобальным представлением g и преобразуется через нелинейное отображение $\tilde{c}_j = \tanh(W_c[c_j \parallel g] + b_c)$.

В случае расширяемого множества этапов может дополнительно применяться cross-attention кандидатов к последовательности Z' для уточнения контекста, а также маска достижимости, ограничивающая невозможные переходы. Такая конструкция позволяет обобщать архитектуру не только на фиксированное число целей, но и на динамически изменяемые множества этапов бизнес-процесса.

Финальный уровень архитектуры поддерживает два варианта вероятностного вывода, позволяя адаптировать модель к различным сценариям прогнозирования.

1. Квантильная голова. Для множества квантилей $\mathcal{Q}$ формируются отдельные линейные проекции: $q_{j,\alpha} = w_\alpha^\top \tilde{c}_j + b_\alpha$, $\alpha \in \mathcal{Q}$, что обеспечивает построение предсказательных интервалов и обучение с помощью pinball-функции потерь [12].

2. Low-rank Gaussian. В многомерной постановке используется параметризация ковариации с низким рангом: $\mu = (\mu_1, \dots, \mu_m)$, $U = [u_1^\top; \dots; u_m^\top] \in R^{m \times r}$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_m^2)$, $\Sigma = UU^\top + \text{diag}(\sigma^2)$. Такое задание матрицы Σ позволяет эффективно моделировать корреляции между задержками при умеренной вычислительной сложности. Для вычисления обратной матрицы и определителя используется формула Вудбери, что согласуется с практиками low-rank ковариации в многомерном прогнозе [4].

Обучение модели осуществляется в различных режимах, соответствующих выбранной вероятностной голове.

1. Квантильный прогноз: минимизация pinball-ошибки:

$$\mathcal{L}_{\mathcal{Q}} = \frac{1}{|\mathcal{Q}|} \sum_{\alpha \in \mathcal{Q}} \frac{1}{m} \sum_{j=1}^m \rho_{\alpha}(D_j - q_{j,\alpha}), \quad \rho_{\alpha}(u) = \begin{cases} \alpha u, & u \geq 0, \\ (\alpha - 1)u, & u < 0. \end{cases} \quad (10)$$

2. Low-rank Gaussian: минимизация отрицательного логарифма правдоподобия (NLL):

$$\mathcal{L}_{NLL} = \frac{1}{2} [(\mathbf{D} - \boldsymbol{\mu})^{\top} \Sigma^{-1} (\mathbf{D} - \boldsymbol{\mu}) + \log |\Sigma| + m \log(2\pi)]. \quad (11)$$

Финальная функция потерь комбинирует несколько компонентов:

$$\mathcal{L} = \lambda_{main} \mathcal{L}_{main} + \lambda_{ES} \mathcal{L}_{ES} + \lambda_{reg} \mathcal{R}, \quad (12)$$

где $\mathcal{R}$ включает L2-регуляризацию весов, dropout и L1-штраф на коэффициенты gating для индукции разреженности.

На этапе предсказания формирование доверительных интервалов зависит от выбранного режима: В квантильном режиме интервалы определяются напрямую: $(q_{j,\alpha_1}, q_{j,\alpha_2})$.

Для повышения покрытия используется conformal calibration: на валидационной подвыборке вычисляются невязки δ_j , после чего верхняя граница корректируется эмпирическим квантилем $\delta_{(1-\epsilon)}$ (split-конформный предиктор [13]).

В гауссовском режиме доверительные интервалы задаются как $\mu_j \pm z_{\beta} \sigma_j$, при учёте корреляций между компонентами анализируются совместные доверительные эллипсоиды.

Таким образом, архитектура поддерживает гибкий вероятностный вывод с возможностью последующей калибровки, что обеспечивает интерпретируемые и статистически обоснованные интервалы предсказаний.

3. Заключение

Предварительный аналитический разбор и тесты демонстрирует концептуальные преимущества по сравнению с Temporal Fusion Transformer и классическими многовыходными регрессорами: снижение латентности инференса при малом числе целевых компонент, прямое моделирование корреляций между задержками вместо латентного извлечения их из временной решётки, устойчивость к нерегулярности событий благодаря

лог-интервальному и деау-кодированию, а также повышенную интерпретируемость для управленческих сценариев (выделение влияния рейтинга, календарных паттернов и загрузки ресурсов). Архитектура сохраняет инженерную простоту (отсутствие сложного декодера), что снижает барьер внедрения в прикладных средах с ограниченными вычислительными бюджетами.

Ограничения текущей версии включают квадратичную сложность self-attention при очень длинных трассах, ограниченную мультимодальность гауссовской low-rank параметризации, отсутствие явной причинной декомпозиции (рейтинг трактуется ассоциативно) и необходимость внешней агрегации ресурсных сигналов. Данная архитектура предложена в перспективе интеграции с CPN IDE (цветные сети Петри) для визуального наложения прогнозов и их неопределённости на формальную модель процесса.

Библиографические ссылки

1. *Aalst W. van der Process Mining : Data Science in Action*. 2nd ed. Berlin ; Heidelberg : Springer-Verlag, 2016. URL: http://repo.darmajaya.ac.id/4319/1/Process%20Mining_%20Data%20Science%20in%20Action%20%28%20PDFDrive%20%29.pdf (date of access: 15.09.2025).
2. Predictive Business Process Monitoring with LSTM Neural Networks / N. Tax [et al.] // *Advanced Information Systems Engineering : 29th International Conference, CAiSE 2017, Essen, Germany, June 12-16, 2017, Proceedings* / eds.: E. Dubois, K. Pohl. Springer Cham, 2017. (Lecture Notes in Computer Science, Vol. 10253.) URL: https://www.researchgate.net/publication/318134519_Predictive_Business_Process_Monitoring_with_LSTM_Neural_Networks (date of access: 15.09.2025).
3. Survey and Cross-benchmark Comparison of Remaining Time Prediction Methods in Business Process Monitoring / I. Verenich [et al.] // *ACM TIST*. 2019. Vol. 10, iss. 4. Article no. 34. URL: https://www.researchgate.net/publication/334581217_Survey_and_Cross-benchmark_Comparison_of_Remaining_Time_Prediction_Methods_in_Business_Process_Monitoring (date of access: 15.09.2025).
4. A survey on multi-output regression / H. Borchani [et al.] // *WIREs Data Mining and Knowledge Discovery*. 2015. Vol. 5, iss. 5. P. 216–232. URL: <https://cig.fi.upm.es/wp-content/uploads/2024/01/Borchani-A-survey-on-multi-output-regression.pdf> (date of access: 18.09.2025).
5. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting / B. Lim [et al.]. 2019. URL: <https://arxiv.org/abs/1912.09363> (date of access: 18.09.2025).
6. Attention Is All You Need. NeurIPS / A. Vaswani [et al.] // *Advances in Neural Information Processing Systems 30 (NIPS 2017)* / eds.: I. Guyon [et al.]. 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (date of access: 19.09.2025).
7. Verbeek E., Fahland D. CPN IDE: An Extensible Replacement for CPN Tools That Uses Access / CPN // *Proceedings of the ICPM Doctoral Consortium and Demo Track 2021 co-located with 10th International Conference on Process Mining (ICPM 2021)* / eds.: M. Jans [et al.]. CEUR-WS.org, 2021. P. 29–30 URL: <https://icpmconference.org/2021/wp->

[content/uploads/sites/5/2021/10/F-CPN-IDE-An-extensible-replacement-for-CPN-Tools-that-uses-AccessCPN.pdf](#) (date of access: 18.09.2025).

8. Recurrent neural network based language model / T. Mikolov [et al.] // Proc. Interspeech. 2010. Vol. 2, no. 3. P. 1045–1048. URL: https://www.fit.vut.cz/research/group/speech/public/publi/2010/mikolov_interspeech2010_IS100722.pdf (date of access: 18.09.2025).

9. Time2Vec: Learning a Vector Representation of Time / S. M. Kazemi [et al.]. 2019. URL: <https://arxiv.org/abs/1907.05321> (date of access: 18.09.2025).

10. Shaw P., Uszkoreit J., Vaswani A. Self-Attention with Relative Position Representations. 2018. URL: <https://arxiv.org/abs/1803.02155> (date of access: 19.09.2025).

11. Mei H., Eisner J. The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process // Advances in Neural Information Processing Systems 30 (NIPS 2017) / eds.: I. Guyon [et al.]. 2017. URL: <https://www.cs.jhu.edu/~jason/papers/mei+eisner.arxiv17.pdf> (date of access: 19.09.2025).

12. Koenker R. Quantile Regression. Cambridge Univ. Press, 2010. DOI: [10.1017/CBO9780511754098](https://doi.org/10.1017/CBO9780511754098).

13. Vovk V., Gammerman A., Shafer G. Algorithmic Learning in a Random World // Springer. 2005. URL: <https://download.e-bookshelf.de/download/0000/0004/78/L-G-0000000478-0002330810.pdf> (date of access: 19.09.2025).