

# Информатика и её применения

Том 17 Выпуск 3 Год 2023

## СОДЕРЖАНИЕ

О формировании множеств прецедентов на основе таблиц разнородных признаков описаний методами топологической теории анализа данных <b>И. Ю. Торшин</b>	2
Оптимальная фильтрация состояния нелинейной динамической системы по наблюдениям со случайными запаздываниями <b>А. В. Босов</b>	8
Рынок с марковской скачкообразной волатильностью II: алгоритм вычисления справедливой цены деривативов <b>А. В. Борисов</b>	18
Моделирование настойчивого поведения пользователей в сетях 5G NR с адаптацией скорости и блокировками <b>Э. С. Сопин, А. Р. Маслов, В. С. Шоргин, В. О. Бегишев</b>	25
Анализ загрузки многопользовательской сети при расщеплении потоков по кратчайшим маршрутам <b>Ю. Е. Малашенко, И. А. Назарова</b>	33
Оптимизация схемы распределения буферной памяти узла пакетной коммутации <b>Я. М. Агаларов</b>	39
О скорости сходимости и предельных характеристиках для одного обобщенного процесса рождения и гибели <b>И. А. Усов, Я. А. Сатин, А. И. Зейфман</b>	49
Метод оценивания параметров гамма-экспоненциального распределения по выборке со слабо зависимыми компонентами <b>А. А. Кудрявцев, О. В. Шестаков</b>	58
Логические методы корректной классификации данных <b>Е. В. Дюкова, Г. О. Масляков, А. П. Дюкова</b>	64
Классификация с помощью причинно-следственных связей <b>А. А. Грушо, Н. А. Грушо, М. И. Забежайло, Д. В. Смирнов, Е. Е. Тимонина</b>	71
О кластеризации объектов сетевой вычислительной инфраструктуры на основе анализа статистических аномалий в трафике <b>А. К. Горшенин, С. А. Горбунов, Д. Ю. Волканов</b>	76
Исследование эффективности применения бинарных нейронных сетей при детектировании объекта на изображении <b>Д. О. Королев, О. Г. Малеев</b>	88
Формализованное описание статистической обработки информации в базах данных <b>В. В. Вакуленко, И. М. Зацман</b>	93
Критерии определения семантической близости дискурсивных отношений <b>О. Ю. Инькова, М. Г. Кружков</b>	100
Трансформация иерархии Акоффа в научной парадигме информатики <b>И. М. Зацман</b>	107
Об авторах	114
Правила подготовки рукописей	116
Requirements for manuscripts	119

# О ФОРМИРОВАНИИ МНОЖЕСТВ ПРЕЦЕДЕНТОВ НА ОСНОВЕ ТАБЛИЦ РАЗНОРОДНЫХ ПРИЗНАКОВЫХ ОПИСАНИЙ МЕТОДАМИ ТОПОЛОГИЧЕСКОЙ ТЕОРИИ АНАЛИЗА ДАННЫХ\*

И. Ю. Торшин<sup>1</sup>

**Аннотация:** Факторизация вкладов различных переменных при анализе разнородных признаков описаний — насущная задача интеллектуального анализа сложных данных. В работе предложено развитие решеточного формализма топологической теории анализа данных, в рамках которого получены новые способы порождения параметрических оценок и метрик на решетках, образованных над топологиями множеств объектов. Формализм был апробирован на задаче формирования множеств прецедентов для проведения хеомикробиомного анализа. Тогда как порождение множества исходных информации на основе регрессионных коэффициентов и разности значений материала обучения соответствовало крайне низкой обобщающей способности настраиваемых алгоритмов (коэффициент корреляции на контроле —  $0,32 \pm 0,20$ ), использование предлагаемых оценок для порождения множеств прецедентов в задачах хеомикробиомики позволило существенно повысить обобщающую способность соответствующих алгоритмов (коэффициент корреляции на контроле —  $0,79 \pm 0,21$ ).

**Ключевые слова:** топологический анализ данных; теория решеток; параметризация решеточных термов; микробиом человека; фармакоинформатика, алгебраический подход Ю. И. Журавлёва

**DOI:** 10.14357/19922264230301

**EDN:** AQEUYO

## 1 Введение

В биомедицинских исследованиях объектом служит формализованное описание состояния пациента, включающее булевы (диагнозы, прием лекарств и др.), числовые (лабораторные анализы) и категориальные (показатели демографии и др.) переменные, графы (формулы лекарств), временные ряды и изображения (результаты обследования пациента аппаратными методами). При формализации таких задач важно выделить независимые вклады переменных-признаков в целевую переменную (отклик) таким образом, чтобы получить высокое качество работы алгоритмов распознавания/классификации. Топологический анализ данных, развиваемый в рамках алгебраического подхода к распознаванию научной школы Ю. И. Журавлёва и К. В. Рудакова [1, 2], позволяет систематически исследовать возможные решения этой задачи.

В настоящей работе данный подход апробирован на задаче формирования множества прецедентов для проведения хеомикробиомного анализа лекарств [3]. В этой прикладной задаче (оценка влияния лекарств на микробиоту) факт использования лекарства, как правило, описан булевым признаком, а откликом служит уровень той или иной

бактерии микробиома (числовая переменная). Для выделения независимых вкладов булевых переменных в числовые целевые переменные предложен системный подход к порождению параметризованных оценок на решетке (решеточных термов) и соответствующих метрик. Такой подход представляет собой теоретическое обобщение «расщепления» эмпирической функции распределения булевым признаком [2] и тесно связан с порождением метрических функций расстояния и с проблематикой так называемых «оценок информативности» (точнее, весовых функций признаков), используемых в комбинаторной теории разрешимости [4].

## 2 Основные понятия

Основы формализма изложены в [2, 5]. Задано множество исходных описаний объектов  $\mathbf{X} = \{x_1, \dots, x_{N_0}\}$ ,  $\mathbf{X} \subseteq S$ , множества значений признаков  $I_k = \{\lambda_{k_1}, \lambda_{k_2}, \dots, \lambda_{k_b}, \dots, \lambda_{k_{|I_k|-1}}, \Delta\}$ ,  $b = 1, \dots, |I_k|$ , функции  $\Gamma_k : S \rightarrow I_k$ ,  $k = 1, \dots, n + l$  ( $n$  — число признаков;  $l$  — число целевых (прогнозируемых) переменных;  $\Delta$  — неопределенность). Определено пространство допустимых признаков описаний объектов  $J_{\text{об}} \subseteq I_i \times I_f$  ( $I_i \subseteq I_1 \times \dots \times I_n$ ,  $I_f \subseteq I_{n+1} \times \dots \times I_{n+l}$ ),

\* Работа выполнена при поддержке гранта РНФ (проект 23-21-00154) с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

<sup>1</sup> Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук, tiy135@yahoo.com

функции  $D : S \rightarrow J_{ob}$ ,  $D(x_\alpha) = (\Gamma_1(x_\alpha) \times \dots \times \Gamma_k(x_\alpha) \times \dots \times \Gamma_{n+1}(x_\alpha))_\Delta$  и  $\varphi(\mathbf{X}) = \{D(x_\alpha) | x_\alpha \in \mathbf{X}\}$ . Принимается, что множества  $\mathbf{X}$  и  $Q = \varphi(\mathbf{X})$ ,  $|Q| = N$ , регулярны, т. е.  $N_0 = N$  и  $\exists D^{-1} : \forall x \in \mathbf{X}$ ,  $x = D^{-1}(D(x))$  [5]. Множество  $U(\mathbf{X}) = \{\Gamma_k^{-1}(\lambda_{k_b})\}$ , образованное функциями полных прообразов значений признаков  $\lambda_{k_b}$ , рассматривается как предбаза топологии  $T(\mathbf{X}) = \{\emptyset, I, a \cup b, a \cap b : a, b \in U(\mathbf{X})\}$ , где  $I = \{\mathbf{X}\}$ . Топологии  $T(\mathbf{X})$  соответствует решетка  $L(T(\mathbf{X})) = \{a \vee b, a \wedge b : a, b \in T(\mathbf{X})\}$ . При регулярности  $\mathbf{X}$  и  $Q$  решетка  $L(T(\mathbf{X}))$  — булева [2].

**Теорема 1.** Произвольной цепи  $\langle u_1, \dots, u_m \rangle$  решетки  $L(T(\mathbf{X}))$  можно сопоставить числовой признак с множеством значений  $I_t = (\lambda_{t_1}, \dots, \lambda_{t_b}, \dots, \lambda_m)$ ,  $\lambda_{t_{i-1}} \leq \lambda_{t_i} \leq \lambda_{t_{i+1}}$ .

Справедливость теоремы следует из того, что линейный порядок любой конечной цепи  $\langle u_1 \supseteq u_2 \dots \supseteq u_i \dots \supseteq u_m \rangle$  изоморфен линейному порядку на конечном множестве чисел  $\{\lambda_{t_i}\}$ ,  $\lambda_{t_{i-1}} \leq \lambda_{t_i}$ ,  $i = 2, \dots, m$ .

**Следствие 1.** Любая цепь  $A_t$  в  $L(T(\mathbf{X}))$  представима в виде  $A_t = \langle u(\lambda_{t_1}), \dots, u(\lambda_{t_i}), \dots, u(\lambda_{t_m}) \rangle$ ,  $u(\lambda_{t_i}) = \bigcup_{\beta=1}^i \Gamma_t^{-1}(\lambda_{t_\beta})$ , где  $I_t = (\lambda_{t_1}, \dots, \lambda_m)$  монотонна.

**Следствие 2.** При дополнении наблюдаемой области  $I_t$  значений  $t$ -го числового признака неопределенностью  $\Delta$  и принятии соглашения  $\Gamma_t(\emptyset) = \Delta$  соответствующая цепь  $A_t$  — максимальная цепь решетки  $L(T(\mathbf{X}))$ .

**Следствие 3.** Для произвольной цепи  $t$ -го числового признака определена эмпирическая функция распределения  $\text{cdf}(\lambda, A_k(\mathbf{X})) = |u(\lambda_{k_b})|/N | \lambda_{k_{b-1}} \leq \lambda \leq \lambda_{k_b}$ .

### 3 Решеточные термы и операции над ними

В топологическом анализе данных булевой решетке сопоставляется метрическое пространство значений признаков  $M_L(L(T(\mathbf{X})), \rho_L)$  с метрикой  $\rho_L : L^2 \rightarrow R^+$  [2]. Метрики  $\rho_L$  могут определяться на основании формализма оценок (см. ниже) или же на основе известных подходов (расстояния Танимото, Амана, Тверского, Сокала—Сниса, Гоуэра—Лежандра и др.) [6]. При заданной  $\rho_L$  расстояние  $\rho_A : \mathbf{A}(\mathbf{X})^2 \rightarrow R^+$  между цепями  $a = \langle a_1, \dots, a_i, \dots \rangle$  и  $b = \langle b_1, \dots, b_j, \dots \rangle$  определено так, что существует однозначное соответствие элементов цепей  $a$  и  $b$  (например,  $\rho_A(a, b) = \min(\sum_{i=1, |a|} \rho_L(a_i, \arg \min_{b_j \in b} \rho_L(a_i, b_j)), \sum_{i=1, |b|} \rho_L(b_j, \arg \min_{a_i \in a} \rho_L(b_j, a_i))$ ). При заданных  $\rho_L$ ,  $\rho_A$ ,  $A_k(\mathbf{X})$  и множества допустимых цепей  $\mathbf{A}(\mathbf{X})_{1,n}$  (содержит цепи с длинами от 1 до  $n$ )

искомый ( $\varepsilon$ -корректный) алгоритм соответствует решению задачи комбинаторной оптимизации [5]:

$$\arg \min_{a \in \mathbf{A}(\mathbf{X})_{1,n}} \rho_A(A_k(\mathbf{X}), a) | A(\mathbf{X})_{1,n} \subset \mathbf{A}(\mathbf{X}).$$

При порождении метрических пространств над произвольной решеткой вводится понятие решеточного терма или оценки  $v : L \rightarrow R^+$ , для которой выполнено условие оценки (**уО**:  $\forall_L a, b : v[a] + v[b] = v[a \wedge b] + v[a \vee b]$ ) и условие изотонности (**уИ**:  $\forall_L a, b : a \supseteq b \Rightarrow v[a] \geq v[b]$ ). Изотонность оценки  $v[\ ]$  важна потому, что позволяет определить метрику  $\rho(x, y) = v[x \vee y] - v[x \wedge y]$  [2].

**Теорема 2.** Линейная комбинация  $v[\ ] = \sum \omega_i v_i[\ ]$ ,  $i = 1, \dots, m$ , произвольного числа изотонных оценок  $v_i[\ ]$  — изотонная оценка при  $v[\ ] \geq 0$ .

Теорема доказывается через линейные преобразования условий **уО**<sub>*i*</sub> и **уИ**<sub>*i*</sub> для  $m$  оценок  $v_i[\ ]$ .

**Следствие 1.** Сумма произвольного числа (изотонных) оценок — оценка.

**Следствие 2.** Разность (изотонных) оценок становится оценкой только при условии положительной определенности.

**Следствие 3.** Пусть индекс  $i$  в выражении для  $v[\ ]$  изменяется от  $i = 0$ ,  $v_0 \equiv 1$ . Тогда можно подобрать такое значение  $\omega_0$ , что  $v[\ ] \geq 0$ .

При произвольном нелинейном преобразовании  $f : R^+ \rightarrow R^+$ , применяемом к (изотонной) оценке  $v[\ ]$ , выполнимость **уО** неочевидна. Если  $v[\ ]$  изотонна, то  $f(v[\ ])$  может быть изотонна только при условии монотонности  $f$  (сигмоида, степенная функция и др.). Для оценки выполнимости **уО**/**уИ**  $f(v[\ ])$  на подмножестве решетки  $L$  возможно применение аналитических или комбинаторных подходов. Например, легко показать, что для  $f(x) = x^\alpha$  при  $\alpha = 2$  или  $0,5$  выражение  $f(v[\ ])$  служит оценкой только на элементах произвольной цепи решетки  $L(T(\mathbf{X}))$ .

### 4 Параметрические оценки на основе «опорного элемента»

Выберем подмножество объектов  $\mathbf{X}$ ,  $\alpha \in L(T(\mathbf{X}))$ . Подмножество  $\alpha$  может соответствовать  $k'$ -му булеву признаку в исходном признаковом описании (тогда  $\alpha = \Gamma_{k'}^{-1}(1)$ ) или синтетическому булеву признаку, что разбивает  $\mathbf{X}$  на  $\alpha$  и  $\bar{\alpha} = \mathbf{X} \setminus \alpha$  и определяет  $\nu_\alpha = |\alpha|/|\mathbf{X}| = |\alpha|/N$  и  $\nu_{\bar{\alpha}} = 1 - \nu_\alpha$ .

«Опираясь» на множество  $\alpha$  как параметр, можно породить несколько изотонных решеточных оценок на основе базовой оценки, равной высоте элемента  $u_i$  в  $L(T(\mathbf{X}))$ ,  $h[u_i] = |u_i|$ , частот  $\nu_\alpha$

и  $\nu_{\bar{\alpha}}$ . Оценка  $v_{\alpha}^{+}[u_i] = |u_i \cap \alpha|/|\alpha|$  соответствует частоте встречаемости элементов из  $u_i \subseteq \mathbf{X}$  в  $\alpha \subset \mathbf{X}$ ; оценка  $v_{\bar{\alpha}}^{-}[u_i] = |u_i \cap \bar{\alpha}|/|\bar{\alpha}| = |u_i \setminus \alpha|/(N - |\alpha|)$  — встречаемость элементов  $u_i$  в  $\bar{\alpha}$ . В соответствии с теоремой 2 линейные комбинации оценок  $v_{\alpha}^{+}$  и  $v_{\bar{\alpha}}^{-}$  также изотонны при условии положительной определенности.

**Теорема 3.** *Изотонная оценка  $d_{\alpha}[\cdot] = v_{\alpha}^{+}[\cdot] - v_{\bar{\alpha}}^{-}[\cdot]$  существует, когда объекты из множества  $\alpha$  встречаются в оцениваемых множествах  $u_i$  не реже, чем в среднем по множеству всех объектов  $\mathbf{X}$ .*

Рассмотрим функционал  $d_{\alpha}[u_i] = v_{\alpha}^{+}[u_i] + \omega v_{\bar{\alpha}}^{-}[u_i]$ ,  $\omega \in R$ , и покажем, что  $d_{\alpha}[\cdot]$  — линейная комбинация метрики  $\rho(u_i, \alpha)$  и оценки  $h[u_i] = |u_i|$ . Приведем  $v_{\alpha}^{+}$  и  $v_{\bar{\alpha}}^{-}$  к общему знаменателю, равному  $|\alpha|(N - |\alpha|)$ . Подставляя

$$|u_i \cap \alpha| = 0,5(|u_i| + |\alpha| - \rho(u_i, \alpha)),$$

после упрощения получим (для произвольного  $u_i$ ), что

$$d_{\alpha}[u_i] = k_{\alpha}|\alpha| + b_{\alpha}|u_i| - k_{\alpha}\rho(u_i, \alpha),$$

где

$$b_{\alpha} = \frac{1/v_{\alpha} + \omega - 1}{2(N - |\alpha|)};$$

$$k_{\alpha} = \frac{1/v_{\alpha} - \omega - 1}{2(N - |\alpha|)}.$$

Рассмотрим функционал  $d_{\alpha}[\cdot]$  (который служит изотонной оценкой для всех  $\omega > 0$ ) в случае  $\omega < 0$  и запишем его в виде  $d_{\alpha}[\cdot] = v_{\alpha}^{+}[\cdot] - |\omega|v_{\bar{\alpha}}^{-}[\cdot]$ . Условию  $d_{\alpha}[u_i] \geq 0$  соответствует

$$\frac{1/v_{\alpha} + |\omega| - 1}{N - |\alpha|} |u_i \cap \alpha| \geq \frac{|\omega||u_i|}{N - |\alpha|},$$

т. е.

$$\frac{|i_i \cap \alpha|}{|u_i|} \geq \frac{|\omega|}{1/v_{\alpha} + |\omega| - 1}.$$

Это условие выделяет подмножество  $\{u_i\} \subset L(T(\mathbf{X}))$ , где  $d_{\alpha}[\cdot]$  изотонна. При  $|\omega| \sim 1$  получаем  $|u_i \cap \alpha|/|u_i| \geq \nu_{\alpha}$ , соответствующее условию теоремы. Теорема доказана.

**Следствие 1.**

$$d'_{\alpha}[u_i] = \begin{cases} v'_{\alpha}[u_i] - v'_{\bar{\alpha}}[u_i] & \text{при } \frac{|u_i \cap \alpha|}{|u_i|} \geq \nu_{\alpha}; \\ 0 & \text{в противном случае} \end{cases}$$

есть изотонная оценка.

Функционал  $d_{\alpha}[u_i]$  интересен тем, что может быть увязан с конструктами непараметрической статистики, разработанными в научной школе

А. Н. Колмогорова [7]. Как известно, для непараметрических тестов используется так называемое максимальное уклонение  $D$ , равное супремуму модуля разности значений двух функций распределения (теоретической и эмпирической или двух эмпирических функций распределения) по всей их области определения.

**Теорема 4.** *Для «опорного» множества  $\alpha$  и произвольной цепи  $U = \langle u_1, \dots, u_i, \dots \rangle$*

$$D = \sup |d_{\alpha}[u_i]|$$

при  $\omega = -1$  — максимальное уклонение Колмогорова.

Доказательство проводится исходя из теоремы 1 для любой цепи  $A = \langle u_i \rangle$ ,  $u_i = u(\lambda_{k_i})$  и анализа числового признака с множеством значений  $I_k$  в двух цепях, образованных пересечением каждого элемента  $u_i$  цепи  $A$  с множествами  $\alpha$  и  $\bar{\alpha}$  соответственно.

**Следствие 1.** *Статистическая достоверность значения  $D$  оценивается по критерию Колмогорова—Смирнова.*

Оценки  $v_{\alpha}^{+}$ ,  $v_{\bar{\alpha}}^{-}$ ,  $d_{\alpha}$  и др. могут строиться для набора «опорных» множеств. Пусть задан набор  $\alpha \subset L(T(\mathbf{X}))$  подмножеств объектов (т. е. элементов решетки):  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_i \dots\}$ . Каждое из множеств в наборе  $\alpha$  порождает (изотонную) параметрическую оценку  $v_{\alpha_i}$  (например,  $v_{\alpha_i}^{+}$ ,  $v_{\alpha_i}^{-}$ , их линейные комбинации). Тогда в соответствии с теоремой 2 может быть порождена и настраиваемая изотонная оценка

$$v_{\alpha} = \sum_{i=0,|\alpha|} \omega_i v_{\alpha_i}, \quad v_{\alpha_0} \equiv 1.$$

Альтернативно на основании набора  $\alpha$  для каждого  $a \in L(T(\mathbf{X}))$  может быть вычислен вектор оценок  $(v_{\alpha_1}[a], v_{\alpha_2}[a], \dots, v_{\alpha_i}[a], \dots)$ ,  $v_{\alpha_i}[a] \in R^+$ , и введенные метрики уже на пространстве таких векторов ( $l_p$ -метрики, расстояния Пенроуза, Мотыки, Брея—Куртиса, корреляционные расстояния) [6]. Формирование наборов  $\alpha$  может проводиться на основе «информативности» соответствующих  $\alpha_i \in \alpha$  методами метрического анализа данных и др. [4].

## 5 О параметрических оценках на основе «опорной цепи»

Крайне перспективным представляется направление порождения оценок, в котором заданная максимальная цепь  $A_t$  используется в качестве «опорной», т. е. для порождения параметрических

оценок. Данное направление будет подробно рассмотрено в отдельной работе. Вкратце: произвольная цепь  $A = \langle u_i \rangle$  представима в виде  $\langle u(\lambda_{t_i}) \rangle$  для некоторого упорядоченного множества чисел  $I_t$  (теорема 1). Значение функции  $\Gamma_t$  (включая неопределенность), вычисляемое для любого объекта в  $\mathbf{X}$ , равно  $\Gamma_t(q)$  для каждого решеточного атома  $\{q\} \in L(T(\mathbf{X}))$ , ( $h[\{q\}] \equiv |\{q\}| \equiv 1$ ), так что любому  $u \in L(T(\mathbf{X}))$  соответствует множество  $\Gamma_t(u) = \{\Gamma_t(q), q \in u\}$ . Определим оператор  $\hat{\phi}(x)$  для формирования эмпирической функции распределения конечного множества  $A = \{a_1, a_2, \dots, a_i, \dots, a_n\}$ ,  $a_i \in R$ , как

$$\hat{\phi}(x)A = \sup \frac{|\{B \subseteq A | \forall a \in B : a \leq x\}|}{|A|}, \quad x \in R,$$

и оператор вычисления математического ожидания  $\hat{\mu}$ .

Итак, при задании опорной цепи  $A_t$  (т.е. опорного числового признака с областью значений  $I_t$ ) любому элементу решетки  $u \in L(T(\mathbf{X}))$  сопоставлено множество чисел  $\Gamma_t(u)$ , числовая функция  $\hat{\phi}(x)\Gamma_t(u)$  одной переменной  $x \in R$ , нетривиально определенная в каждой точке отрезка  $[\lambda_{t_1}, \lambda_{t_{|I_t|-1}}]$ , и ряд функционалов (в частности,  $\hat{\mu}\hat{\phi}(x)\Gamma_t(u)$ ). При выполнении условия регулярности для  $\mathbf{X}/Q$  решетка  $L(T(\mathbf{X}))$  однозначно сопоставлена решетке, образованной числовыми множествами  $\Gamma_t(u)$ , для каждого элемента которой вычислима функция  $\hat{\phi}\Gamma_t(u)$ . Множества  $\Gamma_t(u)$ , функции  $\hat{\phi}\Gamma_t(u)$  и функционалы наподобие  $\hat{\mu}\hat{\phi}\Gamma_t(u)$  могут быть использованы для определения оценок в решетке  $L(T(\mathbf{X}))$ ,

порождаемых на основе выбранного опорного признака.

## 6 Экспериментальная апробация формализма

Формализм был апробирован на задаче формирования множеств прецедентов для проведения хеомикробиомного анализа [3, 8]. В исходной таблице признаков описаний представлены булевы переменные, соответствующие 122 лекарствам, которые влияют на 1533 числовые переменные, соответствующие уровням отдельных бактерий или их групп.

Для квантификации вклада лекарства  $\alpha$  в изменение уровней  $t$ -й бактерии среди разработанных оценок использовалось уклонение  $D$  со знаком (теоремы 3 и 4). Для учета направления изменения уровней бактерий уклонение  $D$  необходимо домножить на знак  $d_\alpha[\cdot]$ . По тесту Колмогорова–Смирнова вычислялось значение статистической достоверности  $p(D)$  для отсекаемых наименее достоверных ассоциаций «лекарство–эффект». Также использовались более очевидные подходы: коэффициенты линейной многопараметрической регрессии (с отбором признаков по модели «лассо»), разность значений  $\hat{\mu}\hat{\phi}\Gamma_t(\alpha) - \hat{\mu}\hat{\phi}\Gamma_t(\bar{\alpha})$ , доля разности  $1 - \hat{\mu}\hat{\phi}\Gamma_t(\bar{\alpha})/\hat{\mu}\hat{\phi}\Gamma_t(\alpha)$ .

Лекарства с известной химической структурой описывались на основе хемографов  $G_j$ , а в качестве множества начальных информационных  $I_i$  использовалось множество хемоинвариан-

Результаты вычислительных экспериментов на выборке 156 327 измерений хеомикробиомной биологической активности (2173 пациентов). Оценены эффекты 122 лекарств. Приведены значения коэффициентов ранговой корреляции для различных моделей формирования множеств прецедентов. Эксперименты проводились в рамках кросс-валидационного дизайна (10 разбиений в соотношении «случай–контроль» 6 : 1). Поиск оптимальных значений параметров проводился мультистартовой стохастической оптимизацией

Вычислительный эксперимент	$n_{\text{акт}}$	$r$	$r_c$
Коэффициент многопараметрической регрессии, $p(D) = 0,20$	1250	0,72 ± 0,23	0,32 ± 0,20
Коэффициент многопараметрической регрессии, $p(D) = 0,05$	730	0,73 ± 0,39	0,38 ± 0,18
Разность значений, $p(D) = 0,20$	1192	0,74 ± 0,19	0,34 ± 0,22
Разность значений, $p(D) = 0,05$	787	0,78 ± 0,36	0,39 ± 0,25
Доля разности значений, $p(D) = 0,20$	1292	0,80 ± 0,22	0,46 ± 0,14
Доля разности значений, $p(D) = 0,10$	1291	0,80 ± 0,27	0,62 ± 0,18
<b>Доля разности значений, <math>p(D) = 0,05</math></b>	<b>868</b>	<b>0,76 ± 0,37</b>	<b>0,77 ± 0,22</b>
Уклонение $D$ со знаком, $p(D) = 0,20$	1286	0,81 ± 0,21	0,48 ± 0,15
Уклонение $D$ со знаком, $p(D) = 0,10$	1306	0,80 ± 0,26	0,63 ± 0,18
<b>Уклонение <math>D</math> со знаком, <math>p(D) = 0,05</math></b>	<b>836</b>	<b>0,80 ± 0,38</b>	<b>0,79 ± 0,21</b>

Обозначения:  $r$  и  $r_c$  — коэффициенты ранговой корреляции на обучении и на контроле соответственно;  $n_{\text{акт}}$  — число различных типов хеомикробиомной активности, по которым проводилось усреднение  $r$  и  $r_c$ ;  $p(D)$  — верхний порог статистической значимости по тесту Колмогорова–Смирнова.

тов над алфавитом элементарных меток. Алгоритмы  $f_{\theta_k} : I_i \rightarrow R$  строились в виде композиций вложенных корректирующих функций нижнего уровня (т.е. порождения синтетических признаков) для фиксированного числа моделей  $n_{\text{mod}} : f_{\theta_k} = g(f_1(\sum \omega_k^j x_k), \dots, f_l(\sum \omega_k^j x_k), \dots)$ ,  $l = 1, \dots, n_{\text{mod}}$ , как в работе [5]. Тестирование алгоритмов  $f_{\theta_k}$  проводилось на выборке данных о пациентах (см. таблицу).

Результаты экспериментов показывают, что порождение множества исходных информации на основе регрессионных коэффициентов и разности значений материала обучения соответствует крайне низкой обобщающей способности настраиваемых алгоритмов. При этом ужесточение порога на значение  $p(D)$ , давая снижение числа различных типов хеомикробиомной активности, не приводило к увеличению качества алгоритмов. Наилучший результат был получен при использовании в качестве исходной информации значения уклонения  $D$  со знаком ( $r = 0,80 \pm 0,38$ ,  $r_c = 0,79 \pm 0,21$ ), а более низкие значения порога  $p(D)$  приводили к повышению качества распознавания. Отметим неплохой результат и для такого простого функционала, как доля разности значений (см. таблицу).

## 7 Заключение

В работе впервые проведено систематическое рассмотрение способов введения оценок на решетке (высота элемента, оценки на основании булевых и числовых переменных, линейные комбинации вышперечисленных оценок и др.). Введение оценок на  $L(X)$ , по аналогии с понятием меры в функциональном анализе, позволяет порождать параметрические решеточные оценки и затем вводить проблемно-ориентированные метрики, оценивающие расстояние между вершинами решетки. Разработанный формализм был применен для достижения практической цели настоящей статьи —

нахождения оптимального способа оценки вкладов переменных при анализе сложных данных хеомикробиомных исследований. Разработанный формализм предоставляет инструментарий для поиска адекватной формализации задач классификации и прогнозирования.

## Литература

1. Журавлёв Ю. И. Избранные научные труды. — М.: Магистр, 1998. 420 с.
2. Torshin I. Yu., Rudakov K. V. On the theoretical basis of metric analysis of poorly formalized problems of recognition and classification // Pattern Recognition Image Analysis, 2015. Vol. 25. No. 4. P. 577–587. doi: 10.1134/S1054661815040252.
3. Торшин И. Ю., Громова О. А., Захарова И. Н., Максимов В. А. Хеомикробиомный анализ Лактитола // Экспериментальная и клиническая гастроэнтерология, 2019. Т. 164. № 4. С. 111–121. doi: 10.31146/1682-8658-ecg-164-4-111-121.
4. Рудаков К. В., Торшин И. Ю. Анализ информативности мотивов на основе критерия разрешимости в задаче распознавания вторичной структуры белка // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 79–90.
5. Торшин И. Ю. О задачах оптимизации, возникающих при применении топологического анализа данных к поиску алгоритмов прогнозирования с фиксированными корректорами // Информатика и её применения, 2023. Т. 17. Вып. 2. С. 2–10. doi: 10.14357/19922264230201. EDN: IGSPWE.
6. Деца Е. И., Деца М. М. Энциклопедический словарь расстояний / Пер. с англ. — М.: Наука, 2008. 444 с. (Deza E., Deza M.-M. Dictionary of distances. — Elsevier B.V., 2006. 412 p.)
7. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. — М.: Наука, 1989. 624 с.
8. Forslund S. K., Chakaroun R., Stumvoll M., Bork P. Combinatorial, additive and dose-dependent drug-microbiome associations // Nature, 2021. Vol. 600. No. 7889. P. 500–505. doi: 10.1038/s41586-021-04177-9.

Поступила в редакцию 02.02.23

# ON THE FORMATION OF SETS OF PRECEDENTS BASED ON TABLES OF HETEROGENEOUS FEATURE DESCRIPTIONS BY METHODS OF TOPOLOGICAL THEORY OF DATA ANALYSIS

I. Yu. Torshin

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** Factorization of the contributions of various variables in the analysis of heterogeneous feature descriptions is an urgent task of complex data mining. The paper proposes the development of the lattice formalism of the

topological theory of data analysis, within which new methods for generating parametric estimates and metrics on lattices formed over the topologies of sets of objects are obtained. The formalism was tested on the problem of forming sets of precedents for conducting chemomicrobiome analysis. Whereas the generation of a set of initial information based on regression coefficients and the difference in the values of the learning material corresponded to an extremely low generalizing ability of custom algorithms (correlation coefficient in the control  $0.32 \pm 0.20$ ), the use of the proposed estimates for generating sets of precedents in chemomicrobiomics problems made it possible to significantly increase the generalizing ability of the corresponding algorithms (correlation coefficient in control  $0.79 \pm 0.21$ ).

**Keywords:** topological data analysis; lattice theory; parametrization of lattice terms; human microbiome; pharmacoinformatics, algebraic approach of Yu. I. Zhuravlev.

**DOI:** 10.14357/19922264230301

**EDN:** AQEUYO

## Acknowledgments

The research was funded by the Russian Science Foundation, project No. 23-21-00154. The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (CKP “Informatics”) of FRC CSC RAS (Moscow).

## References

1. Zhuravlev, Yu. I. 1998. *Izbrannyye nauchnyye trudy* [Selected scientific works]. Moscow: Magistr. 420 p.
2. Torshin, I. Y., and K. V. Rudakov. 2015. On the theoretical basis of metric analysis of poorly formalized problems of recognition and classification. *Pattern Recognition Image Analysis* 25(4):577–587. doi: 10.1134/S1054661815040252.
3. Torshin, I. Yu., O. A. Gromova, I. N. Zakharova, and V. A. Maksimov. 2019. Khemomikrobiomnyy analiz Laktitola [Hemomikrobiomny lactitol analysis]. *Ekspierimental'naya i klinicheskaya gastroenterologiya* [Experimental and Clinical Gastroenterology]. 164(4):111–121. doi: 10.31146/1682-8658-ecg-164-4-111-121.
4. Rudakov, K. V., and I. Yu. Torshin. 2012. Analiz informativnosti motivov na osnove kriteriya razreshimosti v zadache raspoznavaniya vtorichnoy struktury belka [Analysis of the informativeness of motives based on the criterion of solvability in the problem of recognizing the secondary structure of a protein]. *Informatika i ee Primeneniya — Inform Appl.* 6(1):79–90.
5. Torshin, I. Yu. 2023. O zadachakh optimizatsii, voznikayushchikh pri primenenii topologicheskogo analiza dannykh k poisku algoritmov prognozirovaniya s fiksirovannyimi korrektorami [On optimization problems arising from the application of topological data analysis to the search for forecasting algorithms with fixed correctors]. *Informatika i ee Primeneniya — Inform Appl.* 17(2):2–10. doi: 10.14357/19922264230201. EDN: IGSPEW.
6. Deza, E., and M.-M. Deza. 2006. *Dictionary of distances*. Elsevier B.V. 412 p.
7. Kolmogorov, A. N., and S. V. Fomin. 1989. *Elementy teorii funktsiy i funktsional'nogo analiza* [Elements of the theory of functions and functional analysis]. Moscow: Nauka. 624 p.
8. Forslund, S. K., R. Chakaroun, M. Stumvoll, and P. Bork. 2021. Combinatorial, additive and dose-dependent drug-microbiome associations. *Nature* 600(7889):500–505. doi: 10.1038/s41586-021-04177-9.

Received February 2, 2023

## Contributor

**Torshin Ivan Y.** (b. 1972) — Candidate of Science (PhD) in physics and mathematics, Candidate of Science (PhD) in chemistry, senior scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; tiy135@yahoo.com

# ОПТИМАЛЬНАЯ ФИЛЬТРАЦИЯ СОСТОЯНИЯ НЕЛИНЕЙНОЙ ДИНАМИЧЕСКОЙ СИСТЕМЫ ПО НАБЛЮДЕНИЯМ СО СЛУЧАЙНЫМИ ЗАПАЗДЫВАНИЯМИ\*

А. В. Босов<sup>1</sup>

**Аннотация:** Изучается математическая модель нелинейной динамической системы наблюдения с дискретным временем, позволяющая учитывать зависимость времени получения наблюдений от состояния наблюдаемого объекта. Модель реализует предположение о том, что время между моментом, когда формируется измерение состояния, и моментом получения измеренного состояния наблюдателем зависит случайным образом от положения движущегося объекта. Источником такого предположения выступает процесс наблюдения стационарными средствами за автономным подводным аппаратом, в котором время получения актуальных данных зависит от неизвестного расстояния между объектом и наблюдателем. В отличие от детерминированных задержек, формируемых известным состоянием среды наблюдения, для учета зависимости временных задержек от неизвестного состояния объекта наблюдения требуется использовать для их описания случайные функции. Основным результатом исследования предложенной модели стало решение задачи оптимальной фильтрации. Для этого получены рекуррентные байесовские соотношения, описывающие эволюцию апостериорной плотности вероятности. Использование полученного фильтра для практических целей не представляется возможным из-за вычислительной сложности. Предложенная модель проиллюстрирована практическим примером задачи слежения за движущимся подводным объектом по результатам измерений, выполняемых типовыми акустическими сенсорами. Предполагается, что объект движется под водой в плоскости с известной средней скоростью, постоянно выполняет хаотические маневры и наблюдается двумя независимыми комплексами акустических сенсоров, измеряющими дальности до объекта и направляющие косинусы. Сложность определения положения такого объекта иллюстрируется простым фильтром, использующим геометрические свойства измеряемых величин.

**Ключевые слова:** стохастическая динамическая система наблюдения; фильтрация состояния; оптимальный байесовский фильтр; среднеквадратичный критерий оценивания; автономный подводный аппарат; акустический сенсор; слежение за целью

DOI: 10.14357/19922264230302

EDN: CFVYJM

## 1 Введение

Задача слежения за маневрирующим объектом по косвенным наблюдениям с ошибками представляет собой типичное приложение теории и методов стохастической фильтрации [1]. Предложение учитывать в модели наблюдений случайные задержки времени их получения от состояния наблюдаемого объекта исходит от популярной современной прикладной области автономных подводных аппаратов (Autonomous Underwater Vehicle, AUV) [2]. Приложениям AUV посвящено значительное число работ, в большинстве связанных с задачами управления (см., например, специальный выпуск [3]).

Вместе с тем и традиционные задачи слежения за маневрирующим объектом сохраняют актуальность [4–11], а их решение обеспечивают методы стохастической фильтрации.

Наибольшее практическое распространение здесь, как и в других приложениях, получили модель и алгоритм фильтрации Калмана [12] и основанные на нем различные субоптимальные фильтры, начиная с расширенного фильтра Калмана [13] и вплоть до концепций дезодорированного или кубатурного фильтров Калмана [14–16]. Исследования и усовершенствования таких субоптимальных фильтров продолжают и по настоящее время, в том числе с применением к слежению за AUV [17].

Однако принципиально новых идей уже не предлагается, а изначально имевшиеся недостатки так и остались. Суть этих недостатков состоит в том, что ни для какого фильтра нельзя гарантировать точностные характеристики для сколь-нибудь широкого класса моделей, т. е. ни несмещенности оценки, ни ограниченности ошибки хотя бы дисперсией оцениваемого состояния гарантировать

\* Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук; ABosov@frccsc.ru



нельзя, а значит, всегда есть опасность, что в другой модели фильтр будет неустойчивым и оценка будет расходиться. Единственным подтверждением свойств субоптимального фильтра был и остается практический эксперимент с конкретной моделью и фиксированным набором параметров.

Надо отметить, что в обсуждаемой прикладной области и проблема запаздывающих наблюдений хорошо известна. Временные задержки обязаны известному свойству акустических измерителей, а именно: зависимости скорости распространения акустической волны от температуры, солености и давления воды [18]. Именно это привлекло внимание в [11], и решение было найдено путем объединения данных измерений акустических датчиков с информацией от других датчиков в бортовой инерциальной навигационной системе. Более типичный вариант борьбы с запаздывающими наблюдениями состоит в том, чтобы оценивать время запаздывания и учитывать его, корректируя обычный фильтр [19]. Интересно, что запаздывание сигнала объясняется не только природой подводных наблюдений, но присутствует и в химических процессах [20].

Следует учитывать, что модели с детерминированным временем запаздывания имеют ограниченное применение в связи с тем, что в реальности время запаздывания случайно и существенно зависит от условий наблюдения, а главное, оно меняется со временем. Эти изменения могут быть существенны. В модели, которая используется в статье, время запаздывания наблюдений описывается случайным процессом, который является известной функцией состояния системы. Это делает модель максимально адекватной при описании результатов работы акустических сенсоров. При этом формально модель системы наблюдения остается моделью марковского процесса с дискретным временем, к ней может быть применена классическая процедура байесовской фильтрации [21], т.е. записаны рекуррентные соотношения для апостериорной плотности вероятности. Эти соотношения составляют основной теоретический результат статьи. Но этот формальный успех оказывается полезным лишь для того, чтобы обосновать невозможность практической реализации ни самого оптимального фильтра, ни его субоптимальных упрощений. Причины этого состоят в том, что формальное приведение уравнений системы наблюдения к традиционному виду чудовищно увеличивает размерность. Поэтому полученные ниже уравнения оптимальной фильтрации следует рассматривать только как однозначное основание для поиска других, более ориентированных на реализацию методов. Убедиться в том, что такие методы действительно нужны, позволя-

ет описанный в последнем разделе статьи численный эксперимент. Простейшая модель движения AUV в плоскости с известной средней скоростью и постоянными хаотическими маневрами плюс типовые наблюдатели, измеряющие дальности и направляющие косинусы до объекта, позволяют без труда записать «естественный» фильтр, используя геометрическую связь измеряемых величин и координат объекта. Такой фильтр в предложенной модели приводит к неприемлемому падению качества оценивания по сравнению с оцениванием в системе, не предполагающей временные задержки.

## 2 Система наблюдения со случайными временными задержками

В модели используется дискретное время. Предполагается, что процесс фильтрации начинается в момент  $t = 0$ , т.е. в этот момент вычисляется первая оценка состояния. При этом наблюдения поступают с временной задержкой, максимально возможная задержка известна и равна  $T$ , поэтому состояние системы начинает формироваться в моменты  $t = -T, -T + 1, \dots, 0, 1, \dots$  и начальное состояние задается в момент  $-T - 1$ .

Уравнения состояния и наблюдений записываются в самой типичной форме с аддитивными, не зависящими от состояния шумами. Именно такую модель системы наблюдения использует большинство субоптимальных фильтров. Отличие от канонической записи состоит только в описании момента времени, в который было сформировано состояние, измеренное текущим наблюдением:

$$\left. \begin{aligned} x_t &= \varphi_t(x_{t-1}) + w_t, & x_{-T-1} &= \eta; \\ y_t &= \psi_t(x_{t-\tau_t}) + v_t, & \tau_t &= \theta_t(x_t). \end{aligned} \right\} \quad (1)$$

Здесь  $x_t \in \mathbb{R}^{p_x}$  — вектор состояния системы;  $w_t \in \mathbb{R}^{p_x}$  — дискретный белый шум, моделирующий возмущения;  $\eta \in \mathbb{R}^{p_x}$  — вектор начальных условий;  $y_t \in \mathbb{R}^{q_y}$  — вектор косвенных наблюдений;  $v_t \in \mathbb{R}^{q_y}$  — дискретный белый шум, моделирующий ошибки измерений; векторы  $\eta$ ,  $w_t$  и  $v_t$  предполагаются независимыми в совокупности.

Процесс  $\tau_t$  моделирует задержки наблюдений, описывая их известной функцией  $\theta_t$  состояния  $x_t$ . Элементы вектора  $\tau_t \in \mathbb{R}^{q_y}$  задают временные задержки для соответствующих компонентов вектора  $y_t$ , каждый компонент моделируется случайной последовательностью, элементы которой — дискретные случайные величины со значениями

в множестве  $\{0, 1, \dots, T\}$ , т. е.  $(\tau_t)_i \in \{0, 1, \dots, T\}$ ,  $i = 1, \dots, q_y$ . Чтобы это определение было корректным, будем считать, что в (1) и всюду далее через  $x_{t-\tau_t}$  обозначен составной вектор, содержащий состояния  $x_t$ , со всеми сдвигами  $(\tau_t)_i$ , т. е.

$$x_{t-\tau_t} = \left( x'_{t-(\tau_t)_1}, \dots, x'_{t-(\tau_t)_{q_y}} \right)' \in \mathbb{R}^{q_y p_x}.$$

Рассматривается задача оценивания состояния  $x_t$  по наблюдениям  $y_s$ ,  $s = 0, 1, \dots, t$ , критерий точности оценки  $\tilde{x}_t$  среднеквадратичный:  $E\{\|x_t - \tilde{x}_t\|^2\}$ ,  $E\{x\}$  — математическое ожидание  $x$ ,  $\|x\|$  — обычная евклидова норма вектора  $x$ . Здесь использовано обозначение «'» для операции транспонирования.

Таким образом, единственное отличие поставленной задачи от традиционной задачи фильтрации состояния системы наблюдения в дискретном времени состоит в запаздывании наблюдений на случайное время  $\tau_t$ . Формально даже этого отличия нет, потому что в следующем разделе показано, как привести систему (1) к традиционной марковской форме записи. Поэтому оптимальное решение  $\hat{x}_t = \arg \min_{\tilde{x}_t} E\{\|x_t - \tilde{x}_t\|^2\}$  является условным математическим ожиданием  $x_t$  относительно наблюдений  $y_s$ ,  $s = 0, 1, \dots, t$ , а значит, для его вычисления достаточно знать апостериорную плотность вероятности  $x_t$  относительно  $y_s$ ,  $s = 0, 1, \dots, t$ . Далее будут выписаны апостериорные плотности для системы (1) и  $t = 0, 1, \dots$  в форме рекуррентных байесовских соотношений [21] в условиях, когда нужные плотности вероятности существуют, т. е. при определенных ограничениях на непрерывность нелинейных функций и возмущений в (1).

### 3 Оптимальная фильтрация

Немарковская система наблюдения (1) может быть записана в форме марковского процесса, имеющего расширенный вектор состояния. Формально этот расширенный процесс будет иметь тот же вид (1), но с  $\tau_t = 0$ .

Расширенный вектор состояния далее обозначается  $\mathbf{x}_t \in \mathbb{R}^{(T+1)p_x}$  и представляет собой составной вектор, включающий все состояния системы от момента времени  $t - T$  до текущего момента  $t$ , т. е.  $\mathbf{x}_t = (x'_{t-T}, \dots, x'_{t-1}, x'_t)'$ . Отметим, что этот вектор не нужно путать с составным вектором  $x_{t-\tau_t} \in \mathbb{R}^{q_y p_x}$ , использованным в записи (1). В  $x_{t-\tau_t}$  объединены те состояния, которые сдвинуты на фактические задержки  $\tau_t$  для каждого элемента измерений,  $\mathbf{x}_t$  объединяет состояния, сдвинутые на все возможные значения  $0, 1, \dots, T$  задержек. Уравнения для  $\mathbf{x}_t$  имеют вид:

$$\left. \begin{aligned} (\mathbf{x}_t)_{p_x+1}^{p_x} &= (\mathbf{x}_{t-1})_{p_x+1}^{2p_x}; \\ &\dots \\ (\mathbf{x}_t)_{(T-1)p_x+1}^{Tp_x} &= (\mathbf{x}_{t-1})_{Tp_x+1}^{(T+1)p_x}; \\ (\mathbf{x}_t)_{Tp_x+1}^{(T+1)p_x} &= \\ &= \varphi_t \left( (\mathbf{x}_{t-1})_{Tp_x+1}^{(T+1)p_x} \right) + w_t; \end{aligned} \right\} \begin{cases} x_{t-T} \\ \dots \\ x_{t-1} \\ x_t = \\ = \varphi_t(x_{t-1}) + w_t \end{cases},$$

где через  $(\mathbf{x})_i^j$  обозначен подвектор вектора  $\mathbf{x}$  с элементами от  $i$ -го до  $j$ -го.

Обозначив такое преобразование вектора состояния через  $\Phi_t$ , а соответствующий аддитивный шум — через  $\mathbf{w}_t = (0', \dots, 0', w'_t)'$ , получаем уравнение состояния в виде:

$$\mathbf{x}_t = \Phi_t(\mathbf{x}_t) + \mathbf{w}_t. \quad (2)$$

Для записи соотношения для наблюдателя определим матричную функцию  $\Psi_t(\mathbf{x}_t) \in \mathbb{R}^{(T+1)q_y}$  следующим образом:

$$\Psi_t(\mathbf{x}_t) = \begin{pmatrix} \psi'_t((\mathbf{x}_t)_{p_x+1}^{p_x}) \\ \psi'_t((\mathbf{x}_t)_{(T-1)p_x+1}^{Tp_x}) \\ \dots \\ \psi'_t((\mathbf{x}_t)_{Tp_x+1}^{(T+1)p_x}) \end{pmatrix} = \begin{pmatrix} \psi'_t(x_{t-T}) \\ \psi'_t(x_{t-T+1}) \\ \dots \\ \psi'_t(x_t) \end{pmatrix}.$$

Таким образом, в строках этой матрицы собраны все возможные в момент  $t$  из-за задержек  $\tau_t$  наблюдения без шума. Для описания самих задержек также введем матричную функцию  $\Theta_t(\mathbf{x}_t) \in \mathbb{R}^{(T+1)q_y}$ , принимающую значения в соответствии со следующим правилом:

$$\left( \Theta_t(\mathbf{x}_t) \right)_{i,j} = \begin{cases} 1, & \text{если } (\theta_t(x_t))_i = T - j + 1; \\ 0, & \text{в противном случае,} \end{cases} \\ i = 1, \dots, q_y, j = 1, \dots, T + 1.$$

Таким образом, элемент  $(i, j)$  матрицы  $\Theta_t(\mathbf{x}_t)$  принимает значение 1, если наблюдению  $(y_t)_i$  соответствует задержка  $(\tau_t)_i = T - j + 1$ .

Эти два обозначения позволяют записать уравнение наблюдений в виде:

$$y_t = \Theta_t(\mathbf{x}_t)\Psi_t(\mathbf{x}_t) + v_t. \quad (3)$$

Таким образом, (2), (3) — это каноническая форма марковской системы наблюдения с дискретным временем и аддитивными независимыми шумами.

Значения, которые может принимать функция  $\Theta_t(\mathbf{x}_t)$ , обозначим  $[\Theta_t]_k$ ,  $k = 1, \dots, (T + 1)q_y$ . Порядковый номер  $k$  можно задать, положив

$$k = j_{k,1} + (T + 1)j_{k,2} + \dots + (T + 1)^{q_y-1}j_{k,q_y}, \quad (4)$$

где  $j_{k,i}$  — номер строки  $i$ -го столбца матрицы  $[\Theta_t]_k$ , в котором стоит 1.

Эти же обозначения дают возможность пересчитать возможные значения функции наблюдателя  $\Theta_t(\mathbf{x}_t)\Psi_t(\mathbf{x}_t)$ . Обозначим значение этой функции с тем же номером  $k$  из (4) через  $[\Theta_t\Psi_t]_k$ . Тогда

$$[\Theta_t\Psi_t]_k = [\Theta_t\Psi_t(\mathbf{x}_t)]_k = \begin{pmatrix} (\psi_t(x_{t-j_{k,1}}))_1 \\ (\psi_t(x_{t-j_{k,2}}))_2 \\ \dots \\ (\psi_t(x_{t-j_{k,1}}))_{q_y} \end{pmatrix}.$$

Здесь  $(\psi_t)_i$  —  $i$ -я координата вектора  $\psi_t$ . Таким образом,  $[\Theta_t\Psi_t]_k$  состоит из наблюдений без шума, смещенных на величины задержек, и соответствует порядковому номеру  $k$ , заданному для матрицы  $[\Theta_t]_k$  соотношением (4). Кроме того, пронумерованы возможные значения вектора запаздываний  $\tau_t$ :

$$[\tau_t]_k = [\theta_t(x_t)]_k, \quad ([\tau_t]_k)_i = T - j_{k,i} + 1, \\ i = 1, \dots, q_y.$$

Сделанные обозначения несколько громоздки, но они нужны, чтобы представить оптимальный байесовский фильтр. С этой целью введем следующие правила именования вероятностных характеристик. Для случайных векторов  $x \in \mathbb{R}^p$ ,  $y \in \mathbb{R}^q$  через  $(x', y')' \in \mathbb{R}^{p+q}$  обозначим составной вектор. Используемые в соотношениях ниже плотности вероятности будем обозначать так:  $f_x(X)$  — маргинальная плотность  $x$ ,  $f_{x,y}(X, Y)$  — плотность составного вектора  $(x', y')'$ ,  $f_{x|y}(X|Y)$  — условная плотность  $x$  относительно  $y$  при дополнительном предположении  $f_y(Y) > 0$ . Дополнительно отметим, что

- символы  $x$  используются для обозначения случайных векторов в нотациях исходной системы наблюдения (1);
- символы  $X$  используются для обозначения аргументов плотностей вероятности, соответствующих  $x$ ;
- символы  $\mathbf{x}$  используются для обозначения случайных векторов в нотации расширенного состояния в системе наблюдения (2), (3);
- символы  $\mathbf{X}$  используются для обозначения аргументов плотностей вероятности, соответствующих  $\mathbf{x}$ ;
- обозначения  $y^t$  используются для вектора всех наблюдений до момента  $t$  включительно, т. е.  $y^t = (y'_0, \dots, y'_t)'$ , и  $Y^t$  — для соответствующего аргумента плотности вероятности.

Для условного распределения состояния  $\mathbf{x}_t$  системы (2) относительно наблюдений  $y_s$ ,  $s = 0, \dots, t$ , (3) можно формально записать рекуррентные байесовские соотношения для апостериорной плотности вероятности [21]:

$$f_{\mathbf{x}_t|y^t}(\mathbf{X}_t|Y^t) = \\ = \left( \int f_{\mathbf{x}_t|\mathbf{x}_{t-1}}(\mathbf{X}_t|\mathbf{X}_{t-1}) f_{\mathbf{x}_{t-1}|y^{t-1}}(\mathbf{X}_{t-1}|Y^{t-1}) d\mathbf{X}_{t-1} \times \right. \\ \left. \times f_{y_t|\mathbf{x}_t}(Y_t|\mathbf{X}_t) \right) / \left( \iint f_{\mathbf{x}_t|\mathbf{x}_{t-1}}(\mathbf{X}_t|\mathbf{X}_{t-1}) \times \right. \\ \left. \times f_{\mathbf{x}_{t-1}|y^{t-1}}(\mathbf{X}_{t-1}|Y^{t-1}) d\mathbf{X}_{t-1} \cdot f_{y_t|\mathbf{x}_t}(Y_t|\mathbf{X}_t) d\mathbf{X}_t \right).$$

Это выражение можно использовать при условии, что все плотности вероятности в нем существуют. Для системы (2), (3) это не так для переходной плотности  $f_{\mathbf{x}_t|\mathbf{x}_{t-1}}(\mathbf{X}_t|\mathbf{X}_{t-1})$ . Кроме того, нужно получить явный вид плотности  $f_{y_t|\mathbf{x}_t}(Y_t|\mathbf{X}_t)$ .

Выполненные в следующей теореме преобразования позволяют вместо переходной плотности  $f_{\mathbf{x}_t|\mathbf{x}_{t-1}}(\mathbf{X}_t|\mathbf{X}_{t-1})$  расширенного состояния использовать переходную плотность  $f_{x_t|x_{t-1}}(X_t|X_{t-1})$  исходного состояния и, соответственно, вместо  $\int \cdot d\mathbf{X}_{t-1}$ , который вычисляется по пространству  $\mathbb{R}^{(T+1)p_x}$ , считать интеграл  $\int \cdot dX_{t-T-1}$  по пространству  $\mathbb{R}^{p_x}$ . Кроме того, записывается явный вид плотности  $f_{y_t|\mathbf{x}_t}(Y_t|\mathbf{X}_t)$ .

**Теорема.** Пусть для системы (1) задано  $T > 1$  и выполнено:

- плотность вероятности возмущений  $f_{w_t}(W_t)$ ,  $t = -T, -T + 1, \dots$ , непрерывна и векторы  $w_t$  имеют конечные вторые моменты;
- плотность вероятности ошибок наблюдений  $f_{v_t}(V_t)$ ,  $t = 0, 1, \dots$ , непрерывна, неотрицательна и векторы  $v_t$  имеют конечные вторые моменты;
- плотность вероятности начального условия  $f_\eta(X_{-T-1})$  непрерывна и вектор  $\eta$  имеет конечный второй момент;
- функции  $\varphi_t(x)$  и  $\psi_t(x)$  непрерывны и удовлетворяют условию линейного роста, т. е. существует константа  $C$ :

$$\|\varphi_t(x)\|^2 + \|\psi_t(x)\|^2 < C(1 + \|x\|^2), \quad x \in \mathbb{R}^{p_x}.$$

Тогда для апостериорной плотности вероятности  $\rho_t = \rho_t(\mathbf{X}_t|Y^t)$  расширенного состояния  $\mathbf{x}_t = (x'_{t-T}, \dots, x'_t)'$ ,  $t = 0, 1, \dots$ , системы, заданного соотношением (2), относительно наблюдений  $y^t = (y'_0, \dots, y'_t)'$ ,  $t = 0, 1, \dots$ , (3) и оценки оптимального байесовского фильтра  $\hat{x}_t = E\{x_t|Y^t\}$  состояния  $x_t$  выполнены рекуррентные равенства:

$$\rho_t = \left( \begin{aligned} & f_{w_t}(X_t - \varphi_t(X_{t-1})) \sum_k I(\Theta_t(X_t) = [\tau_t]_k) f_{v_t}(Y_t - [\Theta_t \Psi_t(\mathbf{X}_t)]_k) \times \\ & \times \int \rho_{t-1} dX_{t-T-1} \Big/ \left( \int \left( f_{w_t}(X_t - \varphi_t(X_{t-1})) \sum_k I(\theta_t(X_t) = [\tau_t]_k) f_{v_t}(Y_t - [\Theta_t \Psi_t(\mathbf{X}_t)]_k) \int \rho_{t-1} dX_{t-T-1} \right) d\mathbf{X}_t \right); \\ & \hat{\mathbf{x}}_t = \int \mathbf{X}_t \rho_t(\mathbf{X}_t | Y^t) d\mathbf{X}_t, \hat{x}_t = (\hat{\mathbf{x}}_t)_{T p_x + 1}^{(T+1)p_x} \end{aligned} \right) \quad (5)$$

с начальным условием

$$\begin{aligned} \rho_{-1}(\mathbf{X}_{-1} | Y^{-1}) &= \\ &= \rho_{-1}(\mathbf{X}_{-1}) = \rho_{-1}(X_{-T-1}, \dots, X_{-1}) = \\ &= f_\eta(X_{-T-1}) f_{w_{-T}}(X_{-T} - \varphi_{-T}(X_{-T-1})) \dots \\ &\quad \dots f_{w_{-1}}(X_{-1} - \varphi_{-1}(X_{-2})) \end{aligned}$$

и правилом нумерации  $k = 1, \dots, q_y$ , определенным в (4).

В рекуррентных соотношениях (5) опущены аргументы и упрощена запись суммы. Отметим, что не относящиеся к необходимым условия ограниченности вторых моментов наблюдений  $y_t$  сформулированы для того, чтобы в модели можно было использовать помимо байесовского фильтра другие, неоптимальные оценки. Также заметим, что нет формальных причин, препятствующих компьютерной реализации этих формул. Однако отсутствует реалистичность выполнения такого компьютерного расчета из-за роста размерности, масштаб которого определяется значением  $T$ . В численном эксперименте, обсуждаемом далее в статье,  $T = 50$ . Значит, даже при небольших  $p_x = 2$  и  $q_y = 4$  интегралы в (5) будут по пространству  $\mathbb{R}^{100}$ , а слагаемых будет 404. При таких параметрах расчеты согласно (5) невозможно выполнить на практике. Именно этот тезис, т.е. отсутствие практических перспектив использования в рассматриваемой задаче классических алгоритмов фильтрации, хотя они и могут быть записаны в окончательном замкнутом виде, составляет формальный итог исследования. Доказательство. Представим искомую плотность вероятности в виде:

$$\begin{aligned} f_{\mathbf{x}_t | y^t}(\mathbf{X}_t | Y^t) &= \frac{f_{\mathbf{x}_t, y^t}(\mathbf{X}_t, Y^t)}{f_{y^t}(Y^t)} = \\ &= \frac{f_{\mathbf{x}_t, y^{t-1}, y_t}(\mathbf{X}_t, Y^{t-1}, Y^t)}{f_{y^t}(Y^t)} = \\ &= \frac{f_{\mathbf{x}_t, y^{t-1}}(\mathbf{X}_t, Y^{t-1}) f_{y_t | \mathbf{x}_t}(Y_t | \mathbf{X}_t)}{f_{y^t}(Y^t)}. \end{aligned}$$

Далее для первого множителя  $f_{\mathbf{x}_t, y^{t-1}}(\mathbf{X}_t, Y^{t-1})$  имеем:

$$\begin{aligned} f_{\mathbf{x}_t, y^{t-1}}(\mathbf{X}_t, Y^{t-1}) &= \\ &= f_{x_{t-T}, \dots, x_t, y^{t-1}}(X_{t-T}, \dots, X_t, Y^{t-1}) = \\ &= \int f_{x_{t-T-1}, x_{t-T}, \dots, x_t, y^{t-1}}(X_{t-T-1}, X_{t-T}, \dots, \\ &\quad \dots, X_t, Y^{t-1}) dX_{t-T-1} = \\ &= \int f_{x_t | \mathbf{x}_{t-1}, y^{t-1}}(X_t | \mathbf{X}_{t-1}, Y^{t-1}) \times \\ &\quad \times f_{\mathbf{x}_{t-1}, y^{t-1}}(\mathbf{X}_{t-1}, Y^{t-1}) dX_{t-T-1} = \\ &= \int f_{x_t | x_{t-1}}(X_t | X_{t-1}) \times \\ &\quad \times f_{\mathbf{x}_{t-1} | y^{t-1}}(\mathbf{X}_{t-1} | Y^{t-1}) dX_{t-T-1} \cdot f_{y^{t-1}}(Y^{t-1}) = \\ &= \int f_{x_{t-1} | y^{t-1}}(\mathbf{X}_{t-1} | Y^{t-1}) dX_{t-T-1} \times \\ &\quad \times f_{x_t | x_{t-1}}(X_t | X_{t-1}) f_{y^{t-1}}(Y^{t-1}) = \\ &= \int f_{\mathbf{x}_{t-1} | y^{t-1}}(\mathbf{X}_{t-1} | Y^{t-1}) dX_{t-T-1} \times \\ &\quad \times f_{w_t}(X_t - \varphi_t(X_{t-1})) f_{y^{t-1}}(Y^{t-1}). \end{aligned}$$

Для второго множителя  $f_{y_t | \mathbf{x}_t}(Y_t | \mathbf{X}_t)$ :

$$\begin{aligned} f_{y_t | \mathbf{x}_t}(Y_t | \mathbf{X}_t) &= \sum_{k=0}^{(T+1)q_y} I(\Theta_t(\mathbf{X}_t) = [\Theta_t]_k) \times \\ &\quad \times f_{y_t | \mathbf{x}_t}(Y_t | \mathbf{X}_t, \Theta_t(\mathbf{X}_t) = [\Theta_t]_k) = \\ &= \sum_{k=0}^{(T+1)q_y} I(\theta_t(X_t) = [\tau_t]_k) \times \\ &\quad \times f_{y_t | \mathbf{x}_t, \tau_t = [\tau_t]_k}(Y_t | \mathbf{X}_t, \tau_t = [\tau_t]_k) = \\ &= \sum_{k=0}^{(T+1)q_y} I(\theta_t(X_t) = [\tau_t]_k) f_{v_t}(Y_t - [\Theta_t \Psi_t(\mathbf{X}_t)]_k). \end{aligned}$$

И окончательно получаем

$$\begin{aligned} f_{\mathbf{x}_t | y^t} &= \frac{f_{y^{t-1}}(Y^{t-1})}{f_{y^t}(Y^t)} \times \\ &\quad \times \left( \int f_{\mathbf{x}_{t-1} | y^{t-1}} dX_{t-T-1} \cdot f_{w_t}(X_t - \varphi_t(X_{t-1})) \times \right. \\ &\quad \times \left. \sum_k I(\theta_t(X_t) = [\tau_t]_k) f_{v_t}(Y_t - [\Theta_t \Psi_t(\mathbf{X}_t)]_k) \right) = \\ &= \left( f_{w_t}(X_t - \varphi_t(X_{t-1})) \sum_k I(\theta_t(X_t) = [\tau_t]_k) \times \right. \\ &\quad \times \left. f_{v_t}(Y_t - [\Theta_t \Psi_t(\mathbf{X}_t)]_k) \int f_{\mathbf{x}_{t-1} | y^{t-1}} dX_{t-T-1} \right) \times \end{aligned}$$

$$\times \left( \int f_{w_t}(X_t - \varphi_t(X_{t-1})) \sum_k I(\theta_t(X_t) = [\tau_t]_k) \times \right. \\ \left. \times f_{v_t}(Y_t - [\Theta_t \Psi_t(\mathbf{X}_t)]_k) \int f_{\mathbf{x}_{t-1}|y^{t-1}} d\mathbf{X}_{t-T-1} d\mathbf{X}_t \right)^{-1}.$$

В последнем равенстве опущены аргументы у плотностей  $f_{\mathbf{x}_t|y^t}(\mathbf{X}_t|Y^t)$  и  $f_{\mathbf{x}_{t-1}|y^{t-1}}(\mathbf{X}_{t-1}|Y^{t-1})$ , а также суммирование  $\sum_{k=0}^{(T+1)q_y}$  обозначено как  $\sum_k$  и учтено, что коэффициент  $f_{y^t}(Y^t)/f_{y^{t-1}}(Y^{t-1})$  — нормирующий множитель.

Условия ограниченности вторых моментов и линейного роста функций системы очевидным образом достаточны для существования вторых моментов  $x_t$  и  $y_t$ :  $E\{\|x_t\|^2\} + E\{\|y_t\|^2\} < \infty$  и оптимальности оценки  $E\{x_t|Y^t\}$  в среднеквадратичном, что завершает доказательство.

## 4 Компьютерное моделирование

Отсутствие возможности даже в упрощенном модельном эксперименте проводить сложные вычисления (5) не означает, что для рассматриваемой модели требуются специальные алгоритмы оценивания. Вполне возможно, что существенного влияния временная задержка на качество оценивания не оказывает, так что в практических задачах ее можно просто проигнорировать. В данном разделе статьи приводится простой, но практически состоятельный пример движения объекта в водной среде и наблюдения за ним и иллюстрируется низкое качество простого алгоритма фильтрации, не учитывающего задержки. Предполагается, что AUV движется на глубине в горизонтальной плоскости  $Oxy$ . В каждый момент времени к текущей скорости добавляется случайный шум, имитирующий хаотичное маневрирование. Координаты траектории AUV обозначаются, как принято,  $x(t)$  и  $y(t)$ . Обратим внимание на то, что эти обозначения отличны от обозначений вектора состояния  $x_t$  и наблюдений  $y_t$  в исходной модели (1). Единицей измерений положения остаются километры, скорости измеряются в километрах в час, время — в часах. Модель непрерывного движения дискретизована с шагом  $h = 0,0001$  ч, что соответствует частоте около трех измерений в секунду. Для удобства графического представления результатов считается, что движение начинается в момент времени  $t = 0$  и завершается в момент  $t = 1000$ , т.е. движение продолжается 6 мин. Систематические постоянные значения скорости составляют  $v_x = 25$  км/ч и  $v_y = 12,5$  км/ч. За время движения AUV в среднем перемещается на 3 км.

Начальное положение AUV задано гауссовским вектором  $(x(0), y(0))'$ , который имеет среднее  $(6,25; 12,5)'$  и ковариацию  $\text{diag}\{2,5^2; 5^2\}$ . Однотипные наблюдатели расположены в двух точках на той же плоскости  $Oxy$ : первый имеет координаты  $(0, l_y)$ ,  $l_y = 6,25$ , т.е. расположен в 6,25 км от начала координат по оси  $Oy$ . Координаты второго —  $(l_x, 0)$ ,  $l_x = 12,5$ , т.е. он расположен в 12,5 км от начала координат по оси  $Ox$ . В предположении, что для наблюдения используются акустические сонары и скорость звука в воде равна  $v_s = 5400$  км/ч (1500 м/с), можно определить максимально возможную величину задержки наблюдений  $T = 50$ , т.е. максимальная величина, на которую могут запаздывать наблюдения, составляет 18 с. Соответственно, процесс фильтрации и первая оценка появляются в момент  $T$ , через 18 с после начала движения (можно считать, после обнаружения цели). Каждый наблюдатель измеряет дальность и направляющий косинус до цели с аддитивной ошибкой. Вектор ошибок предполагается гауссовским со средним  $(0, 0, 0, 0)'$  и ковариацией  $\text{diag}\{0,001^2, 0,005^2, 0,001^2, 0,005^2\}$ .

Таким образом, имеется следующая система наблюдения. Вектор состояния имеет вид:

$$\left. \begin{aligned} x(t) &= x(t-1) + hv_x + w_x(t); \\ y(t) &= y(t-1) + hv_y + w_y(t). \end{aligned} \right\} \quad (6)$$

Описывающий хаотичные изменения скорости вектор  $(w_x(t), w_y(t))'$  предполагается гауссовским со средним  $(0, 0)'$  и ковариацией  $\text{diag}\{0,01^2; 0,01^2\}$ . Таким образом, AUV может менять скорость на величину порядка 100 км/ч, т.е. это очень быстро маневрирующая цель. Но она интересна для расчета тем, что дает очень разнообразные траектории.

Получаемые измерения обозначаются  $d_1(t)$  и  $d_2(t)$  (дальности, измеренные первым и вторым наблюдателем) и  $c_1(t)$  и  $c_2(t)$  (направляющие косинусы). Вектор ошибок измерений обозначен  $(v_{d_1}(t), v_{d_2}(t), v_{c_1}(t), v_{c_2}(t))'$ . Таким образом, уравнения наблюдений принимают вид:

$$\left. \begin{aligned} d_1(t) &= \\ &= \sqrt{(x(t - (\tau_t)_1))^2 + (y(t - (\tau_t)_1) - l_y)^2} + \\ &\quad + v_{d_1}(t); \\ c_1(t) &= \\ &= \frac{y(t - (\tau_t)_2) - l_y}{\sqrt{(x(t - (\tau_t)_2))^2 + (y(t - (\tau_t)_2) - l_y)^2}} + \\ &\quad + v_{c_1}(t); \end{aligned} \right\} \quad (7)$$

$$\left. \begin{aligned} d_2(t) &= \\ &= \sqrt{(x(t - (\tau_t)_3) - l_x)^2 + (y(t - (\tau_t)_3))^2} + \\ &\quad + v_{d_2}(t); \\ c_2(t) &= \\ &= \frac{x(t - (\tau_t)_4) - l_x}{\sqrt{(x(t - (\tau_t)_4) - l_x)^2 + (y(t - (\tau_t)_4))^2}} + \\ &\quad + v_{c_2}(t). \end{aligned} \right\} (8)$$

С учетом выбранного шага дискретизации  $h$  и скорости звука в воде  $v_s$  элементы  $\tau_t$  выражаются через дальности наблюдателей до AUV, т. е.

$$\left. \begin{aligned} (\tau_t)_1 = (\tau_t)_2 &= \\ &= \min \left\{ T, \left[ \frac{\sqrt{(x(t))^2 + (y(t) - l_y)^2}}{hv_s} \right] \right\}; \\ (\tau_t)_3 = (\tau_t)_4 &= \\ &= \min \left\{ T, \left[ \frac{\sqrt{(x(t) - l_x)^2 + (y(t))^2}}{hv_s} \right] \right\}. \end{aligned} \right\} (9)$$

В (9) использовано обозначение  $[x]$  для целой части  $x$  и учтена потенциальная возможность удаления объекта на дистанцию, для которой величина задержки становится больше заданного максимума  $T$ .

Для реализации «естественного» фильтра учтен геометрический смысл наблюдений, а именно: если предположить, что в (7) и (8) отсутствует шум, то можно очевидным образом определить положение AUV, пересчитав дальности и косинусы в декартовы координаты. Такое преобразование дает две независимые оценки положения AUV  $(\tilde{x}_1(t), \tilde{y}_1(t))$  и  $(\tilde{x}_2(t), \tilde{y}_2(t))$ , вычисленные на основании измерений, выполненных первым и вторым наблюдателем, без учета шумов и временной задержки:

$$\begin{pmatrix} \tilde{x}_1(t) \\ \tilde{y}_1(t) \\ \tilde{x}_2(t) \\ \tilde{y}_2(t) \end{pmatrix} = \begin{pmatrix} \sqrt{(d_1(t))^2 - (c_1(t)d_1(t))^2} \\ c_2(t)d_1(t) + l_y \\ c_2(t)d_2(t) + l_x \\ \sqrt{(d_2(t))^2 - (c_2(t)d_2(t))^2} \end{pmatrix}.$$

С этими выражениями очевидным образом могут быть использованы равенства (9) для получения оценки  $\tilde{\tau}_t$ :

$$\begin{aligned} (\tilde{\tau}_t)_1 = (\tilde{\tau}_t)_2 &= \frac{\sqrt{(\tilde{x}(t))^2 + (\tilde{y}(t) - l_y)^2}}{hv_s}; \\ (\tilde{\tau}_t)_3 = (\tilde{\tau}_t)_4 &= \frac{\sqrt{(\tilde{x}(t) - l_x)^2 + (\tilde{y}(t))^2}}{hv_s}. \end{aligned}$$

Далее в оценках надо учесть изменение положения в силу динамики движения (6) за время задержки

$$\begin{pmatrix} \hat{x}_1(t) \\ \hat{y}_1(t) \\ \hat{x}_2(t) \\ \hat{y}_2(t) \end{pmatrix} = \begin{pmatrix} \tilde{x}_1(t) + (\tilde{\tau}_t)_1 v_x \\ \tilde{y}_1(t) + (\tilde{\tau}_t)_2 v_y \\ \tilde{x}_2(t) + (\tilde{\tau}_t)_3 v_x \\ \tilde{y}_2(t) + (\tilde{\tau}_t)_4 v_y \end{pmatrix}$$

и окончательно скомбинировать оценки двух наблюдателей, учитывая их одинаковую точность:

$$\hat{x}(t) = \frac{\hat{x}_1(t) + \hat{x}_2(t)}{2}; \quad \hat{y}(t) = \frac{\hat{y}_1(t) + \hat{y}_2(t)}{2}.$$

В выполненном компьютерном расчете моделировался пучок из 100 000 траекторий системы наблюдения (6)–(9) и оценки  $(\hat{x}(t), \hat{y}(t))$ . Анализировались точности оценивания координат положения AUV, определяемые среднеквадратичными отклонениями ошибок оценок  $\sigma_{\hat{x}}(t)$  и  $\sigma_{\hat{y}}(t)$ .

На рис. 1 представлены примеры траектории AUV  $(x(t), y(t))'$  и соответствующих оценок  $(\hat{x}(t), \hat{y}(t))'$ . Этот рисунок дает качественное представление о результатах оценивания: фильтр работает, отслеживая цель, но отклонения от истинной траектории кажутся значительными. Подтверждает это рис. 2 с формальными оценками каче-

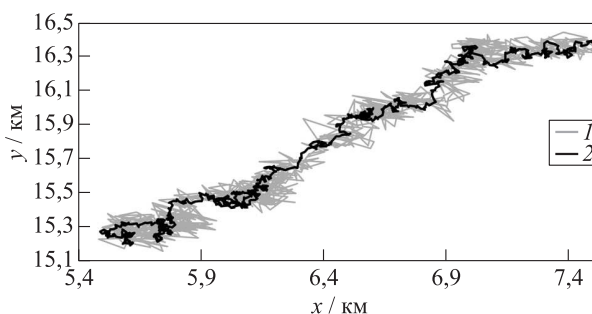


Рис. 1 Пример траекторий: 1 — оценка фильтрации  $(\hat{x}(t), \hat{y}(t))'$ ; 2 — истинная траектория AUV  $(x(t), y(t))'$

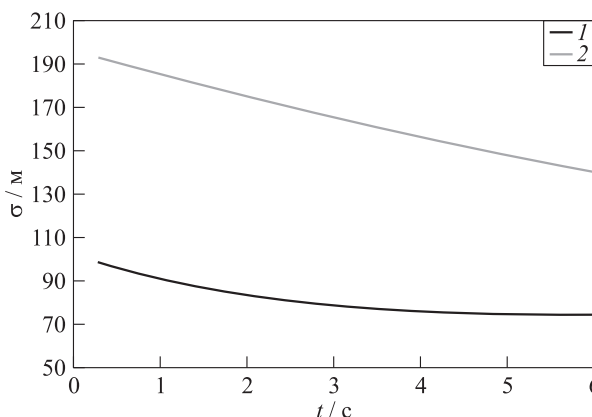


Рис. 2 Статистические оценки качества фильтрации: 1 —  $\sigma_{\hat{x}}(t)$ ; 2 —  $\sigma_{\hat{y}}(t)$

ства фильтрации  $\sigma_{\hat{x}}(t)$  и  $\sigma_{\hat{y}}(t)$ , вычисленными путем осреднения по смоделированному пучку.

## 5 Заключение

Рассмотренная модель динамической системы со случайными задержками наблюдений представляет вызов с точки зрения традиционной задачи фильтрации. Обосновывают это выполненные формальные преобразования модели к традиционной марковской системе и полученные уравнения оптимальной байесовской фильтрации. Также подтверждает этот тезис представленный эксперимент с моделью, близкой к реальности. К качеству «естественного» фильтра, использующего геометрические свойства измерений, есть вопросы. Порядок (среднеквадратичное отклонение) ошибки оценки координат AUV составляет 70 м для  $x(t)$  и 150 м для  $y(t)$ . Эксперименты с условно-оптимальным фильтром Пугачева [22, 23] показывают, что в данном примере при  $T = 0$  можно добиться точности оценок каждой координаты порядка 20 м. Если учесть, что за максимальное время задержки  $T = 50$  AUV в среднем изменяет положение примерно на 15 м в известном направлении, то придется признать, что оценка «естественного» фильтра имеет неприемлемо низкую точность. При этом исследованная оценка  $(\hat{x}(t), \hat{y}(t))'$  не так плоха. Она гораздо лучше тривиальной оценки (безусловного математического ожидания) и обладает устойчивостью, которой нет у большинства известных субоптимальных фильтров. Прямое применение условно-оптимального фильтра также вызывает трудности, но перспективы успеха использования данной концепции к предложенной модели представляются весьма хорошими. В этом направлении предполагаются будущие исследования.

## Литература

1. Bar-Shalom Y., Li X. R., Kirubarajan T. Estimation with applications to tracking and navigation: Theory, algorithms, and software. — New York, NY, USA: J. Wiley & Sons, 2001. 584 p.
2. Autonomous underwater vehicles: Design and practice (radar, sonar & navigation) / Ed. F. Ehlers. — London, U.K.: SciTech Publishing, 2020. 592 p.
3. Advances in marine vehicles, automation and robotics // J. Marine Science Engineering. Special Issue. www.mdpi.com/journal/jmse/special\_issues/advances\_in\_marine\_vehicles\_automation\_and\_robotics.
4. Groen J., Beerens S. P., Been R., Doisy Y., Noutary E. Adaptive port-starboard beamforming of triplet sonar arrays // IEEE J. Oceanic Eng., 2005. Vol. 30. P. 348–359. doi: 10.1109/JOE.2005.850880.
5. Luo J., Han Y., Fan L. Underwater acoustic target tracking: A review // Sensors, 2018. Vol. 18. No. 1. Art. 112. doi: 10.3390/s18010112.
6. Ghaffoor H., Noh Y. An overview of next-generation underwater target detection and tracking: An integrated underwater architecture // IEEE Access, 2019. Vol. 7. P. 98841–98853. doi: 10.1109/ACCESS.2019.2929932.
7. Wolek A., Dzikowicz B. R., McMahon J., Houston B. H. At-Sea evaluation of an underwater vehicle behavior for passive target tracking // IEEE J. Oceanic Eng., 2019. Vol. 44. P. 514–523. doi: 10.1109/JOE.2018.2817268.
8. Su X., Ullah I., Liu X., Choi D. A review of underwater localization techniques, algorithms, and challenges // J. Sensors, 2020. Vol. 2020. Art. 6403161. doi: 10.1155/2020/6403161.
9. Borisov A., Bosov A., Miller B., Miller G. Passive underwater target tracking: Conditionally minimax nonlinear filtering with bearing-Doppler observations // Sensors, 2020. Vol. 20. No. 8. Art. 2257. doi: 10.3390/s20082257.
10. Kumar M., Mondal S. Recent developments on target tracking problems: A review // Ocean Eng., 2021. No. 236. Art. 109558. 20 p. doi: 10.1016/j.oceaneng. 2021.109558.
11. Miller A., Miller B., Miller G. Navigation of underwater drones and integration of acoustic sensing with onboard inertial navigation system // Drones, 2021. Vol. 5. No. 3. Art. 83. doi: 10.3390/drones5030083.
12. Kalman R. E. A new approach to linear filtering and prediction problems // J. Basic Eng. — T. ASME, 1960. Vol. 82. No. 1. P. 35–45. doi: 10.1115/1.3662552.
13. Bernstein I., Friedland B. Estimation of the state of a nonlinear process in the presence of nongaussian noise and disturbances // J. Frankl. Inst., 1966. Vol. 281. No. 6. P. 455–480. doi: 10.1016/0016-0032(66)90434-0.
14. Julier S. J., Uhlmann J. K. New extension of the Kalman filter to nonlinear systems // Proc. SPIE, 1997. Vol. 3068. P. 182–193. doi: 10.1117/12.280797.
15. Julier S. J., Uhlmann J. K. Unscented filtering and nonlinear estimation // P. IEEE, 2004. Vol. 92. No. 3. P. 401–422. doi: 10.1109/JPROC.2003.823141.
16. Arasaratnam I., Haykin S. Cubature Kalman filters // IEEE T. Automat. Contr., 2009. Vol. 54. No. 6. P. 1254–1269. doi: 10.1109/TAC.2009.2019800.
17. Wang T., Zhang L., Liu S. Improved robust high-degree cubature Kalman filter based on novel cubature formula and maximum correntropy criterion with application to surface target tracking // J. Marine Science Engineering, 2022. Vol. 10. No. 8. Art. 1070. doi: 10.3390/jmse10081070.
18. Christ R. D., Wernli R. L. The ROV manual: A user guide for remotely operated vehicles. — 2nd ed. — Oxford, U.K.: Butterworth-Heinemann, 2013. 712 p.
19. Li L., Li Y., Zhang Y., Xu G., Zeng J., Feng X. Formation control of multiple autonomous underwater vehicles under communication delay, packet discreteness and dropout // J. Marine Science Engineering, 2022. Vol. 10. No. 7. Art. 920. doi: 10.3390/jmse10070920.

20. Zhao L., Wang J., Yu T., Chen K., Su A. Incorporating delayed measurements in an improved high-degree cubature Kalman filter for the nonlinear state estimation of chemical processes // *ISA T.*, 2019. Vol. 86. P. 122–133. doi: 10.1016/j.isatra.2018.11.004.
21. Bertsekas D. P., Shreve S. E. Stochastic optimal control: The discrete-time case. — New York, NY, USA: Academic Press, 1978. 330 p.
22. Пугачев В. С., Сеницын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с.
23. Сеницын И. Н., Корепанов Э. Р. Нормальные условно-оптимальные фильтры Пугачёва для дифференциальных стохастических систем, линейных относительно состояния // *Информатика и её применения*, 2015. Т. 9. Вып. 2. С. 30–38. doi: 10.14357/19922264150204.

Поступила в редакцию 26.05.23

## NONLINEAR DYNAMIC SYSTEM STATE OPTIMAL FILTERING BY OBSERVATIONS WITH RANDOM DELAYS

A. V. Bosov

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** A mathematical model of a nonlinear dynamic observation system with a discrete time which allows taking into account the dependence of the time of receiving observations on the state of the observed object is proposed. The model implements the assumption that the time between the moment when the measurement of the state is formed and the moment when the measured state is received by the observer depends randomly on the position of the moving object. Such an assumption source is the process of observation by stationary means of an autonomous underwater apparatus in which the time of obtaining up-to-date data depends on the unknown distance between the object and the observer. Unlike deterministic delays formed by the known state of the observation environment, to account for the dependence of time delays on the unknown state of the object of observation, it is required to use random functions to describe them. The main result of the study of the proposed model is the solution of the optimal filtering problem. For this purpose, recurrent Bayesian relations describing the evolution of the a posteriori probability density are obtained. The difficulties of using a semifinished filter for practical purposes are discussed. The proposed model is illustrated by a practical example of the task of tracking a moving underwater object based on the results of measurements performed by typical acoustic sensors. It is assumed that the object moves under the water in a plane with a known average speed, constantly performs chaotic maneuvers, and is observed by two independent complexes of acoustic sensors measuring the distances to the object and the guiding cosines. The complexity of determining the position of such an object is illustrated by a simple filter using the geometric properties of the measured quantities and the least squares method.

**Keywords:** stochastic dynamic observation system; state filtering; optimal Bayesian filter; mean square evaluation criterion; autonomous underwater vehicle; acoustic sensor; target tracking

DOI: 10.14357/19922264230302

EDN: CFVYJM

### Acknowledgments

The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (СКР “Informatics”) of FRC CSC RAS (Moscow).

### References

1. Bar-Shalom, Y., X. R. Li, and T. Kirubarajan. 2004. *Estimation with applications to tracking and navigation: Theory, algorithms and software*. New York, NY: John Wiley & Sons. 548 p.
2. Ehlers, F., ed. 2020. *Autonomous underwater vehicles: Design and practice (radar, sonar & navigation)*. London, U.K.: SciTech Publishing. 592 p.
3. Advances in marine vehicles, automation and robotics. *J. Marine Science Engineering*. Special Issue. Available at: [www.mdpi.com/journal/jmse/special\\_issues/advances\\_in\\_marine\\_vehicles\\_automation\\_and\\_robotics](http://www.mdpi.com/journal/jmse/special_issues/advances_in_marine_vehicles_automation_and_robotics) (accessed June 27, 2023).
4. Groen, J., S. P. Beerens, R. Been, Y. Doisy, and E. Noutary. 2005. Adaptive port-starboard beamforming of triplet sonar arrays. *IEEE J. Oceanic Eng.* 30:348–359. doi: 10.1109/JOE.2005.850880.



5. Luo, J., Y. Han, and L. Fan. 2018. Underwater acoustic target tracking: A review. *Sensors* 18(1):112. doi: 10.3390/s18010112.
6. Ghafoor, H., and Y. Noh. 2019. An overview of next-generation underwater target detection and tracking: An integrated underwater architecture. *IEEE Access* 7:98841–98853. doi: 10.1109/ACCESS.2019.2929932.
7. Wolek, A., B. R. Dzikowicz, J. McMahon, and B. H. Houston. 2019. At-Sea evaluation of an under-water vehicle behavior for passive target tracking. *IEEE J. Oceanic Eng.* 44:514–523. doi: 10.1109/JOE.2018.2817268.
8. Su, X., I. Ullah, X. Liu, and D. Choi. 2020. A review of underwater localization techniques, algorithms, and challenges. *J. Sensors* 2020:6403161. doi: 10.1155/2020/6403161.
9. Borisov, A., A. Bosov, B. Miller, and G. Miller. 2020. Passive underwater target tracking: Conditionally minimax nonlinear filtering with bearing-Doppler observations. *Sensors* 20(8):2257. doi: 10.3390/s20082257.
10. Kumar, M., and S. Mondal. 2021. Recent developments on target tracking problems: A review. *Ocean Eng.* 236:109558. 20 p. doi: 10.1016/j.oceaneng.2021.109558.
11. Miller, A., B. Miller, and G. Miller. 2021. Navigation of underwater drones and integration of acoustic sensing with onboard inertial navigation system. *Drones* 5(3):83. doi: 10.3390/drones5030083.
12. Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng. — T. ASME* 82(1):35–45. doi: 10.1115/1.3662552.
13. Bernstein, I., and B. Friedland. 1966. Estimation of the state of a nonlinear process in the presence of nongaussian noise and disturbances. *J. Frankl. Inst.* 281(6):455–480. doi: 10.1016/0016-0032(66)90434-0.
14. Julier, S. J., and J. K. Uhlmann. 1997. New extension of the Kalman filter to nonlinear systems. *Proc. SPIE* 3068:182–193. doi: 10.1117/12.280797.
15. Julier, S. J., and J. K. Uhlmann. 2004. Unscented filtering and nonlinear estimation. *P. IEEE* 92(3):401–422. doi: 10.1109/JPROC.2003.823141.
16. Arasaratnam, I., and S. Haykin. 2009. Cubature Kalman filters. *IEEE T. Automat. Contr.* 54(6):1254–1269. doi: 10.1109/TAC.2009.2019800.
17. Wang, T., L. Zhang, and S. Liu. 2022. Improved robust high-degree cubature Kalman filter based on novel cubature formula and maximum correntropy criterion with application to surface target tracking. *J. Marine Science Engineering* 10(8):1070. doi: 10.3390/jmse10081070.
18. Christ, R. D., and R. L. Wernli. 2013. *The ROV manual: A user guide for remotely operated vehicles*. 2nd ed. Oxford, U.K.: Butterworth-Heinemann. 712 p.
19. Li, L., Y. Li, Y. Zhang, G. Xu, J. Zeng, and X. Feng. 2022. Formation control of multiple autonomous underwater vehicles under communication delay, packet discreteness and dropout. *J. Marine Science Engineering* 10(7):920. doi: 10.3390/jmse10070920.
20. Zhao, L., J. Wang, T. Yu, K. Chen, and A. Su. 2019. Incorporating delayed measurements in an improved high-degree cubature Kalman filter for the nonlinear state estimation of chemical processes. *ISA T.* 86:122–133. doi: 10.1016/j.isatra.2018.11.004.
21. Bertsekas, D. P., and S. E. Shreve. 1978. *Stochastic optimal control: The discrete-time case*. New York, NY: Academic Press. 330 p.
22. Pugachev, V. S., and I. N. Sinitsyn. 1990. *Stokhasticheskie differentsial'nye sistemy. Analiz i fil'tratsiya* [Stochastic differential systems. Analysis and filtering]. Moscow: Nauka. 632 p.
23. Sinitsyn, I. N., and E. R. Korepanov. 2015. Normal'nye uslovno-optimal'nye fil'try Pugacheva dlya differentsial'nykh stokhasticheskikh sistem, lineynykh otositel'no sostoyaniya [Normal Pugachev filters for state linear stochastic systems]. *Informatika i ee Primeneniya — Inform Appl.* 9(2):30–38. doi: 10.14357/19922264150204.

Received May 26, 2023

## Contributor

**Bosov Alexey V.** (b. 1969) — Doctor of Science in technology, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AVBosov@ipiran.ru

# РЫНОК С МАРКОВСКОЙ СКАЧКООБРАЗНОЙ ВОЛАТИЛЬНОСТЬЮ II: АЛГОРИТМ ВЫЧИСЛЕНИЯ СПРАВЕДЛИВОЙ ЦЕНЫ ДЕРИВАТИВОВ\*

А. В. Борисов<sup>1</sup>

**Аннотация:** Вторая часть цикла посвящена численной реализации задачи моделирования справедливой цены производных финансовых инструментов (деривативов) в модели неполного рынка с марковской скачкообразной волатильностью. Концепция рыночной цены риска, распространенная в Runggaldier (2004) на класс рисков базовых активов, позволила в первой части цикла получить систему дифференциальных уравнений в частных производных, описывающих временную эволюцию цены деривативов как функцию текущей цены базового актива и скрытой волатильности — обобщение классического уравнения Блэка–Шоулза. В отличие от последнего, полученная система не допускает аналитического решения. Для этого в работе предложено использовать приближенно-аналитический метод дробных шагов. Временная шкала разбивается сеткой, и искомое решение аппроксимируется комбинацией решений классического уравнения теплопроводности и системы обыкновенных линейных дифференциальных уравнений. Свойства полученных решений уравнений и смоделированных с их помощью цен деривативов проиллюстрированы комплексом численных экспериментов.

**Ключевые слова:** марковский скачкообразный процесс; оптимальная фильтрация; стохастическая волатильность; рыночная цена риска; преобладающая мартингальная мера

DOI: 10.14357/19922264230303

EDN: DNJXJB

## 1 Необходимые сведения и постановка задачи

В цикле рассматривается модель неполного финансового рынка со стохастической волатильностью, состоящего из банковского вклада с детерминированной ставкой  $r_t$ ,  $N$  базовых финансовых инструментов  $S$  и  $M$  производных финансовых инструментов (деривативов)  $F$ , определяемых платежным требованием  $H(S)$ . Стохастическая волатильность определяется внешним скрытым марковским скачкообразным процессом (МСП)

$$Z_t \triangleq \text{col}(Z_t^1, \dots, Z_t^L) \in \{e_1, \dots, e_L\}$$

с матричнозначной функцией интенсивности переходов  $\Lambda(\cdot)$  и начальным распределением  $\pi_0^Z$ . Этот процесс является единственным сильным решением линейной стохастической дифференциальной системы (СДС) с мартингалом  $M_t$  в правой части:

$$dZ_t = \Lambda^\top(t)Z_t dt + dM_t, \quad t \in (0, T], \quad Z_0 \sim \pi_0^Z. \quad (1)$$

Эволюция цен базовых финансовых активов  $S_t \triangleq \text{col}(S_t^1, \dots, S_t^N)$  задается единственным сильным решением СДС

$$dS_t = \text{diag } S_t a(t, Z_t) dt + \text{diag } S_t \sigma(t, Z_t) dw_t, \quad t \in (0, T], \quad S_0 \sim \pi_0^S, \quad (2)$$

где  $w_t \triangleq \text{col}(w_t^1, \dots, w_t^K)$  —  $K$ -мерный стандартный винеровский процесс ( $K \geq N$ ), а случайные функции мгновенной процентной ставки  $a$  и внутренней волатильности  $\sigma$  «модулируются» МСП  $Z_t$  как

$$a(t, Z_t) = \sum_{\ell=1}^L Z_t^\ell a^\ell(t); \quad \sigma(t, Z_t) = \sum_{\ell=1}^L Z_t^\ell \sigma^\ell(t)$$

с известными наборами детерминированных функций  $\{a^\ell(t)\}_{\ell=1, \dots, L}$  и  $\{\sigma^\ell(t)\}_{\ell=1, \dots, L}$ .

Срок погашения всех деривативов совпадает с  $T$ , погашение происходит в соответствии с платежным требованием

$$H(S_T) \triangleq \text{col}(H^1(S_T), \dots, H^M(S_T)).$$

Концепция рыночной цены риска, распространенная в [1] на класс рисков базовых активов, позволила в [2] доказать, что в условиях безарбитражности данного рынка при наличии мартингальной меры справедливая цена  $F^m$   $m$ -го дериватива определяется формулой:

\* Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

<sup>1</sup> Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук; aborisov@frccsc.ru

$$F^m = F^m(t, S_t, Z_t) = \sum_{\ell=1}^L Z_t^\ell F^{m\ell}(t, S_t), \quad (3)$$

а функции  $\{F^{m\ell}(t, s)\}_{\ell=1, \dots, L}$  являются решением задачи Коши для следующей системы дифференциальных уравнений в частных производных:

$$\left. \begin{aligned} F_t^{m\ell} &= rF^{m\ell} - \sum_{j=1}^L \Lambda^{\ell j} F^{mj} - \\ &- \sum_{n=1}^N F_{s^n}^{m\ell} s^n (r - a_n^\ell) - \frac{1}{2} \sum_{i,j=1}^N s^i s^j F_{s^i s^j}^{m\ell} B_{ij}^\ell, \\ &\ell = \overline{1, M}, \quad m = \overline{1, M}, \quad t \in [0, T]; \\ F^{m\ell}(T, s) &= H^m(s), \end{aligned} \right\} \quad (4)$$

где

$$B^\ell = \|B_{ij}^\ell\|_{i,j=\overline{1, N}} \triangleq \sigma^\ell (\sigma^\ell)^\top.$$

Эти системы могут трактоваться как некоторые обобщения классического уравнения Блэка–Шоулза. При этом цены  $F^m$  допускают следующий стохастический дифференциал:

$$\begin{aligned} dF^m &= \sum_{\ell=1}^L Z^\ell \left[ rF^{m\ell} + \sum_{n=1}^N F_{s^n}^{m\ell} S^n (a_n^\ell - r) \right] dt + \\ &+ \sum_{\ell=1}^L F^{m\ell} dM^\ell + \sum_{\ell=1}^L Z^\ell \sum_{n=1}^N \sum_{k=1}^K F_{s^n}^{m\ell} S^n \sigma_{nk}^\ell dw^k, \\ &m = \overline{1, M}, \quad \ell = \overline{1, L}, \quad F^{m\ell} = F^{m\ell}(t, S_t). \end{aligned} \quad (5)$$

Задача данной статьи заключается в разработке алгоритма моделирования цены деривативов. Стохастический дифференциал (5) не подходит для этой цели: он необходим при решении задачи оценивания состояния скрытого МСП по наблюдениям  $S$  и  $F$ , что станет предметом исследований последующих частей цикла. Генерация траекторий деривативов выполняется только с помощью (3) и (4), поэтому решение системы (4) представляется актуальной задачей.

## 2 Применение алгоритма дробных шагов при решении обобщения уравнения Блэка–Шоулза

Вообще говоря, (4) представляет собой совокупность  $M$  независимых подсистем с  $L$  уравнениями в каждой, причем внутри каждой подсистемы уравнения не могут быть разделены. При этом все уравнения относятся к классу уравнений теплопроводности.

Для упрощения представления алгоритма решения (4) методом дробных шагов сделаем следующие допущения.

1. Матрица интенсивностей переходов  $\Lambda$ , банковский процент  $r$ , мгновенные процентные ставки по базовым активам  $a$  и внутренние волатильности  $\sigma$  не зависят от времени:

$$\Lambda(t) \equiv \Lambda; \quad r_t \equiv r; \quad a^\ell(t) \equiv a^\ell; \quad \sigma^\ell(t) \equiv \sigma^\ell.$$

2. Платежные требования  $H^m(s^1, \dots, s^N)$ , соответствующие отдельным деривативам, определяются одним базовым активом, т. е. для любого  $1 \leq m \leq M$  существует такой номер  $1 \leq n_m \leq N$ , что  $H^m(s) = H^m(s^{n_m})$ .

Сделанные предположения не слишком обременительны. Во-первых, постоянство коэффициентов требуется лишь на отрезках дискретизации по времени. Порядок длины этого отрезка для расчетов не превышает 1 ч (т. е. 0,001 при нормировании к году), и на нем указанные характеристики весьма близки к константам. Во-вторых, большинство реальных деривативов, например европейских опционов, строятся на отдельных базовых активах, а не на их комбинациях.

Рассмотрим один дериватив и порождающий его базовый актив. Для простоты записи будем опускать зависимость от индексов: номера дериватива  $m$  и базового актива  $n$ . Используя обозначения

$$\overline{F} \triangleq \text{col}(F^1(t, s), \dots, F^L(t, s));$$

$$\overline{F}(t) \triangleq \text{col}(F^1(t, \cdot), \dots, F^L(t, \cdot));$$

$$\overline{H} \triangleq \text{col}(H(\cdot), \dots, H(\cdot));$$

$$A \triangleq \text{diag}(a^1, \dots, a^L), \quad B \triangleq \text{diag}(B^1, \dots, B^L);$$

$$\overline{F}_t \triangleq \frac{\partial}{\partial t} \overline{F}(t, s);$$

$$\mathcal{L}_1 \overline{F} \triangleq \left[ r - (rI - A)s \frac{\partial}{\partial s} - \frac{1}{2} Bs^2 \frac{\partial^2}{\partial s^2} \right] \overline{F};$$

$$\mathcal{L}_2 \overline{F} \triangleq -\Lambda \overline{F},$$

однородную задачу Коши (4) можно записать в виде

$$\overline{F}_t = (\mathcal{L}_1 + \mathcal{L}_2) \overline{F}, \quad t \in [0, T), \quad \overline{F}(T, s) = \overline{H}(s),$$

и ее решение будет представимо в виде

$$\overline{F}(t) = \mathcal{L}(t, T) \overline{H},$$

где  $\mathcal{L}(t, T)$  — линейный оператор перехода на промежутке  $[t, T]$ , соответствующий дифференциальному оператору  $(\mathcal{L}_1 + \mathcal{L}_2)$ . Для данной задачи Коши выполнен принцип Адамара:

$$\overline{F}(t) = \mathcal{L}(t, \tau) \overline{F}(\tau) \text{ для любых } 0 < t \leq \tau \leq T,$$

который можно положить в основу рекуррентной процедуры вычисления решения  $\bar{F}$  на временной сетке  $\{t_j\}_{j=1, \overline{J}}$ :  $t_j = jh$ ,  $h = T/J$ :

$$\begin{aligned} \bar{F}(t_{j-1}) &= \mathfrak{L}(t_{j-1}, t_j) \bar{F}(t_j), \\ j &= \overline{1, J}, \quad \bar{F}(t_J, s) = \bar{H}(s). \end{aligned} \quad (6)$$

Поиск оператора  $\mathfrak{L}(t, T)$  представляется нетривиальной задачей, и исходную задачу Коши можно решать с помощью различных вариантов метода конечных разностей [3]. В данной работе предлагается решать ее методом дробных шагов [4]. Для этого рассмотрим две вспомогательные однородные задачи Коши:

$$\bar{R}_t = \mathcal{L}_1 \bar{R}, \quad t \in [0, T), \quad \bar{R}(T, s) = \bar{H}_R(s); \quad (7)$$

$$\bar{Q}_t = \mathcal{L}_2 \bar{Q}, \quad t \in [0, T), \quad \bar{Q}(T, s) = \bar{H}_Q(s), \quad (8)$$

решения которых также могут быть выражены через соответствующие операторы перехода  $\mathcal{L}_1$  и  $\mathcal{L}_2$  (ниже они также представлены в форме принципа Адамара):

$$\bar{R}(t) = \mathfrak{L}_1(t, \tau) \bar{R}(\tau); \quad \bar{Q}(t) = \mathfrak{L}_2(t, \tau) \bar{Q}(\tau)$$

для любых  $0 < t \leq \tau \leq T$ .

Решение (6)  $\{\bar{F}(t_j, s)\}_{j=0, \overline{J}}$  предлагается аппроксимировать, используя вместо  $\mathfrak{L}$  композицию переходных операторов  $\mathcal{L}_1$  и  $\mathcal{L}_2$ :

$$\begin{aligned} \tilde{F}(t_{j-1}) &= \mathfrak{L}_2(t_{j-1}, t_j) \mathfrak{L}_1(t_{j-1}, t_j) \tilde{F}(t_j), \\ j &= \overline{1, J}, \quad \tilde{F}(t_J) = \bar{H}. \end{aligned} \quad (9)$$

Операторы  $\mathcal{L}_1$  и  $\mathcal{L}_2$  могут быть легко найдены.

Все уравнения системы (7) независимы, и они могут быть преобразованы к стандартному уравнению теплопроводности так же, как это сделано для классического уравнения Блэка–Шоулза [5]. Рассмотрим одну из таких задач Коши (индекс  $\ell$  опустим для простоты записи)

$$\begin{aligned} R_t(t, s) &= rR(t, s) - (r-a)sR_s(t, s) - \frac{b}{2}s^2R_{ss}(t, s), \\ 0 &\leq t < T, \quad R(T, s) = H_R(s), \end{aligned} \quad (10)$$

и решим ее с помощью следующей последовательности замен.

1. Искомая функция  $R$  заменяется на  $V$ :  $V(t, s) \triangleq e^{r(T-t)}R(t, s)$ . Задача Коши (10) для новой функции  $V$  принимает вид:

$$V_t(t, s) = -(r-a)sV_s(t, s) - \frac{b}{2}s^2V_{ss}(t, s), \quad (11)$$

$$0 \leq t < T, \quad V(T, s) = H_R(s).$$

2. Вводится новая переменная  $\tau(t) = T-t$  и функция  $U(\tau(t), s) \triangleq V(t, s)$ . Задача Коши (11) для  $U$  принимает вид:

$$\begin{aligned} U_\tau(\tau, s) &= (r-a)sU_s(\tau, s) + \frac{b}{2}s^2U_{ss}(\tau, s), \\ 0 < \tau &\leq T, \quad U(0, s) = H_R(s). \end{aligned} \quad (12)$$

3. Вводится новая переменная  $x(\tau, s) = \ln s + (r-a-(b/2))\tau$  и функция  $G(\tau, x(\tau, s)) \triangleq U(\tau, s)$ . Задача Коши (12) для  $G$  принимает вид:

$$\begin{aligned} G_\tau(\tau, x) &= \frac{b}{2}x^2G_{xx}(\tau, x), \\ 0 < \tau &\leq T, \quad G(0, x(0, s)) = H_R(s). \end{aligned} \quad (13)$$

Решение (13) известно:

$$G(\tau, x) = \int_{-\infty}^{\infty} \Phi(\tau, x-y)H_R(y) dy,$$

где

$$\Phi(u, x) \triangleq \frac{1}{\sqrt{2\pi bu}} \exp\left(-\frac{x^2}{2bu}\right).$$

Производя обратные подстановки, можно получить окончательный покомпонентный вид переходного оператора  $\mathcal{L}_1 = \text{col}(\mathcal{L}_1^1, \dots, \mathcal{L}_1^L)$  на одном временном шаге:

$$\begin{aligned} \mathfrak{L}_1^\ell(t_{j-1}, t_j)R^\ell &\triangleq \\ &\triangleq e^{-r\tau} \int_0^\infty \frac{\Phi(h, \ln s/y + (r-a^\ell - B^\ell/2)h)}{y} R^\ell(y) dy, \\ &\ell = \overline{1, L}. \end{aligned}$$

Система (8) состоит из однородных обыкновенных линейных дифференциальных уравнений, оператор перехода  $\mathcal{L}_2$  для нее определен следующим образом:

$$\mathfrak{L}_2(t_{j-1}, t_j)Q = e^{-h\Lambda}Q$$

для любых  $0 \leq t \leq \tau \leq T$ .

Окончательно рекуррентная процедура (9) послойного вычисления приближенного решения  $\bar{F}$  принимает вид

$$\begin{aligned} \tilde{F}(t_{j-1}, s) &= e^{-h(rI+\Lambda)} \int_0^\infty \Psi(s, y) \tilde{F}(t_j, y) dy, \\ j &= \overline{1, J}, \quad \tilde{F}(t_J, s) = \bar{H}(s), \end{aligned} \quad (14)$$

где

$$\Psi(s, y) \triangleq \text{diag} \left( \frac{1}{y\sqrt{2\pi B^1 h}} \times \right. \\ \left. \times e^{-((\ln s/y + (r - a^1 - B^1/2)h)^2)/(2B^1 h)}, \dots \right. \\ \left. \dots, \frac{1}{y\sqrt{2\pi B^L h}} e^{-((\ln s/y + (r - a^L - B^L/2)h)^2)/(2B^L h)} \right).$$

Интеграл (14) не может быть вычислен аналитически и заменяется некоторой аппроксимацией. В данном цикле работ используется метод трапеций с равномерной сеткой с шагом  $\Delta$ . Область интегрирования — положительная полуось — заменяется конечным отрезком  $[\underline{s}, \bar{s}]$ , который выбирается, исходя из условия  $E \{S_T(1 - \mathbf{I}_{[\underline{s}, \bar{s}]}(S_T))\} \leq \Delta^2$ . В этом случае аппроксимация  $\tilde{F}$  обеспечивает на каждом шаге точность вычисления слоя  $O(h^2 + \Delta^2)$  [4]. Следующий раздел содержит результаты численного эксперимента, иллюстрирующего качество предложенного алгоритма вычисления справедливой цены дериватива и его совместное поведение с базовым активом.

### 3 Комплекс численных экспериментов

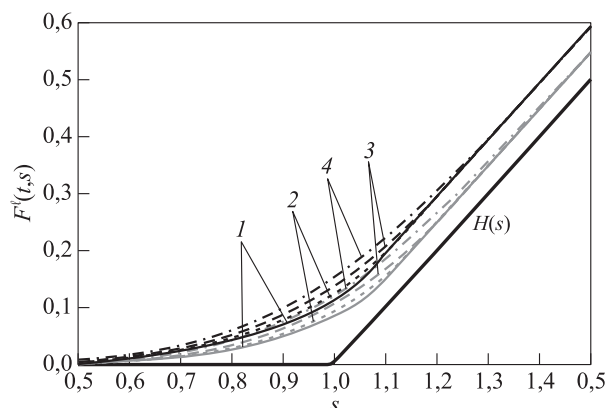
Первый эксперимент посвящен численному анализу эволюции во времени функций  $\{F^\ell\}$ . Параметры финансовой системы в данном примере искусственно выбраны так, чтобы наглядно продемонстрировать разницу  $F^\ell$  для разных значений  $\ell$ .

На отрезке времени  $[0, 1]$  (1 год, 250 торговых дней по 8 ч каждый) моделируется поведение одного базового актива (2) ( $N = 1$ ), стохастическая волатильность которого представляет собой МСП (1) с четырьмя возможными состояниями ( $L = 4$ ). Ставка банковского вклада  $r = 10\%$  годовых, на рынке присутствует один дериватив ( $M=1$ ) — call-опцион с ценой исполнения  $K = 1$ . Для системы (2), (1) были выбраны следующие параметры:

$$S_0 \equiv 1; a = (0,8; 0,6; 0,4; 0,2); \sigma = (0,2; 0,4; 0,6; 0,8);$$

$$\Lambda = \begin{bmatrix} -15 & 5 & 5 & 5 \\ 5 & -15 & 5 & 5 \\ 5 & 5 & -15 & 5 \\ 5 & 5 & 5 & -15 \end{bmatrix}; \pi_0^Z = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}.$$

Интегрирование СДС (2) выполнялось методом Эйлера–Маруямы [6], адаптированным к скачкам [7], с основным шагом  $\tau = 10^{-7}$ , интегрирование системы уравнений в частных производных (4)



**Рис. 1** Зависимость  $F^\ell(t, s)$  от цены  $s$  для некоторых значений времени  $t$ : чёрные кривые —  $t = 0$ ; серые кривые —  $t = 0,5$ ; 1 —  $\ell = 1$ ; 2 — 2; 3 — 3; 4 —  $\ell = 4$

выполнялось с шагом по времени  $h = 0,0001$  (шаг соответствует 6 мин торгового времени) и шагом по координате  $\delta = 0,002$  (шаг соответствует 0,2 процентным пунктам начальной цены базового актива).

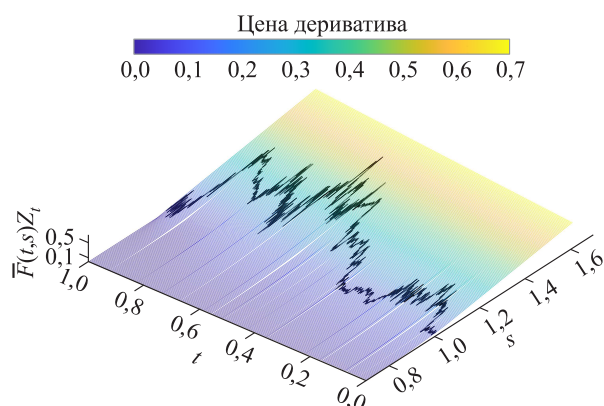
На рис. 1 представлены сечения решений  $\{F^\ell(t, s)\}_{\ell=1,4}$  системы (4) при фиксированных значениях времени:  $t = 0$  — начало года и  $t = 0,5$  — полгода до погашения. Для сравнения на рисунке представлена и функция платежного требования  $H(s)$ , с которой совпадают все функции  $F^\ell(t, s)$  в момент погашения  $t = 1$ . Действительно, при разных значениях  $\ell$  и фиксированных моментах  $t$  графики  $F^\ell(\cdot, s)$  значительно отличаются друг от друга. Помимо этого рисунок также демонстрирует изменение функций  $F^\ell(t, \cdot)$  с течением времени.

Второй численный эксперимент призван продемонстрировать сущность совместной генерации цены базового актива и его производного финансового инструмента. Случайность цены опциона  $F_t$  — это результат влияния двух случайных факторов: текущей цены базового актива  $S_t$  и скрытого состояния рынка  $Z_t$  (3):

$$F_t = \bar{F}(t, S_t) Z_t.$$

Графическая иллюстрация этой комбинации представлена на рис. 2. Трёхмерная поверхность представляет собой произведение  $\bar{F}(t, s)Z_t$ , и на ней траектория базового актива «вычерчивает» пространственную кривую, аппликата которой и представляет собой цену дериватива. Следует отметить, что эта поверхность не непрерывна: ее разрывы вызваны скачками МСП  $Z_t$ . Эти разрывы также порождают скачки дериватива  $F$ .

Третий эксперимент иллюстрирует совместное поведение базового актива и различных деривативов — call-опционов с разными ценами исполне-



**Рис. 2** Зависимость  $F^\ell(t, s)$  от цены  $s$  для некоторых значений времени  $t$

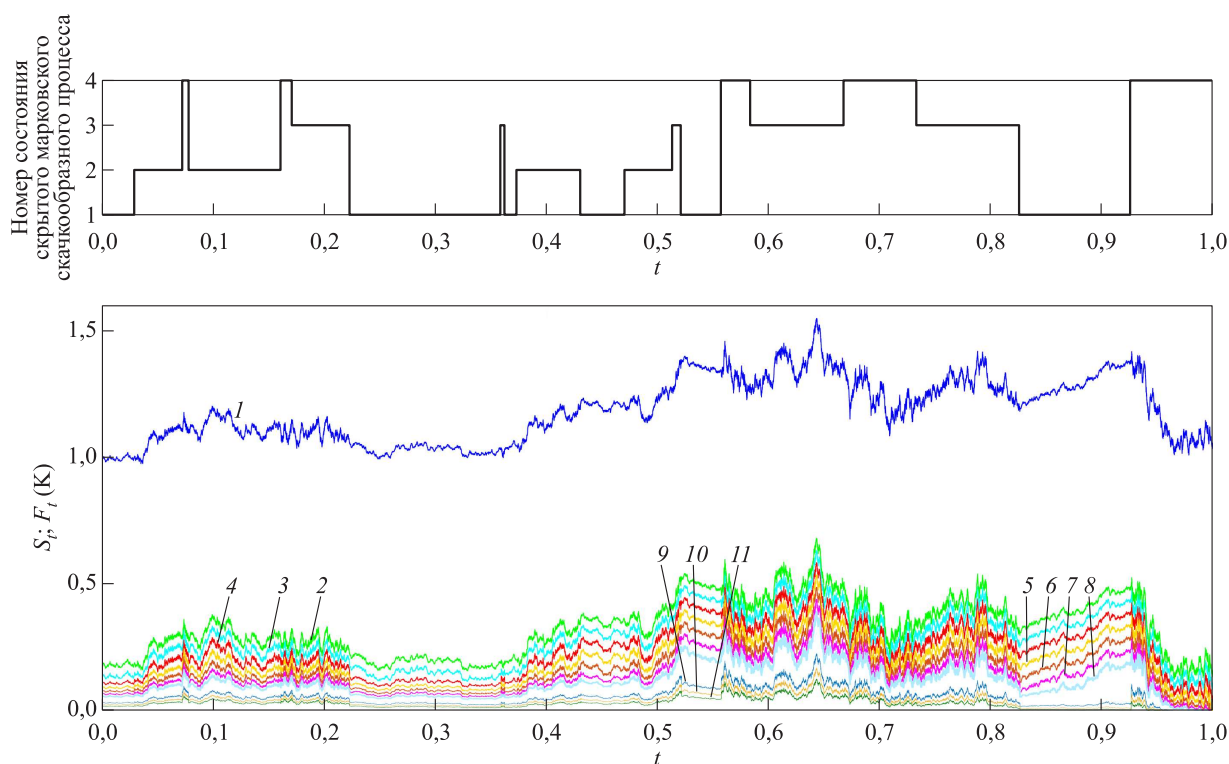
ния  $K$ . Колебания цен деривативов, представленные на рис. 3, похожи на колебания котировок базового актива, однако не прямо пропорциональны им. Особенно наглядно это видно в случае с опционами, цена исполнения которых выше текущей цены базового актива (так называемые опционы *out of the money*).

Формула (3), определяющая цену деривативов, подразумевает также наличие ее скачков, совпадающих по времени со скачками внешнего МСП  $Z_t$ .

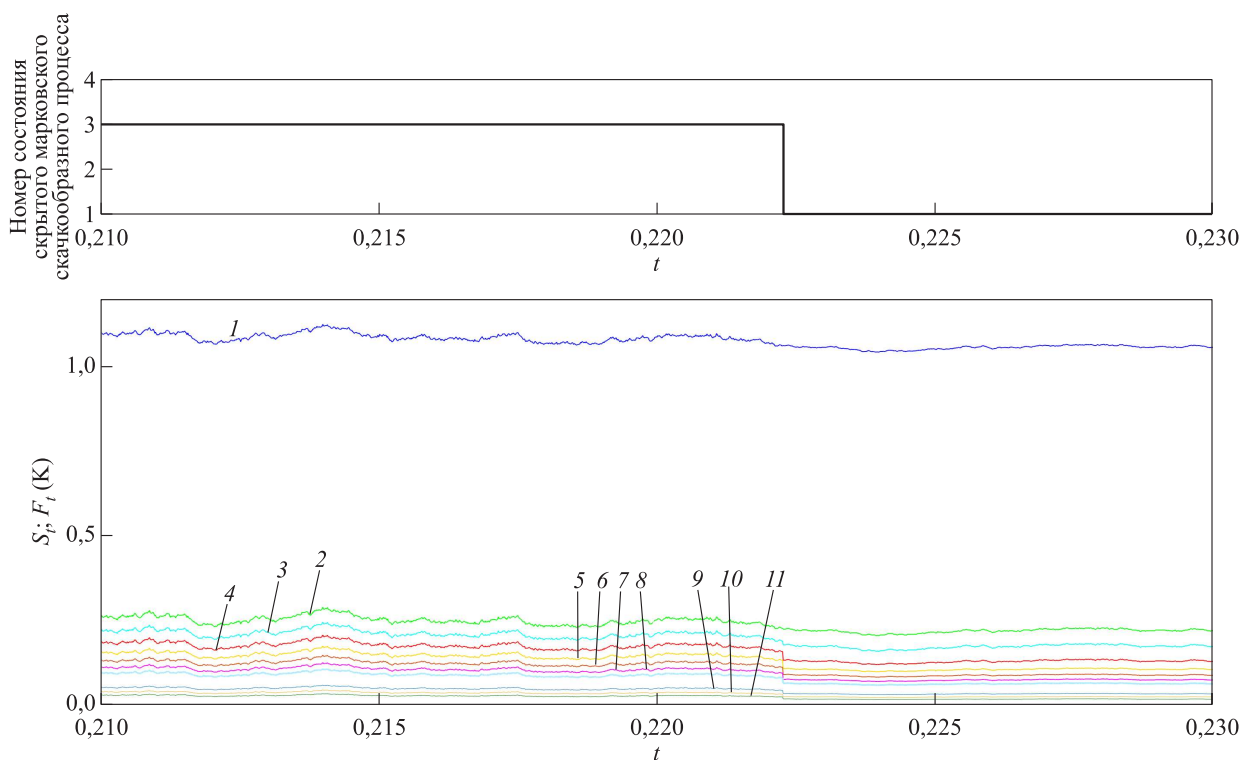
На рис. 4 представлен увеличенный фрагмент рис. 3: на нем на интервале  $[0,22; 0,225]$  виден скачок МСП и порожденные им скачки цен деривативов. Примечательно, что не для всех цен исполнения скачки имеют явно выраженный характер. Причину этому легко понять, анализируя рис. 1. Дело в том, что амплитуда скачка зависит от того, насколько отличаются значения функций  $F^\ell(\cdot, \cdot)$  для различных  $\ell$ . Из рис. 4 можно сделать вывод, что это различие незначительно при достаточно малых значениях аргумента  $s$  и исчезающе мало при  $s \rightarrow +\infty$ . Таким образом, различия между  $F^\ell(\cdot, \cdot)$ , а значит, и величины скачков цены дериватива существенны, когда цена базового актива близка к цене исполнения (опционы с такими страйками называются *at the money*).

## 4 Промежуточные выводы

Вторая часть цикла статей представляет алгоритм вычисления справедливой цены производных финансовых инструментов в моделях рынка с марковской скачкообразной волатильностью. Во-первых, данный алгоритм необходим при решении широкого класса задач анализа цен деривативов в предложенной модели финансовой системы. Во-



**Рис. 3** Скрытый марковский скачкообразный процесс и цены базового актива  $S_t$  ( $I$ ) и *call*-опционов  $F_t(K)$  в зависимости от разных цен исполнения  $K$ : 2 —  $K = 0,9$ ; 3 —  $0,95$ ; 4 —  $1$ ; 5 —  $1,05$ ; 6 —  $1,1$ ; 7 —  $1,15$ ; 8 —  $1,2$ ; 9 —  $1,4$ ; 10 —  $1,5$ ; 11 —  $K = 1,6$



**Рис. 4** Скрытый марковский скачкообразный процесс и цены базового актива  $S_t$  (1) и  $call$ -опционов  $F_t(K)$  в зависимости от разных значений цены исполнения  $K$ : 2 —  $K = 0,9$ ; 3 —  $0,95$ ; 4 —  $1$ ; 5 —  $1,05$ ; 6 —  $1,1$ ; 7 —  $1,15$ ; 8 —  $1,2$ ; 9 —  $1,4$ ; 10 —  $1,5$ ; 11 —  $K = 1,6$ . Увеличенный фрагмент со скачком

вторых, он становится ключевым при совместном численном моделировании эволюции цен базовых финансовых инструментов и их деривативов. Эта возможность нужна для проведения сравнительного численного анализа решения задачи мониторинга рыночной цены риска как частного случая задачи фильтрации МСП по имеющимся наблюдениям цен бумаг. В последующих частях цикла будут представлены алгоритмы решения этой задачи оценивания в зависимости от структуры имеющихся наблюдений: доступными будут цены инструментов, дискретизованные по времени с известным детерминированным шагом [8, 9], либо высокочастотные потоки наблюдений цен в случайные моменты времени [10].

## Литература

1. *Runggaldier W.* Estimation via stochastic filtering in financial market models // *Contemp. Math.*, 2004. Vol. 351. P. 309–318. doi: 10.1090/conm/351/06412.
2. *Борисов А.* Рынок с марковской скачкообразной волатильностью I: мониторинг цены риска как задача оптимальной фильтрации // *Информатика и её применения*, 2023. Т. 17. Вып. 2. С. 27–33. doi: 10.14357/19922264230204. EDN: GAXCHQ.
3. *Тихонов А., Самарский А.* Уравнения математической физики. — М.: Наука, 2004. 798 с.
4. *Яненко Н.* Метод дробных шагов решения многомерных задач математической физики. — Новосибирск: Наука, 1967. 197 с.
5. *Shiryayev A.* Essentials of stochastic finance: Facts, models, theory. — Hackensack, NJ, USA: World Scientific, 1999. 834 p.
6. *Kloeden P., Platen E.* Numerical solution of stochastic differential equations. — Berlin: Springer, 1992. 636 p.
7. *Platen E., Bruti-Liberati N.* Numerical solution of stochastic differential equations with jumps in finance. — Berlin: Springer, 2010. 884 p.
8. *Борисов А.* Численные схемы фильтрации марковских скачкообразных процессов по дискретизованным наблюдениям II: случай аддитивных шумов // *Информатика и её применения*, 2020. Т. 14. Вып. 1. С. 17–23. doi: 10.14357/19922264200103.
9. *Борисов А.* Численные схемы фильтрации марковских скачкообразных процессов по дискретизованным наблюдениям III: случай мультипликативных шумов // *Информатика и её применения*, 2020. Т. 14. Вып. 2. С. 10–18. doi: 10.14357/19922264200202.
10. *Борисов А.* Алгоритм робастной фильтрации марковских скачкообразных процессов по высокочастотным считающим наблюдениям // *Автоматика и телемеханика*, 2020. Вып. 4. С. 3–20.

Поступила в редакцию 29.12.22

# MARKET WITH MARKOV JUMP VOLATILITY II: ALGORITHM OF DERIVATIVE FAIR PRICE CALCULATION

A. V. Borisov

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The second part of the series is devoted to the numerical realization of the fair derivative price in the case of the incomplete financial market with a Markov jump stochastic volatility. The concept of the market price of risk applied to this model by Runggaldier (2004) allows derivation of a system of partial differential equations describing the temporal evolution of the derivative price as a function of the current underlying price and the implied stochastic volatility. This system represents a generalization of the classic Black–Sholes equation. By contrast with this classic version, the proposed system of equations does not permit an analytical solution. The paper presents an approximate analytical method of the fractional steps. The author equips the temporal axis with a grid, then the author approximates the required solution as a combination of the solution to the classical heat equation and the system of the ordinary linear differential equations. The paper contains the results of the numerical experiments illustrating the properties of both the Black–Sholes generalization and the joint evolution of the derivative and corresponding underlying prices.

**Keywords:** Markov jump process; optimal filtering; stochastic volatility; market price of risk; prevailing martingale measure

**DOI:** 10.14357/19922264230303

**EDN:** DNXJGB

## Acknowledgments

The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (CKP “Informatics”) of FRC CSC RAS (Moscow).

## References

1. Runggaldier, W. 2004. Estimation via stochastic filtering in financial market models. *Contemp. Math.* 351:309–318. doi: 10.1090/conm/351/06412.
2. Borisov, A. 2023. Rynok s markovskoy skachkoobraznoy volatil’nost’yu I: monitoring tseny riska kak zadacha optimal’noy fil’tratsii [Market with Markov jump volatility I: Price of risk monitoring as an optimal filtering problem]. *Informatika i ee primeneniya — Inform. Appl.* 17(2): 27–33. doi: 10.14357/19922264230204. EDN: GAXCHQ.
3. Tikhonov, A., and A. Samarskiy. 2004. *Urvneniya matematicheskoy fiziki* [Equations of mathematical physics]. Moscow: Nauka. 798 p.
4. Yanenko, N. 1967. *Metod drobnykh shagov resheniya mnogomernykh zadach matematicheskoy fiziki* [The fractional steps method for solving multidimensional problems of mathematical physics]. Novosibirsk: Nauka. 197 p.
5. Shirayev, A. 1999. *Essentials of stochastic finance: Facts, models, theory*. Hackensack, NJ: World Scientific. 834 p.
6. Kloeden, P., and E. Platen. 1992. *Numerical solution of stochastic differential equations*. Berlin: Springer. 636 p.
7. Platen, E., and N. Bruti-Liberati. 2010. *Numerical solution of stochastic differential equations with jumps in finance*. Berlin: Springer. 884 p.
8. Borisov, A. 2020. Chislennye skhemy fil’tratsii markovskikh skachkoobraznykh protsessov po diskretizovannym nablyudeniya II: sluchay additivnykh shumov [Numerical schemes of Markov jump process filtering given discretized observations II: Additive noises case]. *Informatika i ee primeneniya — Inform. Appl.* 14(1): 17–23. doi: 10.14357/19922264200103.
9. Borisov, A. 2020. Chislennye skhemy fil’tratsii markovskikh skachkoobraznykh protsessov po diskretizovannym nablyudeniya III: sluchay mul’tiplikativnykh shumov [Numerical schemes of Markov jump process filtering given discretized observations III: Multiplicative noises case]. *Informatika i ee primeneniya — Inform. Appl.* 14(2): 10–18. doi: 10.14357/19922264200202.
10. Borisov, A. 2020. Robust filtering algorithm for Markov jump processes with high-frequency counting observations. *Automat. Rem. Contr.* 81(4):575–588. doi: 10.1134/S0005117920040013.

Received December 29, 2022

## Contributor

**Borisov Andrey V.** (b. 1965) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; aborisov@frcsc.ru



# МОДЕЛИРОВАНИЕ НАСТОЙЧИВОГО ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ В СЕТЯХ 5G NR С АДАПТАЦИЕЙ СКОРОСТИ И БЛОКИРОВКАМИ\*

Э. С. Сопин<sup>1</sup>, А. Р. Маслов<sup>2</sup>, В. С. Шоргин<sup>3</sup>, В. О. Бегишев<sup>4</sup>

**Аннотация:** Технология радиодоступа 5G NR (New Radio), работающая в диапазоне миллиметровых волн, и будущие терагерцевые системы 6G предназначены для приложений, чувствительных к скорости. Такие приложения характеризуются адаптивностью, позволяющей снизить скорость передачи в соответствии с текущими условиями сети. Ненадежный характер сетей 5G/6G может вызывать повторные попытки продолжить обслуживание. Предложена модель обслуживания абонентов с нетерпеливым поведением на основе ресурсной системы массового обслуживания (РесМО) с орбитой. В качестве характеристик обслуживания рассматриваются вероятности блокировки доступа на обслуживание и прерывания заявки, а также коэффициент использования ресурсов системы. Показано, что настойчивость пользователей позволяет понизить рассматриваемые вероятности: выполнение в среднем двух повторных попыток снижает оба показателя на 20%–70%. Настойчивость увеличивает использование системных ресурсов на 20%–40% и снижает долю потраченных впустую ресурсов в 2–3 раза.

**Ключевые слова:** 5G; NR (New Radio); ресурсная СМО; повторные вызовы; блокировки распространения; прерывание обслуживания

**DOI:** 10.14357/19922264230304

**EDN:** ENSHKV

## 1 Введение

Технология 5G NR, работающая в диапазоне миллиметровых волн (28–100 ГГц), разработана для удовлетворения потребностей мультимедийных приложений. В будущих системах 6G эти системы будут дополнены технологиями радиодоступа, работающими в диапазоне частот 100–300 ГГц и обладающими схожими характеристиками распространения радиоволн.

Мультимедийные приложения требуют услуги поддержания постоянной скорости передачи данных на радиоинтерфейсе. Однако даже эти приложения часто снабжены возможностями адаптации скорости. Приложение может отслеживать состояние соединения и реагировать на снижение пропускной способности сети и/или временную потерю соединения путем снижения требуемой скорости. В системах связи с прерываниями активного соединения, таких как 5G/6G сети, пользователи будут воспринимать услугу с пониженным качеством.

Принципиальное различие между микроволновыми технологиями, такими как LTE (long-term evo-

lution), и системами миллиметровых волн и терагерцевого диапазона частот заключается в ненадежном характере беспроводного канала вследствие блокировки путей распространения радиоволн, что приводит к потерям соединения [1]. На ранних этапах развертывания потери соединения будут приводить к «нетерпеливому» поведению пользователей — попыткам восстановления активной сессии передачи данных.

Цель работы состоит в анализе эффектов влияния нетерпеливого поведения пользователей в случае потери связи на характеристики обслуживания абонентов в сетях радиодоступа 5G/6G в условиях динамической адаптации скорости передачи и динамической блокировки. Предложена математическая модель процесса обслуживания абонентов на основе РесМО с орбитой и повторными вызовами.

В отличие от похожей модели с неоднородными заявками и орбитой [2], в данной работе модель учитывает специфику миллиметровых волн / терагерцевых частот, включая модели распространения, антенны и блокировки прямой видимости. В качестве характеристик обслуживания рассматриваются вероятность блокировки доступа на обслуживание,

\* Исследование выполнено за счет гранта Российского научного фонда № 22-79-10128.

<sup>1</sup> Российский университет дружбы народов им. Патриса Лумумбы; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sopin-es@rudn.ru

<sup>2</sup> Российский университет дружбы народов им. Патриса Лумумбы, maslov-ar@rudn.ru

<sup>3</sup> Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vshorgin@ipiran.ru

<sup>4</sup> Российский университет дружбы народов им. Патриса Лумумбы, begishev-vo@rudn.ru

вероятность прерывания сессии, а также коэффициент использования ресурсов системы.

## 2 Системная модель

Рассматривается базовая станция (БС) миллиметрового/терагерцевого диапазона с радиусом покрытия  $R_M$  (рис. 1). Зона покрытия соты равна  $d_E$ , зависит от используемой схемы модуляции и кодирования и рассчитывается с использованием модели распространения 3GPP (3rd Generation Partnership Project) [3]. Доступная полоса пропускания —  $B$  МГц. В зоне обслуживания БС находятся пешеходы с плотностью  $\lambda_B$  пешеход/км<sup>2</sup>, движущиеся согласно модели случайных направлений (RDM — random direction model) [4]. Базовая станция обслуживает абонентские устройства (АУ), ассоциированные с пешеходами. Высоты БС и АУ равны  $h_A$  и  $h_U$  соответственно. Высота пешеходов равна  $h_P$ ,  $h_P > h_U$ .

Пусть  $\lambda_U$  — интенсивность поступления заявок от одного АУ. Предполагая, что АУ генерируют запросы на установление сессий независимо, процесс поступления можно считать пуассоновским с интенсивностью  $\lambda_A = \lambda_U \pi d_E^2$ . Предполагается равномерное распределение положения пешеходов в радиусе обслуживания соты.

Рассматриваются услуги с высокой скоростью передачи данных, создающие неэластичный, но адаптирующийся к доступной скорости трафик. Каждая заявка запрашивает постоянную скорость  $C_{max}$ . В случае прерывания обслуживания, вызванного блокировкой, механизм адаптации снижает скорость до  $C_{min}$ ,  $C_{min} < C_{max}$ .

Вследствие случайного расположения АУ в зоне обслуживания фактический объем ресурса, требуемый для достижения скоростей  $C_{min}$  и  $C_{max}$ , считается случайным. Предполагается, что время обслуживания запроса на установление сессии подчиняется экспоненциальному распределению с параметром  $\mu$ .

В работе учитывается специфика миллиметрового/терагерцевого диапазонов, включая модели распространения, антенны и блокировки, аналогичные рассмотренным в [5–7].

## 3 Анализ математической модели

### 3.1 Описание модели

Рассматривается РеСМО с  $N$  приборами,  $R$  единицами ресурса и орбитой емкостью  $M$  заявок. Поступающий поток заявок пуассоновский с интенсивностью  $\lambda$ , время их обслуживания распределено экспоненциально с параметром  $\mu$ . Для каждой поступающей в систему заявки требуется прибор и случайный объем ресурса, определяемый в соответствии с распределением  $\{p_{1,j}\}$ ,  $j = 1, 2, \dots, R$ . Поступающая заявка принимается на обслуживание, если в системе есть свободный прибор и достаточный объем ресурса. Если хотя бы одно из условий не выполняется, то заявка либо уходит на орбиту с вероятностью  $\theta$ , либо покидает систему с вероятностью  $1 - \theta$ . Заявки проводят на орбите экспоненциально распределенное время с интенсивностью  $\alpha$  и затем пытаются повторно поступить на обслуживание с новыми требованиями к ресурсу в соответствии с распределением  $\{p_{2,j}\}$ ,

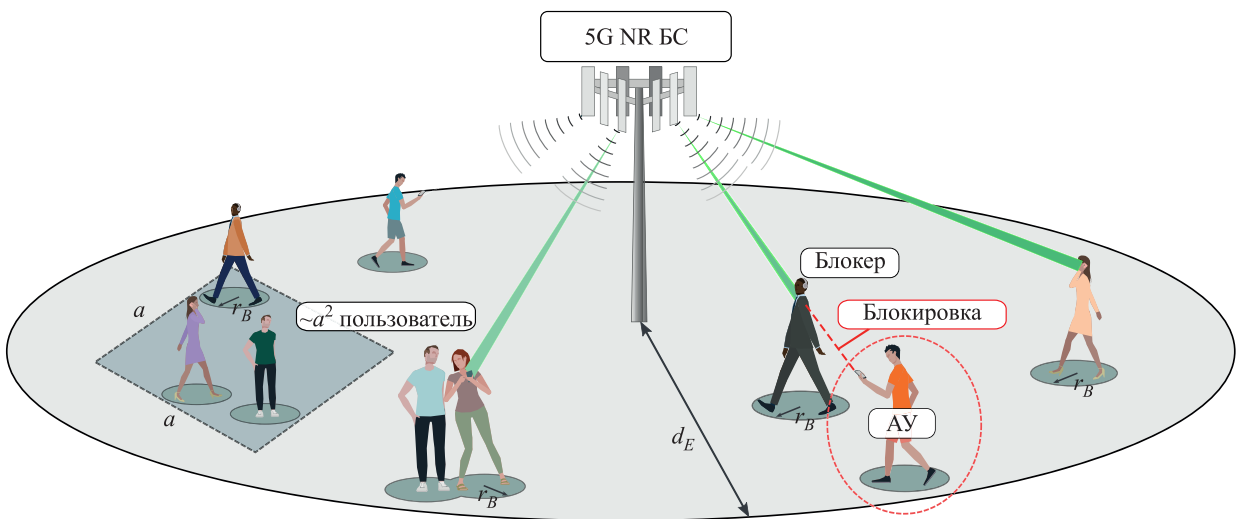


Рис. 1 Рассматриваемый сценарий развертывания

$j = 1, 2, \dots, R$ , что отражает снижение скорости передачи при неудачной попытке продолжения сессии. Далее вновь поступающие заявки будем называть первичными, а поступающие с орбиты — повторными заявками.

Каждая обслуживаемая заявка порождает пуассоновский поток сигналов интенсивностью  $\gamma$ , при поступлении которых заявка освобождает ранее занятый объем ресурса, формирует новые требования согласно тому же распределению и пытается занять новый объем. Если свободного ресурса достаточно, то продолжается ее обслуживание. В противном случае она либо уходит на орбиту с вероятностью  $\theta$ , либо покидает систему с дополнительной вероятностью  $1 - \theta$ .

Поведение системы описывается трехмерным случайным процессом  $X(t) = (\xi(t), \delta(t), \varphi(t))$ , где  $\xi(t)$  — число заявок на обслуживании;  $\delta(t)$  — общее число занятых ими единиц ресурса;  $\varphi(t)$  — число заявок на орбите. В этом случае нельзя точно определить объем высвободившегося ресурса в момент ухода заявки, но его можно аппроксимировать условным вероятностным распределением [8, 9].

Пространство состояний случайного процесса  $X(t)$  имеет вид:

$$* S = \bigcup_{n=0}^N S_n, \quad S_n = \left\{ (n, r, m) : m = 0, 1, \dots, \dots, M, \sum_{i=0}^n \sum_{j=0}^r p_{1,j}^{(i)} p_{2,r-j}^{(n-i)} > 0 \right\},$$

где распределение  $\{p_{s,j}^{(i)}\}$ ,  $j = 0, 1, \dots, R$ ,  $s = \{1, 2\}$ , есть  $i$ -кратная свертка распределения  $\{p_{s,j}\}$ ,  $j = 1, 2, \dots, R$ , а  $p_{s,j}^{(i)}$  обозначает вероятность того, что  $i$  заявок типа  $s$  суммарно занимают  $j$  единиц ресурса.

Предполагается, что требования заявок к ресурсу определяются согласно распределению  $\{p_{3,j}\}$ ,  $j = 1, 2, \dots, R$ , представляющему собой смесь распределений  $\{p_{1,j}\}$  и  $\{p_{2,j}\}$ :

$$p_{3,j} = \bar{N}_1 p_{1,j} + \bar{N}_2 p_{2,j},$$

где  $\bar{N}_1$  и  $\bar{N}_2$  — доли первичных и повторных заявок на обслуживании.

Пусть  $q(n, r, m)$  — стационарная вероятность состояния  $(n, r, m)$  процесса  $X(t)$ . Для рассматриваемой РеСМО можно вывести систему уравнений равновесия:

$$(\lambda + m\alpha)q(0, 0, m) = \mu \sum_{j:(1,j,m) \in S_1} q(1, j, m), \quad 0 \leq m \leq M; \quad (1)$$

$$\begin{aligned} & q(n, r, m) \left[ \lambda \left( \sum_{j=0}^{R-r} p_{1,j} + u(M-m)\theta \sum_{j=R-r+1}^R p_{1,j} \right) + \right. \\ & \left. + n\mu + n\gamma + m\alpha \left( \sum_{j=1}^{R-r} p_{2,j} + (1-\theta) \sum_{j=R-r+1}^R p_{2,j} \right) \right] = \\ & = \lambda \sum_{j < r, (n-1, j, m) \in S_{n-1}} q(n-1, j, m) p_{1, r-j} + \\ & + u(m)\lambda\theta q(n, r, m-1) \sum_{j=R-r+1}^R p_{1,j} + \\ & + (n+1)\mu \sum_{j > r, (n+1, j, m) \in S_{n+1}} q(n+1, j, m) \frac{p_{3, j-r} p_{3, r}^{(n)}}{p_{3, j}^{(n+1)}} + \\ & + n\gamma \sum_{(n, j, m) \in S_n} q(n, j, m) \sum_{s=\max(0, j-r)}^j \frac{p_{3, s} p_{3, j-s}^{(n-1)}}{p_{3, j}^{(n)}} p_{3, r-j+s} + \\ & + (n+1)\gamma \left( \sum_{j > r, (n+1, j, m) \in S_{n+1}} (u(m)\theta q(n+1, j, m-1) + \right. \\ & \left. + (1-\theta)q(n+1, j, m)) \frac{p_{3, j-r} p_{3, r}^{(n)}}{p_{3, j}^{(n+1)}} \sum_{s=R-r+1}^R p_{3, s} \right) + \\ & + u(M-m)(m+1)\alpha \left( (1-\theta)q(n, r, m+1) \sum_{j=R-r+1}^R p_{2, j} + \right. \\ & \left. + \sum_{j < r, (n-1, j, m+1) \in S_{n-1}} q(n-1, j, m+1) p_{2, r-j} \right), \\ & 0 < n < N, \quad 0 \leq m \leq M, \quad (n, r, m) \in S; \quad (2) \end{aligned}$$

$$\begin{aligned} & q(N, r, m) [u(M-m)\theta\lambda + N\mu + N\gamma + \\ & + m\alpha(1-\theta)] = \lambda \sum_{j < r, (N-1, j, m) \in S_{N-1}} q(N-1, j, m) p_{1, r-j} + \\ & + u(m)\lambda\theta q(N, r, m-1) + \\ & + N\gamma \sum_{(N, j, m) \in S_N} q(N, j, m) \sum_{s=\max(0, j-r)}^j \frac{p_{3, s} p_{3, j-s}^{(N-1)}}{p_{3, j}^{(N)}} p_{3, r-j+s} + \\ & + u(M-m)(m+1)\alpha \left( (1-\theta)q(N, r, m+1) + \right. \\ & \left. + \sum_{j < r, (N-1, j, m+1) \in S_{N-1}} q(N-1, j, m+1) p_{2, r-j} \right), \\ & 0 \leq m \leq M, \quad (N, r, m) \in S_N, \quad (3) \end{aligned}$$

где  $u(\cdot)$  — функция Хевисайда. Для вычисления стационарного распределения система уравнений (1)–(3) решается численно с использованием условия нормировки. При этом для весов смешанного распределения требований к ресурсу  $\{p_{3,j}\}$ ,  $j = 1, 2, \dots, R$ , предлагается использовать метод

простых итераций с начальными значениями  $\bar{N}_1 = 1g, \bar{N}_2 = 0$ .

### 3.2 Вероятностные характеристики

Время пребывания заявки в системе может включать множество периодов обслуживания и ожидания. Под периодом обслуживания понимается интервал времени, в течение которого заявка непрерывно обслуживается и по окончании либо покидает систему, либо уходит на орбиту. Аналогично период ожидания — это временной интервал, который заявка непрерывно проводит на орбите, в конце которого она либо покидает систему, либо вновь принимается на обслуживание.

Начнем с вероятности блокировки доступа на обслуживание  $\pi_N$ , которая представляет собой вероятность того, что заявка ни разу не попала на обслуживание и в конечном итоге покинула систему. Для этого заявка либо сбрасывается сразу в момент первого поступления с вероятностью  $\pi_1$ , либо сначала переходит на орбиту с вероятностью  $\pi_2$ , а в конце периода ожидания сбрасывается. Вероятности  $\pi_1$  и  $\pi_2$  имеют вид:

$$\begin{aligned} \pi_1 &= \sum_{r=1}^R \left( q(N, r, M) + (1 - \theta) \sum_{m=0}^{M-1} q(N, r, m) \right) + \\ &+ \sum_{n=1}^{N-1} \sum_{r=1}^R \left( q(n, r, M) + (1 - \theta) \sum_{m=0}^{M-1} q(n, r, m) \right) \times \\ &\quad \times \sum_{j=R-r+1}^R p_{1,j}; \\ \pi_2 &= \theta \sum_{n=1}^{N-1} \sum_{r=1}^R \sum_{m=0}^{M-1} q(n, r, m) \sum_{j=R-r+1}^R p_{1,j} + \\ &\quad + \theta \sum_{r=1}^R \sum_{m=0}^{M-1} q(N, r, m). \end{aligned}$$

Отметим, что дополнительная вероятность  $1 - \pi_1 - \pi_2$  имеет смысл вероятности того, что первичная заявка принимается на обслуживание.

При попытке попасть на обслуживание с орбиты (поступление повторной заявки) заявка может быть сброшена с вероятностью  $\pi_3$ , может вернуться на орбиту с вероятностью  $\pi_4$  либо быть принята на обслуживание с вероятностью  $1 - \pi_3 - \pi_4$ . Эти вероятности будем определять как отношения соответствующих интенсивностей. Так, средняя интенсивность поступления повторных заявок равна  $\alpha \bar{M}$ , где  $\bar{M}$  — среднее число заявок на орбите:

$$\bar{M} = \sum_{n=0}^N \sum_{r=0}^R \sum_{m=1}^M mq(n, r, m).$$

Тогда вероятность  $\pi_3$  есть отношение средней интенсивности  $\nu_3$  сброса повторных заявок к средней интенсивности их поступления:

$$\pi_3 = \frac{\nu_3}{\alpha \bar{M}},$$

где

$$\begin{aligned} \nu_3 &= (1 - \theta) \alpha \left( \sum_{n=1}^{N-1} \sum_{r=1}^R \sum_{m=1}^M mq(n, r, m) \sum_{j=R-r+1}^R p_{2,j} + \right. \\ &\quad \left. + \sum_{r=1}^R \sum_{m=1}^M mq(N, r, m) \right). \end{aligned}$$

Аналогично вероятность  $\pi_4$  равна отношению средней интенсивности  $\nu_4$  возвратов на орбиту повторных заявок к средней интенсивности их поступления. Поскольку  $\nu_4$  отличается от  $\nu_3$  только множителем, получаем

$$\pi_4 = \frac{\theta}{1 - \theta} \pi_3.$$

Тогда вероятность блокировки доступа на обслуживание  $\pi_N$  имеет вид:

$$\pi_N = \pi_1 + \pi_2 \left( \sum_{k=0}^{\infty} \pi_4^k \right) \pi_3 = \pi_1 + \frac{\pi_2 \pi_3}{1 - \pi_4}.$$

Перейдем к анализу вероятности прерывания  $\pi_O$ . По завершении периода обслуживания заявка либо прерывается с вероятностью  $\pi_5$ , либо переходит на орбиту с вероятностью  $\pi_6$ , либо успешно завершается ее обслуживание с вероятностью  $1 - \pi_5 - \pi_6$ . Вероятность  $\pi_5$  определяется как отношение средней интенсивности  $\nu_5$  прерываний заявок к средней интенсивности приема на обслуживание:

$$\pi_5 = \frac{\nu_5}{\lambda(1 - \pi_1 - \pi_2) + \alpha \bar{M}(1 - \pi_3 - \pi_4)},$$

где

$$\begin{aligned} \nu_5 &= \gamma \sum_{n=2}^N \sum_{r=1}^R n \left( q(n, r, M) + \right. \\ &\quad \left. + \sum_{m=0}^{M-1} (1 - \theta) q(n, r, m) \right) \sum_{j < r, r-j+s > R} p_{3,s} \frac{p_{3,j} p_{3,r-j}^{(n-1)}}{p_r^{(n)}}. \end{aligned}$$

Аналогично  $\pi_6$  определяется как отношение средней интенсивности  $\nu_6$  перехода заявок с обслуживания на орбиту к средней интенсивности приема на обслуживание:

$$\pi_6 = \frac{\nu_6}{\lambda(1 - \pi_1 - \pi_2) + \alpha \bar{M}(1 - \pi_3 - \pi_4)},$$

где

$$\nu_6 = \theta \gamma \sum_{n=2}^N \sum_{r=1}^R n \sum_{m=0}^{M-1} q(n, r, m) \sum_{j=1}^r \frac{p_{3,j} p_{3,r-j}^{(n-1)}}{p_r^{(n)}} \times \\ \times \sum_{s=R-r+j+1}^R p_{3,s}.$$

Период ожидания заявки завершается либо прерыванием с вероятностью  $\pi_7$ , либо возвратом на обслуживание с вероятностью  $1 - \pi_7$ . Тогда

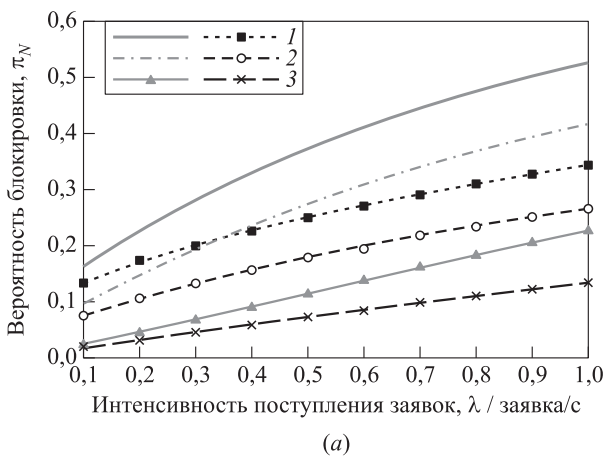
$$\pi_7 = \sum_{k=0}^{\infty} \pi_4^k \pi_3 = \frac{\pi_3}{1 - \pi_4}.$$

Заявка прерывается, если она получила хотя бы один период обслуживания (с вероятностью  $1 - \pi_N$ ), затем получает  $k \geq 0$  периодов ожидания и обслуживания, а затем либо прерывается в конце последнего периода обслуживания с вероятностью  $\pi_5$ , либо уходит на орбиту и прерывается в конце периода ожидания с вероятностью  $\pi_6 \pi_7$ . Тогда

$$\pi_O = (1 - \pi_N) \sum_{k=0}^{\infty} (\pi_6(1 - \pi_7))^k (\pi_5 + \pi_6 \pi_7) = \\ = (1 - \pi_N) \frac{\pi_5 + \pi_6 \pi_7}{1 - \pi_6(1 - \pi_7)}.$$

Средняя доля занятых ресурсов вычисляется по формуле:

$$U = \frac{1}{R} \sum_{n=1}^N \sum_{r=1}^R \sum_{m=0}^M r q(n, r, m).$$



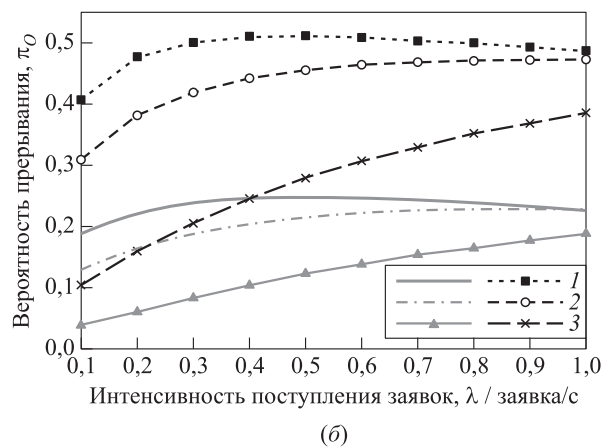
Важным показателем для систем 5G/6G с ненадежным обслуживанием служит доля ресурсов  $U_W$ , предоставленных сессиям, которые в конечном итоге были прерваны:

$$U_W = \frac{\pi_O}{1 - \pi_N} U.$$

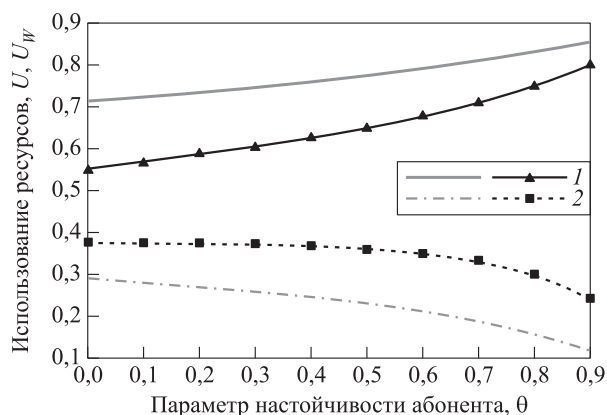
## 4 Численные результаты

В качестве численного исследования рассмотрим основные характеристики обслуживания абонентов — вероятность блокировки доступа на обслуживание и вероятность прерывания заявки, представленные на рис. 2, где среднее время пребывания на орбите составляет  $1/\alpha = 1$  с. Методы расчета интенсивности сигналов и распределения требований первичных и повторных заявок к ресурсу представлены в [7]. Анализируя вероятность блокировки доступа на обслуживание, отметим, что она растет по мере увеличения интенсивности поступления заявок. Отметим, что увеличение плотности блокеров приводит к уменьшению вероятности блокировки доступа на обслуживание для всех значений коэффициента настойчивости пользователей. Примечательно, что настойчивость пользователя позволяет значительно уменьшить исследуемую вероятность. В частности, при  $\lambda_B = 0,01$  бл./м<sup>2</sup> и  $\theta = 0$  вероятность блокировки доступа на обслуживание равна 0,17 и уменьшается до 0,1 при совершении двух повторных попыток в среднем.

Рассмотрим теперь вероятность прерывания сессий (см. рис. 2, б). Здесь наблюдается резкое



**Рис. 2** Основные характеристики обслуживания абонентов: (а) вероятность блокировки доступа на обслуживание; (б) вероятность прерывания заявки: серые кривые —  $\lambda_B = 0,01$  бл./м<sup>2</sup>; черные кривые —  $\lambda_B = 0,1$  бл./м<sup>2</sup>; 1 —  $\theta = 0$ ; 2 — 0,5; 3 —  $\theta = 0,9$



**Рис. 3** Коэффициент использования ресурса системы: серые кривые —  $\lambda_B = 0,01$  бл./м<sup>2</sup>; черные кривые —  $\lambda_B = 0,1$  бл./м<sup>2</sup>; 1 — использование ресурсов системы; 2 — доля потраченного впустую ресурса

различие между кривыми, соответствующими  $\lambda_B = 0,1$  и  $0,01$  бл./м<sup>2</sup>, достигающее 0,2–0,3. Увеличение плотности блокеров приводит к росту исследуемой вероятности и снижению вероятности блокировки доступа на обслуживание (см. рис. 2, а). В целом при рассматриваемых значениях интенсивности поступления большая часть заявок, принятых к обслуживанию, в итоге теряется. Настойчивость пользователя значительно повышает вероятность успешного обслуживания. В частности, при  $\lambda = 0,1$  з./с,  $\lambda_B = 0,1$  бл./м<sup>2</sup> настойчивость снижает вероятность прерывания с 0,4 до 0,1.

Наконец, рассмотрим коэффициент использования ресурса системы и долю ресурса, потраченного системой впустую, представленные на рис. 3. Анализируя представленные данные, можно заметить, что настойчивость пользователей положительно влияет на обе метрики, позволяя использовать больше системного ресурса и делать это более эффективно. В частности, при  $\lambda = 0,1$  бл./м<sup>2</sup> и  $\theta = 0$  около 40% ресурса из занятых 55% тратится впустую. При  $\theta = 0,9$  система не только использует 80% ресурса, но и только 15% из этого числа тратится на заявки, которые не будут успешно обслужены.

## 5 Заключение

В статье предложена модель процесса обслуживания пользователей на БС с учетом ненадежного характера миллиметровых и субтерагерцевых сред передачи систем 5G/6G, которые подвержены прерываниям соединения, вызванным блокировкой, и настойчивому поведению абонен-

тов. Предложенная модель описана в виде РЕСМО с орбитой и позволяет получить основные характеристики обслуживания абонентов и производительности системы — вероятности блокировки доступа на обслуживание и прерывания, а также коэффициент использования ресурса системы.

## Литература

1. Begishev V., Sopin E., Moltchanov D., Kovalchukov R., Samuylov A., Andreev S., Koucheryavy Y., Samouylov K. Joint use of guard capacity and multiconnectivity for improved session continuity in millimeter-wave 5G NR systems // IEEE T. Veh. Technol., 2021. Vol. 70. Iss. 3. P. 2657–2672. doi: 10.1109/TVT.2021.3061906.
2. Власкина А. С., Бурцева С. А., Кочеткова И. А., Шоргин С. Я. Управляемая система массового обслуживания с эластичным трафиком и сигналами для анализа нарезки ресурсов в сети радиодоступа // Информатика и её применения, 2022. Т. 16. Вып. 3. С. 90–96. doi: 10.14357/19922264220312.
3. Kovalchukov R., Moltchanov D., Gaidamaka Y., Bobrikova E. An accurate approximation of resource request distributions in millimeter wave 3GPP new radio systems // Internet of things, smart spaces, and next generation networks and systems / Eds. O. Galinina, S. Andreev, S. Balandin, Y. Koucheryavy. — Lecture notes in computer science ser. — Cham: Springer, 2019. Vol. 11660. P. 572–585. doi: 10.1007/978-3-030-30859-9\_50.
4. Nain P., Towsley D., Liu B., Liu Z. Properties of random direction models // 24th Annual Joint Conference of the IEEE Computer and Communications Societies Proceedings. — Piscataway, NJ, USA: IEEE, 2005. Vol. 3. P. 1897–1907. doi: 10.1109/INFCOM.2005.1498468.
5. Petrov V., Komarov M., Moltchanov D., Jornet J., Koucheryavy Y. Interference and SINR in millimeter wave and terahertz communication systems with blocking and directional antennas // IEEE T. Wirel. Commun., 2017. Vol. 16. Iss. 3. P. 1791–1808. doi: 10.1109/TWC.2017.2654351.
6. Gapeyenko M., Samuylov A., Gerasimenko M., Moltchanov D., Singh S., Akdeniz M., Aryafar E., Himayat N., Andreev S., Koucheryavy E. On the temporal effects of mobile blockers in urban millimeter-wave cellular scenarios // IEEE T. Veh. Technol., 2017. Vol. 66. Iss. 11. P. 10124–10138. doi: 10.1109/TVT.2017.2754543.
7. Бегишев В. О., Сопин Э. С., Молчанов Д. А., Самуйлов А. К., Гайдамака Ю. В., Самуйлов К. Е. Оценка эффективности механизма резервирования полосы пропускания для технологии mmWave в сетях связи пятого поколения // Информационно-управляющие системы, 2019. № 5(102). С. 51–63. doi: 10.31799/1684-8853-2019-5-51-63.

8. Горбунова А. В., Наумов В. А., Гайдамака Ю. В., Самуйлов К. Е. Ресурсные системы массового обслуживания как модели беспроводных систем связи // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 48–55. doi: 10.14357/19922264180307.
9. Moltchanov D., Sopin E., Begishev V., Samuylov A., Koucheryavy Y., Samouylov K. A tutorial on mathematical modeling of 5G/6G millimeter wave and terahertz cellular systems // IEEE Commun. Surv. Tut., 2022. Vol. 24. Iss. 2. P. 1072–1116. doi: 10.1109/COMST.2022.3156207.

Поступила в редакцию 08.02.23

## MODELING INSISTENT USER BEHAVIOR IN 5G NEW RADIO NETWORKS WITH RATE ADAPTATION AND BLOCKAGE

E. S. Sopin<sup>1,2</sup>, A. R. Maslov<sup>1</sup>, V. S. Shorgin<sup>2</sup>, and V. O. Begishev<sup>1</sup>

<sup>1</sup>RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

<sup>2</sup>Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** Millimeter wave 5G New Radio access technology and future 6G terahertz systems are designed for rate-demanding multimedia applications. Such applications are adaptive to the network transmission rate. The unreliable nature of 5G/6G communications networks, subject to loss of connection due to dynamic blockage of radio propagation paths, can cause impatient user behavior leading to retries to continue service. For such systems, the authors propose a service model with impatient behavior and service interruption based on a resource queuing system with an orbit. The authors derive the new and ongoing session loss probabilities as well as characterize the efficiency of resource usage. Numerical results show that user persistence reduces the probabilities under consideration: performing an average of two retries reduces both indicators by 20%–70% while with an average of 10 attempts, the gain can reach 200%–300%. Persistence not only increases the use of system resources by 20%–40% but also allows them to be used efficiently reducing the percentage of wasted resources by two to three times.

**Keywords:** 5G; New Radio; resource queuing system; callbacks; propagation blocking; service interruption

**DOI:** 10.14357/19922264230304

**EDN:** ENSHKV

## Acknowledgments

The research was funded by the Russian Science Foundation, project No. 22-79-10128.

## References

- Begishev, V., E. Sopin, D. Moltchanov, R. Kovalchukov, A. Samuylov, S. Andreev, Y. Koucheryavy, and K. Samouylov. 2021. Joint use of guard capacity and multiconnectivity for improved session continuity in millimeter-wave 5G NR systems. *IEEE T. Veh. Technol.* 70(3): 2657–2672. doi: 10.1109/TVT.2021.3061906.
- Vlaskina, A. S., S. A. Burtseva, I. A. Kochetkova, and S. Ya. Shorgin. 2022. Upravlyаемая система массового обслуживания с эластичным трафиком и сигналами для анализа нarezки ресурсов в сети радиодоступа [Controllable queuing system with elastic traffic and signals for analyzing network slicing]. *Informatika i ee Primeneniya — Inform. Appl.* 16(3): 90–96. doi: 10.14357/19922264220312.
- Kovalchukov, R., D. Moltchanov, Y. Gaidamaka, and E. Bobrikova. 2019. An accurate approximation of resource request distributions in millimeter wave 3GPP new radio systems. *Internet of things, smart spaces, and next generation networks and systems*. Eds. O. Galinina, S. Andreev, S. Balandin, and Y. Koucheryavy. Lecture notes in computer science ser. Cham: Springer. 11660:572–585. doi: 10.1007/978-3-030-30859-9\_50.
- Nain, P., D. Towsley, B. Liu, and Z. Liu. 2005. Properties of random direction models. *24th Annual Joint Conference of the IEEE Computer and Communications Societies Proceedings*. Piscataway, NJ: IEEE. 3:1897–1907. doi: 10.1109/INFCOM.2005.1498468.
- Petrov, V., M. Komarov, D. Moltchanov, J. Jornet, and Y. Koucheryavy. 2017. Interference and SINR in millimeter wave and terahertz communication systems with blocking and directional antennas. *IEEE T. Wirel. Commun.* 16:1791–1808. doi: 10.1109/TWC.2017.2654351.
- Gapeyenko, M., A. Samuylov, M. Gerasimenko, D. Moltchanov, S. Singh, M. Riza, E. Aryafar, N. Himayat, S. Andreev, and Y. Koucheryavy. 2017. On the temporal effects of mobile blockers in urban millimeter-wave cellular scenarios. *IEEE T. Veh. Technol.* 66(11):10124–10138. doi: 10.1109/TVT.2017.2754543.
- Begishev, V. O., E. S. Sopin, D. A. Moltchanov, A. K. Samuylov, Yu. V. Gaydamaka, and K. E. Samouylov. 2019. Otsenka effektivnosti mekhanizma rezervirovaniya po-

- losy propuskaniya dlya tekhnologii mmWave v setyakh svyazi pyatogo pokoleniya [Performance evaluation of bandwidth reservation for mmWave in 5G NR systems]. *Informatsionno-upravlyayushchie sistemy* [Information and Control Systems] 5(102):51–63. doi: 10.31799/1684-8853-2019-5-51-63.
8. Gorbunova, A. V., V. A. Naumov, Yu. V. Gaydamaka, and K. E. Samuylov. 2018. Resursnye sistemy massovogo obsluzhivaniya kak modeli besprovodnykh sistem svyazi [Resource queuing systems as models of wireless communication systems]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):48–55. doi: 10.14357/19922264180307.
9. Moltchanov, D., E. Sopin, V. Begishev, A. Samuylov, Y. Koucheryavy, and K. Samouylov. 2022. A tutorial on mathematical modeling of 5G/6G millimeter wave and terahertz cellular systems. *IEEE Commun. Surv. Tut.* 24(2):1072–1116. doi: 10.1109/COMST.2022.3156207.

*Received February 8, 2023*

## Contributors

**Sopin Eduard S.** (b. 1987) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Applied Probability and Informatics, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sopin-es@rudn.ru

**Maslov Alexander R.** (b. 1998) — PhD student, Department of Applied Probability and Informatics, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; maslov-ar@rudn.ru

**Shorgin Vsevolod S.** (b. 1978) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vshorgin@ipiran.ru

**Begishev Vyacheslav O.** (b. 1988) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Applied Probability and Informatics, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; begishevu-vo@rudn.ru



# АНАЛИЗ ЗАГРУЗКИ МНОГОПОЛЬЗОВАТЕЛЬСКОЙ СЕТИ ПРИ РАСЩЕПЛЕНИИ ПОТОКОВ ПО КРАТЧАЙШИМ МАРШРУТАМ

Ю. Е. Малащенко<sup>1</sup>, И. А. Назарова<sup>2</sup>

**Аннотация:** В рамках вычислительных экспериментов на многопродуктовой сетевой модели изучаются два способа передачи потоков различных видов по кратчайшим маршрутам. В первом случае передаваемые межузловые потоки равны по величине. В другом — определяется недискриминирующее распределение, при котором всем парам корреспондентов выделяются одинаковые ресурсы. Суммарная нагрузка ребер сети, возникающая при одновременной передаче всех межузловых информационных потоков, считается заданной. Предложенный метод позволяет получить гарантированные оценки удельных затрат ресурсов сети и предельно допустимых нагрузок ребер при одновременной передаче расщепленных межузловых потоков по всем найденным кратчайшим маршрутам. Приводятся результаты сравнительного анализа уравнивающего распределения потоков и ресурсов в сетях с различными структурными особенностями. Алгоритмическая схема имеет полиномиальную оценку требуемого числа операций.

**Ключевые слова:** многопродуктовая потоковая модель; распределение межузловых потоков и нагрузок; предельная нагрузка сети

DOI: 10.14357/19922264230305

EDN: NLUSQJ

## 1 Введение

Данная работа продолжает исследование проблем управления потоками в телекоммуникационных территориально распределенных системах связи [1–3]. В рамках вычислительных экспериментов на многопродуктовой сетевой модели анализируются результирующие загрузки ребер при уравнительном способе распределения потоков. Под межузловой нагрузкой понимается суммарная величина пропускных способностей, выделяемых в сети для обеспечения передачи информационного потока определенного вида. Сумма всех дуговых потоков разных видов, проходящих по некоторому ребру, трактуется как *реберная нагрузка*. Общая допустимая нагрузка ребер сети, возникающая при одновременной передаче межузловых информационных потоков для всех пар корреспондентов, считается заданной.

Анализируются две стратегии управления и диспетчерские правила маршрутизации. В первом случае осуществляется распределение межузловых потоков, равных друг другу, во втором — определяется недискриминирующее распределение, при котором всем парам корреспондентов выделяются одинаковые ресурсы. При передаче каждого вида потока по всем кратчайшим маршрутам проводится оценка удельных затрат. Анализ результатов экспериментов позволяет оценить распределение

*реберных нагрузок* для описанных стратегий управления в многопользовательских сетях с различными структурными особенностями.

При создании, развитии и эксплуатации телекоммуникационных систем в настоящее время используются потоковые модели [4]. В частности, указанные модели применяются для поиска стратегий управления и построения диспетчерских правил распределения потоков, нагрузок и ресурсов в многопользовательских сетях [5, 6]. В русле исследований по разработке SPLIT-методов [7–10] лежат представленные в разд. 2–4 алгоритмические схемы выравнивания нагрузок и получения уравнительных распределений межузловых потоков по различным маршрутам. В данной работе используются процедуры поиска всех кратчайших путей с полиномиальными оценками вычислительных затрат [11].

## 2 Математическая модель

Для описания многопользовательской сетевой системы связи воспользуемся следующей математической записью модели передачи многопродуктового потока. Сеть  $G$  задается множествами  $\langle V, R, U, P \rangle$ : узлов (вершин) сети  $V = \{v_1, v_2, \dots, v_n, \dots, v_N\}$ ; неориентированных ребер  $R = \{r_1, r_2, \dots, r_k, \dots, r_E\}$ ; ориентированных

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, malash09@ccas.ru

<sup>2</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, irina-nazar@yandex.ru

дуг  $U = \{u_1, u_2, \dots, u_k, \dots, u_{2E}\}$ ; пар узлов-корреспондентов  $P = \{p_1, p_2, \dots, p_M\}$ . Предполагается, что в сети отсутствуют петли и сдвоенные ребра. В многопользовательской сети  $G$  рассматривается  $M = N(N - 1)$  независимых, невзаимозаменяемых и равноправных межузловых потоков различных видов.

Ребро  $r_k \in R$  соединяет смежные вершины  $v_{n_k}$  и  $v_{j_k}$ . Каждому ребру  $r_k$  ставятся в соответствие две ориентированные дуги  $u_k$  и  $u_{k+E}$  из множества  $U$ . Дуги  $\{u_k, u_{k+E}\}$  определяют прямое и обратное направление передачи потока по ребру  $r_k$  между концевыми вершинами  $v_{n_k}$  и  $v_{j_k}$ .

Каждой паре узлов-корреспондентов  $p_m$  из множества  $P$  ставятся в соответствие: вершина-источник с номером  $s_m$ , из  $s_m$  входной поток  $m$ -го вида поступает в сеть; вершина-приемник с номером  $t_m$ , из  $t_m$  поток  $m$ -го вида покидает сеть. Обозначим через  $z_m$  величину межузлового потока  $m$ -го вида, поступающего в сеть через узел с номером  $s_m$  и покидающего сеть из узла с номером  $t_m$ ;  $x_{mk}$  и  $x_{m(k+E)}$  — поток  $m$ -го вида, который передается по дугам  $u_k$  и  $u_{k+E}$  согласно направлению передачи,  $x_{mk} \geq 0$ ,  $x_{m(k+E)} \geq 0$ ,  $m = \overline{1, M}$ ,  $k = \overline{1, E}$ ;  $S(v_n)$  — множество номеров исходящих дуг, по ним поток покидает узел  $v_n$ ;  $T(v_n)$  — множество номеров входящих дуг, по ним поток поступает в узел  $v_n$ . Состав множеств  $S(v_n)$  и  $T(v_n)$  однозначно формируется в ходе выполнения следующей процедуры. Пусть некоторое ребро  $r_k \in R$  соединяет вершины с номерами  $n$  и  $j$ , такими что  $n < j$ . Тогда ориентированная дуга  $u_k = (v_n, v_j)$ , направленная из вершины  $v_n$  в  $v_j$ , считается *исходящей* из вершины  $v_n$  и ее номер  $k$  заносится в множество  $S(v_n)$ , а дуга  $u_{k+E}$ , направленная из  $v_j$  в  $v_n$ , — *входящей* для  $v_n$  и ее номер  $k + E$  помещается в список  $T(v_n)$ . Дуга  $u_k$  является *входящей* для  $v_j$ , и ее номер  $k$  попадает в  $T(v_j)$ , а дуга  $u_{k+E}$  — *исходящей*, и номер  $k + E$  вносится в список исходящих дуг  $S(v_j)$ .

Во всех узлах сети  $v_n \in V$ ,  $n = \overline{1, N}$ , для каждого вида потока должны выполняться условия сохранения потоков:

$$\sum_{i \in S(v_n)} x_{mi} - \sum_{i \in T(v_n)} x_{mi} = \begin{cases} z_m, & \text{если } v_n = v_{s_m}; \\ -z_m, & \text{если } v_n = v_{t_m}; \\ 0 & \text{в остальных случаях,} \end{cases} \\ n = \overline{1, N}, \quad m = \overline{1, M}, \quad x_{mi} \geq 0, \quad z_m \geq 0. \quad (1)$$

Величина  $z_m$  равна входному межузловому потоку  $m$ -го вида, проходящему от источника к приемнику пары  $p_m$  при распределении потоков  $x_{mi}$  по дугам сети.

Каждому ребру  $r_k \in R$  приписывается неотрицательное число  $d_k$ , определяющее суммарный предельно допустимый поток, который можно передать по ребру  $r_k$  в обоих направлениях. Компоненты вектора пропускных способностей  $\mathbf{d} = (d_1, d_2, \dots, d_k, \dots, d_E)$  — положительные числа  $d_k > 0$ . Вектор  $\mathbf{d}$  задает следующие ограничения на сумму потоков всех видов, передаваемых по ребру  $r_k$  одновременно:

$$\sum_{m=1}^M (x_{mk} + x_{m(k+E)}) \leq d_k, \\ x_{mk} \geq 0, \quad x_{m(k+E)} \geq 0, \quad k = \overline{1, E}. \quad (2)$$

Ограничения (1) и (2) задают выпуклое многогранное множество допустимых значений компонент вектора межузловых потоков  $\mathbf{z} = (z_1, z_2, \dots, z_m, \dots, z_M)$ :

$$\mathcal{Z}(\mathbf{d}) = \{ \mathbf{z} \geq 0 \mid \exists \mathbf{x} \geq 0 : (\mathbf{z}, \mathbf{x}) \text{ удовлетворяют (1), (2)} \}.$$

В рамках данной модели пропускная способность ребер сети измеряется в условных единицах потока и трактуется как *ресурсное* ограничение. Суммарное значение пропускной способности сети  $D(0) = \sum_{k=1}^E d_k$  считается заданным. Для каждой пары узлов-корреспондентов  $p_m \in P$ , для некоторого допустимого межузлового потока  $\tilde{z}_m$  и соответствующих дуговых потоков  $\tilde{x}_{mk}$ ,  $k = \overline{1, 2E}$ , величина

$$\tilde{y}_m = \sum_{i=1}^{2E} \tilde{x}_{mi}, \quad m = \overline{1, M},$$

характеризует *нагрузку* на сеть при передаче межузлового потока величиной  $\tilde{z}_m$  из узла-источника  $s_m$  в узел-приемник  $t_m$ . Величина  $\tilde{y}_m$  показывает, какая суммарная пропускная способность сети потребуется для передачи дуговых потоков  $\tilde{x}_{mk}$ . В рамках модели отношение реберных и межузловых потоков  $\tilde{w}_m = \tilde{y}_m / \tilde{z}_m$ ,  $m = \overline{1, M}$ , можно трактовать как удельные *затраты* ресурсов сети при передаче единичного потока  $m$ -го вида между узлами  $s_m$  и  $t_m$  при дуговых потоках  $\tilde{x}_{mi}$ . Величины  $\tilde{z}_m = \tilde{z}_m / \tilde{y}_m$  и  $\tilde{x}_{mi} = \tilde{x}_{mi} / \tilde{y}_m$ ,  $m = \overline{1, M}$ ,  $i = \overline{1, E}$ , соответствуют межузловому потоку при единичной нагрузке для пары  $p_m$ .

Вводится величина  $\tilde{\Delta}_k$ , характеризующая реберную суммарную *загрузку* (РС-загрузку)  $k$ -го ребра при одновременной передаче всех межузловых потоков  $\tilde{z}_m$ :

$$\tilde{\Delta}_k = \sum_{m=1}^M (\tilde{x}_{mk} + \tilde{x}_{m(k+E)}), \quad k = \overline{1, E}.$$

### 3 Распределение межузловых потоков по кратчайшим маршрутам

Для оценки минимальных удельных затрат при различных уравнивающих стратегиях управления использовалась SFSR-процедура (*англ.* split flow shortest route — расщепление потока по всем кратчайшим маршрутам), которая позволяет распределять межузловые потоки по маршрутам, состоящим из минимального числа ребер.

При проведении эксперимента на первом этапе для каждой пары узлов  $p_m = (s_m, t_m)$  в сети  $G(0)$  определяется множество всех кратчайших путей, которые далее используются как маршруты передачи  $m$ -го вида потока

$$H^0(m) = \{h^1(m), h^2(m), \dots, h^j(m), \dots, h^{J(m)}(m)\}.$$

Здесь  $h^j(m) = \{k_1^j(m), k_2^j(m), \dots, k_{\mu(m)}^j(m)\}$  — список номеров ребер в  $j$ -м кратчайшем пути между узлами  $s_m$  и  $t_m$ , где  $\mu(m)$  — число ребер в кратчайшем маршруте  $h^j(m)$ ;  $J(m)$  — число кратчайших маршрутов для  $m$ -й пары.

Для оценки возможности «расщепления» потока по различным маршрутам на первом этапе для каждой пары  $p_m \in P$  по каждому маршруту  $h^j(m)$  из  $H^0(m)$  передается межузловой поток  $z_m^j = 1$ . Подсчитывается величина  $z_m^0$ , численно равная сумме единичных потоков, которые одновременно передаются по всем кратчайшим маршрутам из узла  $s_m$  в узел  $t_m$ :

$$z_m^0 = \sum_{j=1}^{J(m)} z_m^j = J(m).$$

Вычисляется нормирующий коэффициент

$$\xi_m^0 = \frac{1}{z_m^0}, \quad z_m^0 \neq 0, \quad m = \overline{1, M},$$

значения индикаторной функции

$$\eta_k^j(m) = \begin{cases} 1, & \text{если } r_k \in R, \quad k \in h^j(m); \\ 0 & \text{в остальных случаях} \end{cases}$$

и подсчитываются реберные загрузки

$$\Delta_k^0(m) = \xi_m^0 \sum_{j=1}^{J(m)} \eta_k^j(m), \quad m = \overline{1, M}, \quad k = \overline{1, E},$$

которые возникнут при передаче единичного межузлового потока из узла  $s_m$  в узел  $t_m$  одновременно по всем кратчайшим маршрутам из  $H^0(m)$ .

В ходе вычислительных экспериментов предполагалось, что суммарная загрузка всех ребер сети не

может превышать  $D(0)$ . Для заданного  $D(0)$  подсчитываются предельно допустимые РС-загрузки:

$$\alpha^{**}(0) \sum_{m=1}^M \Delta_k^0(m) = \Delta_k^*(z), \quad \sum_{k=1}^E \Delta_k^*(z) = D(0).$$

Реберные суммарные загрузки  $\Delta_k^*(z)$  возникают в сети при одновременной передаче межузловых потоков  $z_m^{**} = \alpha^{**}(0)$  для всех пар  $p_m \in P$  и суммарного межузлового потока  $\sum_{m=1}^M z_m^{**} = M\alpha^{**}(0)$ .

Для оценки РС-загрузок при *уравнительном* распределении результирующих межузловых нагрузок для всех пар узлов  $p_m \in P$  также используется SFSR-процедура. На первом этапе определяются реберные загрузки для  $m = \overline{1, M}$  и  $k = \overline{1, E}$  при передаче единичных потоков по всем кратчайшим маршрутам из  $H^0(m)$ ,  $m = \overline{1, M}$ . Для каждой пары  $p_m \in P$  вычисляются загрузки

$$y_m^0 = \sum_{k=1}^E \sum_{j=1}^{J(m)} \eta_k^j(m) = \mu(m)J(m), \quad m = \overline{1, M},$$

нормирующие коэффициенты

$$\theta_m^0 = \frac{1}{y_m^0}, \quad y_m^0 \neq 0, \quad m = \overline{1, M},$$

и загрузки

$$\Delta_k^0(m) = \theta_m^0 \sum_{j=1}^{J(m)} \eta_k^j(m), \quad m = \overline{1, M}, \quad k = \overline{1, E},$$

при которых по всем маршрутам  $H^0(m)$ ,  $m = \overline{1, M}$ , результирующие межузловые нагрузки  $y_m^1 = \theta_m^0 y_m^0$  будут равны единице. Определяются РС-загрузки

$$\beta^{**} \sum_{j=1}^M \Delta_k^0(m) = \Delta_k^*(y), \quad \sum_{k=1}^E \Delta_k^*(y) = D(0)$$

и межузловые потоки

$$z_m(y^{**}) = \beta^{**} \theta_m^0 z_m^0, \quad m = \overline{1, M}.$$

### 4 Вычислительный эксперимент

Вычислительный эксперимент проводился на моделях сетевых систем, представленных на рис. 1. В каждой сети 69 узлов. В ходе вычислительного эксперимента проводилась нормировка, и суммарная пропускная способность в обеих сетях была одинакова:

$$\sum_{k=1}^E \Delta_k^*(\cdot) = D(0) = 68\,256.$$

Результаты вычислительных экспериментов с использованием SFSR-процедуры представлены

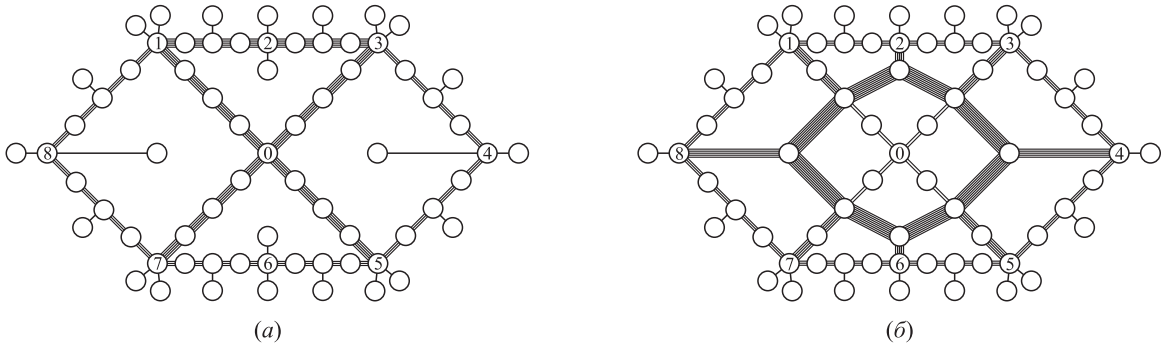


Рис. 1 Результирующие реберные загрузки в базовой (а) и в кольцевой сети (б)

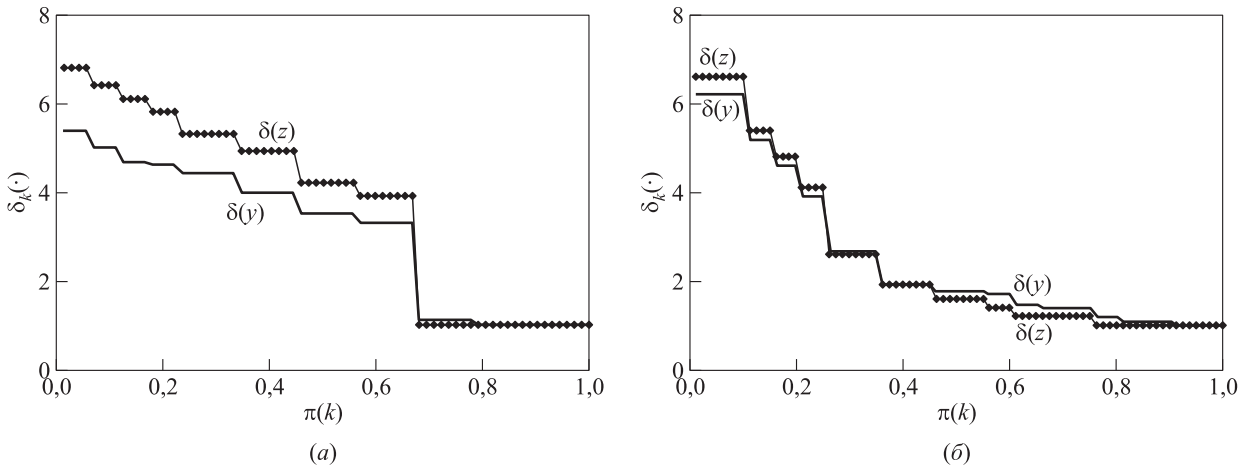


Рис. 2 Распределение реберных загрузок в базовой (а) и в кольцевой сети (б)

Значения потоков и загрузок в базовой и кольцевой сети (SFSR-стратегия)

Сеть	Медиана потоков	Сумма межузловых потоков	Удельные затраты	Норма вектора РС-загрузок
Базовая				
$z^{**}$	1,85	8 670	7,9	9210
$z(y^{**})$	1,82	12 810	5,3	9000
Кольцевая				
$z^{**}$	2,5	11 750	5,8	9520
$z(y^{**})$	2,4	15 650	4,4	9160

на рис. 1 и 2. В таблице и на рис. 1 значения потоков и загрузок указаны в единицах потока, а на рис. 2 — в относительных единицах. По завершении экспериментов на основе полученных значений  $\Delta_k^*(z)$  (загрузки сети при уравнительном распределении межузловых потоков) и  $\Delta_k^*(y)$  (загрузок при уравнительном распределении межузловых нагрузок) вычисляются

$$\Delta_{\min}(z) = \min_k \{\Delta_k^*(z)\}, \quad \Delta_{\min}(y) = \min_k \{\Delta_k^*(y)\},$$

$$k = \overline{1, E};$$

$$\delta_k(z) = \frac{\Delta_k^*(z)}{\Delta_{\min}(z)}, \quad \delta_k(y) = \frac{\Delta_k^*(y)}{\Delta_{\min}(y)},$$

$$k = \overline{1, E}, \quad \Delta_{\min}(\cdot) \neq 0.$$

Результирующие диаграммы представлены на рис. 2. На диаграммах по вертикальной оси откладываются значения  $\delta_k(\cdot)$  (упорядоченные по невозрастанию), а по горизонтальной указываются порядковые номера ребер в данной упорядоченной последовательности  $\pi(k) = k/E$  для  $k = \overline{1, E}$ , где  $E = 72$  — число ребер для базовой сети (см. рис. 2, а)

и  $E = 80$  — для кольцевой (см. рис. 2, б). Значения  $\delta_k(\cdot) = 1$  в правой части диаграммы относятся к ребрам, инцидентным висячим узлам, а в левой части диаграммы — к ребрам с максимальной загрузкой.

Схематичные изображения сетей на рис. 1 соответствуют распределению относительных загрузок  $\delta_k(\cdot)$  на ребрах сети. В базовой сети (см. рис. 1, а) наибольшая нагрузка на кольцо и ребрах, исходящих из центра сети. В кольцевой сети (см. рис. 1, б) максимальная нагрузка приходится на дополнительные ребра внутреннего кольца и на ребра, соединяющие внешнее кольцо с внутренним. В обеих сетях единичная относительная нагрузка на висячих ребрах соответствует собственным информационным потокам при отсутствии транзитных.

Результаты, собранные в таблице, свидетельствуют о том, что как суммарное, так и медианное значение межузловых потоков значительно больше в кольцевой сети при равной суммарной допустимой нагрузке в обеих сетях. Удельные затраты на передачу потоков при равных межузловых нагрузках достигают минимального значения в кольцевой сети. Медианные значения реберной загрузки в базовой сети на 50% больше, чем в кольцевой.

## 5 Заключение

В работе анализируются распределения реберных загрузок и проводится оценка удельных затрат ресурсов при одновременной передаче межузловых потоков разных видов и уравнильном способе управления. Полученные значения медиан распределений межузловых потоков можно интерпретировать как представительные оценки функциональных возможностей сети в условиях неопределенности, невзаимозаменяемости потоков и равноправности корреспондентов. Все межузловые потоки, не превышающие медианных значений, можно передать в сети одновременно. В кольцевой сети удельные затраты на передачу потоков и медианные значения реберных загрузок значительно меньше, чем в базовой. Средние показатели удельных затрат позволяют оценивать эффективность использования ресурсов при изменении структуры сети.

Предложенный метод расщепления потоков по всем кратчайшим маршрутам дает возможность анализировать загрузку сети с учетом расщепления каждого вида потока. Полученные агрегированные показатели для узлов сети могут быть использованы при построении маршрутно-адресных таблиц, подготовке графиков обходов и замен для работы

в аварийных режимах и при критически опасных повреждениях.

Описанный метод позволяет проводить предварительную оценку проекта сети, построенной на базе арендованных каналов связи при сохранении их общего числа. Вычислительные эксперименты показали, что изменение структуры и загрузки ребер при переходе от базовой к кольцевой позволяет значительно увеличить межузловые потоки, хотя общее число арендуемых каналов совпадает в обеих сетях.

## Литература

1. Малащенко Ю. Е., Назарова И. А. Оценка предельных распределений пропускной способности в многопользовательской сети при передаче межузловых потоков по кратчайшим маршрутам // Известия РАН. Теория и системы управления, 2022. № 5. С. 79–89.
2. Малащенко Ю. Е., Назарова И. А. Анализ распределения нагрузки и межузловых потоков при различных стратегиях маршрутизации в многопользовательской сети // Известия РАН. Теория и системы управления, 2022. № 6. С. 112–122.
3. Малащенко Ю. Е., Назарова И. А. Сравнительный анализ стратегий управления потоками в многопользовательской телекоммуникационной сети // Известия РАН. Теория и системы управления, 2023. № 2. С. 164–176.
4. Salimifard K., Bigharaz S. The multicommodity network flow problem: State of the art classification, applications, and solution methods // Operational Research, 2020. Vol. 22. Iss. 1. P. 1–47. doi: 10.1007/s12351-020-00564-8.
5. Luss H. Equitable resource allocation: Models, algorithms, and applications. — Hoboken, NJ, USA: John Wiley & Sons, 2012. 376 p.
6. Balakrishnan A., Li G., Mirchandani P. Optimal network design with end-to-end service requirements // Oper. Res., 2017. Vol. 65. Iss. 3. P. 729–750. doi: 10.1287/opre.2016.1579.
7. Caramia M., Sgalambro A. A fast heuristic algorithm for the maximum concurrent k-splittable flow problem // Optim. Lett., 2010. Vol. 4. Iss. 1. P. 37–55. doi: 10.1007/s11590-009-0147-4.
8. Kabadurmus O., Smith A. E. Multicommodity k-splittable survivable network design problems with relays // Telecommun. Syst., 2016. Vol. 62. Iss. 1. P. 123–133. doi: 10.1007/s11235-015-0067-9.
9. Bialon P. A randomized rounding approach to a k-splittable multicommodity flow problem with lower path flow bounds affording solution quality guarantees // Telecommun. Syst., 2017. Vol. 64. Iss. 3. P. 525–542. doi: 10.1007/s11235-016-0190-2.
10. Melchiori A., Sgalambro A. A branch and price algorithm to solve the quickest multicommodity k-splittable flow problem // Eur. J. Oper. Res., 2020. Vol. 282. Iss. 3. P. 846–857. doi: 10.1016/j.ejor.2019.10.016.

11. Кормен Т. Х., Лейзерсон Ч. И., Ривест Р. Л., Штайн К. Алгоритмы: построение и анализ / Пер. с англ. — М.: Вильямс, 2005. 1296 с. (Cormen T. H., Leiserson C. E., Rivest R. L., Stein C. Introduction to algorithms. — 2nd ed. — Cambridge, London: The MIT Press, 2001. 1180 p.)

Поступила в редакцию 31.05.23

## MULTIUSER NETWORK LOAD ANALYSIS BY SPLITTING FLOWS ALONG THE SHORTEST ROUTES

Yu. E. Malashenko and I. A. Nazarova

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** In computational experiments on a multicommodity network model, two ways of transmitting flows of different types along shortest routes are investigated. In the first case, the transmitted internodal flows are equal in magnitude. In the other — a nondiscriminatory distribution is defined in which all pairs of nodes are distributed the same resources. The total load of the network edges resulting from the simultaneous transmission of all internodal flows is considered to be given. The proposed method allows one to obtain guaranteed estimates of the specific resource costs of the network and the maximum feasible load of the edges while transmitting split internodal flows along the shortest routes found. The results of a comparative analysis of the equalization distribution of flows and resources in networks with different structural features are given. The algorithmic scheme has a polynomial estimate of the required number of operations.

**Keywords:** multicommodity flow model; distribution of internodal flows and loads; network peak load

**DOI:** 10.14357/19922264230305

**EDN:** NLUSQJ

### References

1. Malashenko, Y. E., and I. A. Nazarova. 2022. Estimate of resource distribution with the shortest paths in the multiuser network. *J. Comput. Sys. Sc. Int.* 61(4):599–610. doi: 10.1134/S106423072204013X.
2. Malashenko, Y. E., and I. A. Nazarova. 2022. Analysis of the load distribution and internodal flows under different routing strategies in a multiuser network. *J. Comput. Sys. Sc. Int.* 61(6): 970–980. doi: 10.1134/S1064230722060132.
3. Malashenko, Y. E., and I. A. Nazarova. 2023. Quantitative analysis of flow distributions in a multi-user telecommunication network. *J. Comput. Sys. Sc. Int.* 62(2): 324–335.
4. Salimifard, K., and S. Bigharaz. 2020. The multicommodity network flow problem: State of the art classification, applications, and solution methods. *Operational Research* 22(1):1–47. doi: 10.1007/s12351-020-00564-8.
5. Luss, H. 2012. *Equitable resource allocation: Models, algorithms, and applications*. Hoboken, NJ: John Wiley & Sons. 376 p.
6. Balakrishnan, A., G. Li, and P. Mirchandani. 2017. Optimal network design with end-to-end service requirements. *Oper. Res.* 65(3):729–750. doi: 10.1287/opre.2016.1579.
7. Caramia, M., and A. Sgalambro. 2010. A fast heuristic algorithm for the maximum concurrent k-splittable flow problem. *Optim. Lett.* 4(1):37–55. doi: 10.1007/s11590-009-0147-4.
8. Kabadurmus, O., and A. E. Smith. 2016. Multicommodity k-splittable survivable network design problems with relays. *Telecommun. Syst.* 62(1):123–133. doi: 10.1007/s11235-015-0067-9.
9. Bialon, P. 2017. A randomized rounding approach to a k-splittable multicommodity flow problem with lower path flow bounds affording solution quality guarantees. *Telecommun. Syst.* 64(3):525–542. doi: 10.1007/s11235-016-0190-2.
10. Melchiori, A., and A. Sgalambro. 2020. A branch and price algorithm to solve the quickest multicommodity k-splittable flow problem. *Eur. J. Oper. Res.* 282(3):846–857. doi: 10.1016/j.ejor.2019.10.016.
11. Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein. 2001. *Introduction to algorithms*. 2nd ed. Cambridge, London: The MIT Press. 1180 p.

Received May 31, 2023

### Contributors

**Malashenko Yuri E.** (b. 1946) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; malash09@ccas.ru

**Nazarova Irina A.** (b. 1966) — Candidate of Science (PhD) in physics and mathematics, scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; irina-nazar@yandex.ru

# ОПТИМИЗАЦИЯ СХЕМЫ РАСПРЕДЕЛЕНИЯ БУФЕРНОЙ ПАМЯТИ УЗЛА ПАКЕТНОЙ КОММУТАЦИИ

Я. М. Агаларов<sup>1</sup>

**Аннотация:** Рассматривается буфер узла коммутации (УК) пакетов, совместно используемый несколькими выходными линиями связи. Совместное использование буферной памяти (БП) несколькими пользователями позволяет уменьшить объем памяти, необходимый для удовлетворения требований к задержке и вероятности потерь пакетов. Однако возникает проблема распределения БП между пользователями, поскольку отдельные пользователи, заняв всю память, могут ограничить (или закрыть) доступ к линиям связи другим пользователям, что может значительно снизить производительность УК в целом. Существует множество различных схем распределения БП, одна из которых, называемая SMA (Sharing with Minimum Allocation), исследуется в данной работе с целью снижения затрат, связанных с отклонением и задержкой пакетов и эксплуатацией накопителя и линий связи. В качестве модели УК используется многопоточковая система массового обслуживания (СМО) с параллельными приборами типа  $M/M/s/K$  с совместно используемым по схеме SMA буфером с фиксированным числом зарезервированных за каждым прибором мест хранения. Сформулирована математическая постановка задачи оптимизации схемы SMA по объему общедоступных мест буфера с целью минимизации потерь системы, возникающих из-за отклонения заявок, задержки заявок в очереди и эксплуатации буфера и приборов. Доказана теорема о границах области, содержащей точку глобального оптимума. Приведен также ряд вытекающих из теоремы утверждений о точке глобального оптимума целевой функции для других моделей УК и частных случаев SMA.

**Ключевые слова:** узел коммутации; распределение буферной памяти; оптимизация; система массового обслуживания

**DOI:** 10.14357/19922264230306

**EDN:** QLXCKV

## 1 Введение

В сетях связи с коммутацией пакетов УК имеют БП, в которой пакеты, поступающие со всех входных линий связи, сохраняются в ожидании, прежде чем передаются на соответствующие выходы. Качество обслуживания в таких сетях в значительной степени зависит от объема памяти УК. Возникает эта зависимость из-за того, что при ограниченной пропускной способности каналов связи и небольшом объеме памяти могут быть недопустимо большие потери пакетов или рост числа повторных передач пакетов и, как следствие, недопустимо сильное снижение производительности сети, а при большом объеме памяти возможны длинные очереди и, как следствие, недопустимо большие задержки пакетов в УК. Совместное использование общей памяти в такой многопользовательской системе может значительно повысить производительность УК, но создает новую проблему — требует управление БП [1]. Отсюда и возникает задача выбора оптимального объема и схемы распределения БП УК с целью выполнения определенных требований к качеству обслуживания пакетов и затратам, связанным со стоимостью оборудования и технической эксплуатации УК [1].

Сразу же после появления сетей связи с КП был опубликован ряд работ, в которых рассматривалась проблема совместного использования БП. Например, в [2] была предложена схема совместного использования БП, которая использует пороговое правило, близкое к оптимальному, а в работе [3] было исследовано несколько схем совместного использования, а именно: CS (Complete Sharing), при которой поступающий пакет принимается, если доступно какое-либо место для хранения; CP (Complete Partitioning), при которой все хранилище постоянно распределяется между выходными линиями; SMQ (Sharing with Maximum Allocation), при которой ограничивается число мест хранения, выделенных каждой выходной линии; SMA, при которой для каждой выходной линии всегда зарезервировано минимальное число мест хранения, а остальные одинаково доступны всем выходным линиям; SMQMA (Sharing with Maximum Queue Length and Minimum Allocation), которая представляет собой комбинацию схем SMQ и SMA.

Наряду с выбором эффективной схемы важно еще оптимизировать параметры этой схемы, в частности общий объем памяти, что, за исключением единичных частных случаев (ниже на некоторые из

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, agglar@yandex.ru

них сделаны ссылки), остается задачей открытой (в смысле ее точного решения) из-за ее сложности. Как правило, разумно ожидать роста пропускной способности (здесь и ниже пропускная способность определяется как интенсивность потока пакетов на выходе системы) с увеличением емкости буфера. Это проверено для многих стандартных СМО, таких как очередь  $M/M/s/K$  (см., например, [4]), и в [5] показано, что результат справедлив и при более общих условиях. Однако в целом результат неверен (см., например, [6]).

Интерес к управлению БП исследователи проявляют и в настоящее время, что видно по множеству научных статей в отечественных и зарубежных научных журналах, опубликованных за последнее время (см., например, [7–17]). В основном они посвящены анализу различных схем распределения БП, однако работ, в которых получены точные решения по оптимизации схем, в литературе встречается мало. Ниже приводится обзор некоторых из них, в которых проводились исследования зависимости показателей качества работы УК от типа схемы распределения БП и значений ее параметров. В [10] авторы рассматривают очередь  $M/G/1/K$ -PS, выводят выражения для различных показателей производительности, связанных с пропускной способностью и средней скоростью передачи и приводят несколько структурных результатов о взаимосвязях между этими показателями. Показано, что изменения емкости буфера оказывают наиболее значительное влияние, когда нагрузка на систему не слишком низкая и не слишком высокая. Для систем  $M/M/1/K$  и  $M/G/1/K$ -PS показано, что увеличение пропускной способности (как номинальное, так и относительное), которое было бы получено за счет добавления дополнительного буферного пространства, унимодально по нагрузке. В работе [11] предложен эвристический алгоритм решения задачи совместного управления объемом БП и скоростью передачи линий связи, который рассматривается в следующих аспектах:

- (1) при заданной степени загрузки каналов и заданной скорости передачи найти необходимый объем БП УК, обеспечивающий заданную среднюю задержку в сети и требуемую вероятность потери ячеек вследствие переполнения буферов;
- (2) при заданной степени загрузки каналов и известных объемах БП найти необходимую скорость передачи, обеспечивающую заданную среднюю задержку в сети и требуемую вероятность потери ячеек вследствие переполнения буфера.

В работе [12] авторы решили задачу выбора объема БП для СМО типа  $M/M/1/K$ , сформулировав ее как задачу нелинейного программирования:

- (1) для фиксированной входной нагрузки найти объем накопителя (БП), при котором средняя задержка достигает минимума, а интенсивность потерь не превышает заданной величины;
- (2) для фиксированного объема накопителя (БП) найти значение входной нагрузки, при которой средняя задержка достигает минимума, а интенсивность потерь не превышает заданной величины.

В работе [13] в предположении пуассоновских входящих потоков, экспоненциального времени обслуживания и одноканальных линий передачи рассмотрено динамическое распределение (управление) БП и получен для случая трех каналов вид допустимого пространства состояний, соответствующий оптимальному решению. Схема динамического совместного использования БП, которая позволяет каждой очереди увеличиваться до динамически назначаемого порога, предложена в работе [9]. Этот порог вычисляется как произведение оставшегося буфера с предопределенным параметром. В работе [7] предложена схема (названная drop-on-demand, или сокращенно DoD) совместного использования, которая допускает отбрасывание принятых пакетов и, следовательно, не относится к классу упомянутых выше схем. Согласно этой схеме поступающий пакет всегда принимается, если имеется пустой буфер. Если поступает пакет, предназначенный для выходной линии  $i$ , и обнаруживается, что буфер заполнен, а выходная линия  $I$  содержит больше пакетов в общей памяти, чем любые другие порты, выполняется следующее действие:

- если  $i = I$ , то поступающий пакет отбрасывается;
- если  $i \neq I$ , то поступающий пакет принимается в буфер и один пакет к линии  $I$  отбрасывается.

Были приведены численные примеры, показывающие, что схема DoD обеспечивает лучшую пропускную способность, чем схемы CS и CP. В работе [8] на примере двухканальной СМО показано, что эта политика оптимальна только для симметричных систем. Там же утверждается, что динамические схемы распределения БП трудно реализуемы по сравнению со статическими. Для СМО с конечной совместно используемой БП и несколькими очередями пакетов в [15] рассмотрена схема с динамическими индивидуальными потолками, зависящими от числа свободных мест в БП (названная



FAB — Flow Aware Buffer). Приведены результаты численного сравнительного анализа схем полного разделения, полного совместного использования, динамического совместного использования и FAB на примерах модели УК в виде параллельных очередей  $M/G/1/K$  с ограниченным накопителем. В [16] сформулирована задача оптимизации общих стоимостных затрат (стоимость потери клиента, стоимость хранения буфера и эксплуатационные расходы) СМО типа  $M/G/1/K$  с несколькими отпусками и численно исследовано влияние параметров системы, таких как размер буфера, продолжительность отпуска и распределение времени обслуживания, на показатели производительности и общую стоимость затрат. Алгоритм оптимизации схемы СР при целевой функции дохода, учитывающей среднюю задержку и вероятность потерь пакетов, предложен в [14] в рамках модели УК, представленной параллельными СМО типа  $M/G/1/K$ . В рамках этой модели УК доказана унимодальность целевой функции по объему БП. Аналогичное утверждение для схемы СР доказано и в случае УК, представленного параллельными СМО типа  $G/M/1/K$  с резервными каналами [17].

Ниже исследуется задача оптимизации схемы SMA при стоимостной целевой функции, учитывающей стоимостные затраты из-за потерь пакетов, задержки пакетов в очереди и эксплуатации накопителя и приборов.

## 2 Модель узла коммутации и постановка задачи

В качестве модели УК рассматривается многопоточковая и многоканальная СМО с общим накопителем (БП) емкости  $N$ , на которую поступает  $n$  пуассоновских потоков заявок (пакетов) с интенсивностями  $\lambda_j > 0$ ,  $j = 1, \dots, n$ , в которой к каждому  $j$ -му потоку для обслуживания заявок прикреплен соответствующая  $j$ -я линия из  $s_j$  однотипных приборов (каналов). При этом каждая заявка может занимать только одно свободное место (ячейку) в накопителе, на  $j$ -линию поступают только заявки  $j$ -потока.

Введем обозначения:  $k_j$  — число  $j$ -заявок в накопителе,  $\bar{k} = (k_1, \dots, k_n)$  — вектор состояния системы;  $\bar{a} = (a_1, \dots, a_n)$ ,  $a_j \geq 0$ , — число закрепленных за  $j$ -заявками мест в накопителе (которые могут занять только  $j$ -заявки),  $\sum_{j=1}^n a_j \leq N$ ;  $L$  — число общедоступных мест в накопителе:

$$L = N - \sum_{j=1}^n a_j;$$

$K^L = \{\bar{k} : \sum_{j=1}^n k_j \leq \sum_{j=1}^n a_j + L\}$  — множество всех возможных состояний системы;  $\bar{K}_j^L = \{\bar{k} \in K^L : k_j \geq a_j \text{ и } \sum_{j=1}^n (k_j - a_j) = L\}$  — множество состояний, при которых  $j$ -заявка не допускается в накопитель (теряется);  $K_j^L = \{\bar{k} \in K^L : k_j < a_j \text{ или } \sum_{j=1}^n (k_j - a_j)^+ < L\}$  — множество состояний, при которых  $j$ -заявке доступно место в накопителе, где

$$(k_j - a_j)^+ = \begin{cases} k_j - a_j & \text{при } k_j > a_j; \\ 0 & \text{при } k_j \leq a_j. \end{cases}$$

Согласно введенным обозначениям заявки допускаются в накопитель по схеме SMA [3]:  $j$ -заявка допускается в накопитель, если выполняется условие  $\bar{k} \in K_j^L$ , а при  $\bar{k} \in \bar{K}_j^L$  отклоняется (теряется).

Каждая поступившая  $j$ -заявка занимает одно из доступных и свободных мест в накопителе и один свободный прибор в  $j$ -линии сразу, если  $k_j < s_j$ , или после освобождения, если  $k_j \geq s_j$ , а после завершения обслуживания покидает систему, освободив одновременно прибор и место в накопителе. Время обслуживания  $j$ -заявки — экспоненциальная случайная величина с заданным параметром  $\mu_j$ ,  $j = 1, \dots, n$ .

Предположим, что  $j$ -заявки в первую очередь занимают закрепленные за ними места и, если есть общедоступное место, занятое  $j$ -заявкой, то освободившееся от  $j$ -заявки закрепленное место становится общедоступным, а это занятое общедоступное место становится закрепленным. Таким образом, число закрепленных за каждым потоком мест остается постоянным. Если вновь поступившая заявка застает все доступные ей места в накопителе занятыми, то она теряется безвозвратно.

Процесс перехода описанной СМО из состояния в состояние — марковский и имеет следующее распределение стационарных вероятностей состояний [18]:

$$\pi_{\bar{k}}(L) = \pi_{\bar{0}}(L) \prod_{j=1}^n Z_j(k_j). \quad (1)$$

Здесь

$$\pi_{\bar{0}}(L) = \left[ \sum_{\bar{k} \in K} \prod_{j=1}^n Z_j(k_j) \right]^{-1},$$

$$\bar{0} = \{\bar{k} : k_i = 0, i = 1, \dots, n\};$$

$$Z_j(k_j) = \begin{cases} \frac{\rho_j^{k_j}}{k_j!}, & 0 \leq k_j \leq s_j; \\ \frac{\rho_j^{s_j}}{s_j!} \left( \frac{\rho_j}{s_j} \right)^{k_j - s_j}, & s_j < k_j \leq a_j + L, \end{cases}$$

где

$$\rho_j = \frac{\lambda_j}{\mu_j}.$$

В дальнейшем систему рассматриваем только в стационарном режиме работы и считаем, что значения параметров  $a_j = s_j, j = 1, \dots, n$ , и являются постоянными величинами.

В качестве показателя эффективности системы используется функция предельного дохода в единицу времени, учитывающая плату за обслуживание, потери из-за ожидания в очереди, отклонения заявок и технического обслуживания системы. Считается, что плату за обслуживание система получает в момент приема заявки в накопитель. Доход измеряется в стоимостных единицах и зависит также от следующих стоимостных параметров:  $C_{0,i} \geq 0$  — плата, получаемая системой, если поступившая  $i$ -заявка будет обслужена системой;  $C_{1,i} \geq 0$  — штраф за отклонение поступившей  $i$ -заявки;  $C_{2,i} \geq 0$  — штраф за единицу времени ожидания  $i$ -заявки в очереди;  $C_3 \geq 0$  — затраты системы в единицу времени на техническое обслуживание одного места в накопителе;  $C_4 \geq 0$  — затраты системы в единицу времени на техническое обслуживание всех приборов системы.

Доход выражается функцией следующего вида:

$$Q(L) = \lambda \sum_{\bar{k} \in K} \pi_{\bar{k}}(L) q_{\bar{k}}, \quad \lambda = \sum_{j=1}^n \lambda_j,$$

где  $L$  — переменная величина;  $q_{\bar{k}}$  — средний доход, получаемый системой за период между соседними моментами поступления извне заявок, если в момент поступления в начале периода система оказалась в состоянии  $\bar{k}$ .

Задача оптимизации объема общедоступных мест в накопителе сформулирована в виде математической задачи

$$L^* = \arg \max_{L \geq 0} Q(L).$$

### 3 Метод решения

Введем обозначения:

$$\bar{K}_{j,m}^L = \left\{ \bar{k} \in \bar{K}_j^L : k_j = m \geq a_j, \sum_{i=1, i \neq j}^n (k_i - a_i)^+ = L - (m - a_j) \right\};$$

$$K_{j,m}^L = \left\{ \bar{k} \in K_j^L : k_j = m, \sum_{i=1, i \neq j}^n (k_i - a_i)^+ < L - (k_j - a_j)^+ \text{ или } k_j = m < a_j \right\};$$

$\pi_{j,m}^-(L) = \sum_{\bar{k} \in K_{j,m}^L} \pi_{\bar{k}}(L)$  — стационарная вероятность того, что  $\bar{k} \in K_{j,m}^L$ ;  $\pi_{j,m}^+(L) = \sum_{\bar{k} \in \bar{K}_{j,m}^L} \pi_{\bar{k}}(L)$  — стационарная вероятность того, что  $\bar{k} \in \bar{K}_{j,m}^L$ ;  $\pi_{j,m}(L) = (\pi_{j,m}^-(L) + \pi_{j,m}^+(L))$  — стационарная вероятность того, что  $k_j = m$ ;  $\chi_j(\bar{k})$  — функция Хевисайда:

$$\chi_j(\bar{k}) = \begin{cases} 1, & \bar{k} \in K_j^L; \\ 0, & \bar{k} \in \bar{K}_j^L; \end{cases}$$

$d_{j,m}$  — стоимость средних суммарных потерь из-за ожидания и отклонения  $j$ -заявок за период между соседними моментами поступления извне заявок, если в начале периода число  $j$ -заявок в накопителе было равно  $m$ .

**Теорема 1.** Функцию  $Q(L)$  можно записать в виде:

$$Q(L) = \sum_{j=1}^n [(\lambda - \lambda_j) Q_j^+(L) + \lambda_j Q_j^-(L)] - C_3 N - C_4, \quad (2)$$

где  $Q_j^+(L)$  и  $Q_j^-(L), j = 1, \dots, n$ , — унимодальные по  $L \geq 0$  функции:

$$Q_j^+(L) = \sum_{m=0}^{a_j+L-1} d_{j,m} \pi_{j,m}^-(L) + \sum_{m=0}^{a_j+L} d_{j,m} \pi_{j,m}^+(L);$$

$$Q_j^-(L) = \sum_{m=0}^{a_j+L-1} (d_{j,m+1} + C_{0,j}) \pi_{j,m}^-(L) + \sum_{m=0}^{a_j+L} (d_{j,m} - C_{1,j}) \pi_{j,m}^+(L).$$

**Доказательство.** Легко видеть, что для дохода  $q_{\bar{k}}$  и величин  $d_{j,m}, j = 1, \dots, n, m = s_j - 1, \dots, \dots, a_j + L$  имеет место следующее соотношение:

$$q_{\bar{k}} = \begin{cases} \sum_{\substack{l=1, \\ l \neq j}}^n d_{l,k_l} + d_{j,m+1} + C_{0,j} - \frac{C_3 N + C_4}{\lambda}, & \text{если извне поступила } j\text{-заявка и } \bar{k} \in K_{j,m}^L; \\ \sum_{l=1}^n d_{l,k_l} - C_{1,j} - \frac{C_3 N + C_4}{\lambda}, & \text{если извне поступила } j\text{-заявка и } \bar{k} \in \bar{K}_{j,m}^L. \end{cases}$$

Тогда для целевой функции справедливо выражение:

$$\begin{aligned}
 Q(L) &= \lambda \sum_{\bar{k} \in K^L} \pi_{\bar{k}}(L) \sum_{j=1}^n \frac{\lambda_j}{\lambda} \left[ \sum_{\substack{l=1, \\ l \neq j}}^n d_{l,k_l} + \right. \\
 &\quad \left. + (d_{j,k_{j+1}} + C_{0,j}) \chi_j(\bar{k}) + \right. \\
 &\quad \left. + (d_{j,k_j} - C_{1,j}) [1 - \chi_j(\bar{k})] - \frac{C_3 N + C_4}{\lambda} \right] = \\
 &= \sum_{j=1}^n \lambda_j \left[ \sum_{l=1}^n \sum_{m=0}^{a_l+L} \pi_{l,m}(L) d_{l,m} - \right. \\
 &\quad \left. - \sum_{m=0}^{a_j+L} \pi_{j,m}(L) d_{j,m} + \sum_{m=0}^{a_l+L-1} \pi_{j,m}^-(L) d_{l,m+1} + \right. \\
 &\quad \left. + \sum_{m=0}^{a_j+L} \pi_{j,m}^+(L) d_{j,m} + C_{0,j} \sum_{m=0}^{a_j+L-1} \pi_{j,m}^-(L) - \right. \\
 &\quad \left. - C_{1,j} \sum_{m=0}^{a_j+L} \pi_{j,m}^+(L) \right] - C_3 N - C_4 = \\
 &= \sum_{j=1}^n (\lambda - \lambda_j) \left[ \sum_{m=0}^{a_j+L-1} d_{j,m} \pi_{j,m}^-(L) + \right. \\
 &\quad \left. + \sum_{m=0}^{a_j+L} d_{j,m} \pi_{j,m}^+(L) \right] + \\
 &\quad + \sum_{j=1}^n \lambda_j \left[ \sum_{m=0}^{a_j+L-1} (d_{j,m+1} + C_{0,j}) \pi_{j,m}^-(L) + \right. \\
 &\quad \left. + \sum_{m=0}^{a_j+L} (d_{j,m} - C_{1,j}) \pi_{j,m}^+(L) \right] - C_3 N - C_4.
 \end{aligned}$$

Следовательно, равенство (2) выполняется.

Приведем аналитические выражения для параметра  $d_{j,m}$ . Воспользуемся результатами работы [10]. Согласно этой работе (см. вывод формулы (5) в [10]),

$$d_{j,m} = \begin{cases} 0, & m \leq s_j; \\ -\frac{C_{2,j}}{2\mu_j s_j} \left[ \sum_{l=1}^{m+1-s_j} l(l+2s_j-2m-1) r_{j,l} - \right. \\ \left. - (m-s_j)(m+1-s_j) \sum_{l=m+2-s_j}^{\infty} r_{j,l} \right], & s_j \leq m \leq a_j + L; \\ d_{j,m-1}, & m = a_j + L, \end{cases}$$

где  $r_{j,l}$  — вероятность того, что за период между соседними поступлениями заявок на  $j$ -линии

завершат обслуживание ровно  $l$  заявок при условии, что в начале периода в очереди не менее  $l$  заявок:

$$r_{j,l} = \lambda \int_0^{\infty} \frac{(\mu_j s_j t)^l}{l!} e^{-(\mu_j s_j + \lambda)t} dt \text{ при } l \geq 0.$$

Согласно работе [10], справедливо также соотношение:

$$\begin{aligned}
 d_{j,m} &= d_{j,m+1} - \frac{C_{2,j}}{\mu_j s_j} \sum_{l=1}^{m+1-s_j} l r_{j,l} - \\
 &\quad - \frac{C_{2,j}(m+1-s_j)}{\mu_j s_j} \sum_{l=m+2-s_j}^{\infty} r_{j,l}, \\
 s_j \leq m \leq a_j + L - 1. \quad (3)
 \end{aligned}$$

**Лемма 1.** Для вероятностей  $\pi_{j,m}^-(L)$  и  $\pi_{j,m}^+(L)$ ,  $j = 1, \dots, n$ ,  $m = s_j - 1, \dots, a_j + L - 1$ , справедливы равенства:

$$\left. \begin{aligned} \pi_{j,m+1}^-(L+1) &= \pi_{j,m}^-(L) A_j(L+1); \\ \pi_{j,m+1}^+(L+1) &= \pi_{j,m}^+(L) A_j(L+1), \end{aligned} \right\} \quad (4)$$

где

$$A_j(L+1) = \frac{1 - P_{j,s_j-1}(L+1)}{1 - P_{j,s_j-2}(L)}, \quad P_{j,m}(L) = \sum_{l=0}^m \pi_{i,j}(L).$$

**Доказательство.** Из (1) и равенства  $K_{j,m}^L = K_{j,m+1}^{L+1}$  следует

$$\begin{aligned}
 &\pi_{j,m+1}^-(L+1) - \pi_{j,m}^-(L) = \\
 &= \pi_0(L+1) \frac{\rho_j^{s_j}}{s_j!} \left( \frac{\rho_j}{s_j} \right)^{m+1-s_j} \sum_{\substack{\bar{k} \in K_{j,m+1}^{L+1} \\ l \neq j}} \prod_{l=1}^n Z_j(k_j) - \\
 &\quad - \pi_0(L) \frac{\rho_j^{s_j}}{s_j!} \left( \frac{\rho_j}{s_j} \right)^{m-s_j} \sum_{\substack{\bar{k} \in K_{j,m}^L \\ l \neq j}} \prod_{l=1}^n Z_j(k_j) = \\
 &= \pi_0(L+1) \pi_0(L) \frac{\rho_j^{s_j}}{s_j!} \left( \frac{\rho_j}{s_j} \right)^{m-s_j} \sum_{\substack{\bar{k} \in K_{j,m}^L \\ l \neq j}} \prod_{l=1}^n Z_j(k_j) \times \\
 &\quad \times \left[ \frac{\rho_j}{s_j} \left[ \sum_{m=0}^{s_j-1} \frac{\rho_j^m}{m!} \sum_{\substack{\bar{k} \in K_{j,m}^L \\ l \neq j}} \prod_{l=1}^n Z_j(k_j) - \right. \right. \\
 &\quad \left. \left. - \frac{\rho_j^{s_j}}{s_j!} \sum_{m=s_j}^{a_j+L} \left( \frac{\rho_j}{s_j} \right)^{m-s_j} \sum_{\substack{\bar{k} \in K_{j,m}^L \\ l \neq j}} \prod_{l=1}^n Z_j(k_j) \right] - \right. \\
 &\quad \left. - \sum_{m=0}^{s_j-1} \frac{\rho_j^m}{m!} \sum_{\substack{\bar{k} \in K_{j,m+1}^{L+1} \\ l \neq j}} \prod_{l=1}^n Z_j(k_j) + \right. \\
 &\quad \left. + \frac{\rho_j^{s_j}}{s_j!} \sum_{m=s_j}^{a_j+L+1} \left( \frac{\rho_j}{s_j} \right)^{m-s_j} \sum_{\substack{\bar{k} \in K_{j,m+1}^{L+1} \\ l \neq j}} \prod_{l=1}^n Z_j(k_j) \right] =
 \end{aligned}$$

$$\begin{aligned}
 &= \pi_0(L+1) \frac{\rho_j^{s_j}}{s_j!} \left(\frac{\rho_j}{s_j}\right)^{m+1-s_j} \times \\
 &\quad \times \sum_{\bar{k} \in K_{j,m+1}^{L+1}} \prod_{\substack{l=1, \\ l \neq j}}^n Z_j(k_j) \pi_0(L) \sum_{m=0}^{s_j-2} \frac{\rho_j^m}{m!} \times \\
 &\quad \times \sum_{\bar{k} \in K_{j,m}^L} \prod_{\substack{l=1, \\ l \neq j}}^n Z_j(k_j) - \pi_0(L) \frac{\rho_j^{s_j}}{s_j!} \left(\frac{\rho_j}{s_j}\right)^{m-s_j} \times \\
 &\quad \times \sum_{\bar{k} \in K_{j,m}^L} \prod_{\substack{l=1, \\ l \neq j}}^n Z_j(k_j) \times \\
 &\quad \times \pi_0(L+1) \sum_{m=0}^{s_j-1} \frac{\rho_j^m}{m!} \sum_{\bar{k} \in K_{j,m+1}^{L+1}} \prod_{\substack{l=1, \\ l \neq j}}^n Z_j(k_j) = \\
 &= \pi_{j,m+1}^-(L+1) P_{j,s_j-2}(L) - \pi_{j,m}^-(L) P_{j,s_j-1}(L+1).
 \end{aligned}$$

Следовательно, первое равенство в (4) выполняется. Точно так же доказывается второе равенство в (4).

**Лемма 2.** Для функций  $Q_j^+(L)$  и  $Q_j^-(L)$  справедливы соотношения:

$$\left. \begin{aligned}
 Q_j^-(L) - Q_j^-(L+1) &= \\
 &= [1 - A_j(L+1)] [Q_j^-(L) - G_j^-(L)]; \\
 Q_j^+(L) - Q_j^+(L+1) &= \\
 &= [1 - A_j(L+1)] [Q_j^+(L) - G_j^+(L)],
 \end{aligned} \right\} \quad (5)$$

где

$$\begin{aligned}
 G_j^-(L) &= C_{0,j} + \frac{C_{2,j}}{1 - A_j(L+1)} \left[ \sum_{m=s_j}^{a_j+L} \left[ \frac{1}{\lambda} - \right. \right. \\
 &\quad \left. \left. - \frac{1}{\mu_j s_j} \sum_{l=m+2-s_j}^{\infty} (l - m - 1 + s_j) r_{j,l} \right] \pi_{j,m}^-(L+1) + \right. \\
 &\quad \left. + \sum_{m=s_j}^{a_j+L+1} \left[ \frac{1}{\lambda} - \right. \right. \\
 &\quad \left. \left. - \frac{1}{\mu_j s_j} \sum_{l=m+1-s_j}^{\infty} (l - m + s_j) r_{j,l} \right] \pi_{j,m}^+(L+1) \right]; \\
 G_j^+(L) &= \frac{C_{2,j}}{1 - A_j(L+1)} \left[ \sum_{m=s_j}^{a_j+L} \left[ \frac{1}{\lambda} - \right. \right. \\
 &\quad \left. \left. - \frac{1}{\mu_j s_j} \sum_{l=m+1-s_j}^{\infty} (l - m + s_j) r_{j,l} \right] \pi_{j,m}^-(L+1) + \right. \\
 &\quad \left. + \sum_{m=s_j}^{a_j+L+1} \left[ \frac{1}{\lambda} - \right. \right. \\
 &\quad \left. \left. - \frac{1}{\mu_j s_j} \sum_{l=m+1-s_j}^{\infty} (l - m + s_j) r_{j,l} \right] \pi_{j,m}^+(L+1) \right].
 \end{aligned}$$

**Доказательство.** Докажем первое равенство в (5). Обозначим  $\Delta_{j,m+1} = d_{j,m+1} - d_{j,m}$ . Из (2) и (4) после несложных преобразований получим равенства:

$$\begin{aligned}
 Q_j^-(L) - Q_j^-(L+1) &= \sum_{m=0}^{s_j-2} (d_{j,m+1} + C_{0,j}) \pi_{j,m}^-(L) - \\
 &\quad - \sum_{m=0}^{s_j-1} (d_{j,m+1} + C_{0,j}) \pi_{j,m}^-(L+1) + \\
 &\quad + \sum_{m=s_j-1}^{a_j+L-1} (d_{j,m+1} + C_{0,j}) \pi_{j,m}^-(L) - \\
 &\quad - \sum_{m=s_j}^{a_j+L} (d_{j,m+1} + C_{0,j}) \pi_{j,m}^-(L+1) + \\
 &\quad + \sum_{m=0}^{s_j-2} (d_{j,m} - C_{1,j}) \pi_{j,m}^+(L) - \\
 &\quad - \sum_{m=0}^{s_j-1} (d_{j,m} - C_{1,j}) \pi_{j,m}^+(L+1) + \\
 &\quad + \sum_{m=s_j-1}^{a_j+L} (d_{j,m} - C_{1,j}) \pi_{j,m}^+(L) - \\
 &\quad - \sum_{m=s_j}^{a_j+L+1} (d_{j,m} - C_{1,j}) \pi_{j,m}^+(L+1) = \\
 &= C_{0,j} [P_{j,s_j-2}(L) - P_{j,s_j-1}(L+1)] + \\
 &\quad + \sum_{m=s_j-1}^{a_j+L-1} (d_{j,m+1} + C_{0,j}) \pi_{j,m}^-(L) - \\
 &\quad - A_j(L+1) \sum_{m=s_j-1}^{a_j+L-1} (d_{j,m+1} + \Delta_{j,m+2} + C_{0,j}) \pi_{j,m}^-(L) + \\
 &\quad + \sum_{m=s_j-1}^{a_j+L} (d_{j,m} - C_{1,j}) \pi_{j,m}^+(L) - \\
 &\quad - A_j(L+1) \sum_{m=s_j-1}^{a_j+L} (d_{j,m} + \Delta_{j,m+1} - C_{1,j}) \pi_{j,m}^+(L) = \\
 &= -C_{0,j} [1 - A_j(L+1)] + [1 - A_j(L+1)] Q_j^-(L) - \\
 &\quad - A_j(L+1) \sum_{m=s_j-1}^{a_j+L-1} \Delta_{j,m+2} \pi_{j,m}^-(L) - \\
 &\quad - A_j(L+1) \sum_{m=s_j-1}^{a_j+L} \Delta_{j,m+1} \pi_{j,m}^+(L) = \\
 &= [1 - A_j(L+1)] \left[ Q_j^-(L) - C_{0,j} - \right.
 \end{aligned}$$

$$- \frac{1}{1 - A_j(L+1)} \left[ \sum_{m=s_j}^{a_j+L} \Delta_{j,m+1} \pi_{j,m}^-(L+1) + \sum_{m=s_j}^{a_j+L+1} \Delta_{j,m} \pi_{j,m}^+(L+1) \right].$$

Применив формулу (3), получим

$$Q_j^-(L) - Q_j^-(L+1) = [1 - A_j(L+1)] \left[ Q_j^-(L) - C_{0,j} - \frac{C_{2,j}}{1 - A_j(L+1)} \left[ \sum_{m=s_j}^{a_j+L} \left[ \frac{1}{\lambda} - \frac{1}{\mu_j s_j} \sum_{l=m+2-s_j}^{\infty} (l - m - 1 + s_j) r_{j,l} \right] \pi_{j,m}^-(L+1) + \sum_{m=s_j}^{a_j+L+1} \left[ \frac{1}{\lambda} - \frac{1}{\mu_j s_j} \sum_{l=m+1-s_j}^{\infty} (l - m + s_j) r_{j,l} \right] \pi_{j,m}^+(L+1) \right] \right]. \quad (6)$$

Следовательно, первое равенство в (5) выполняется. Точно так же для второго равенства в (5) получим:

$$Q_j^+(L) - Q_j^+(L+1) = [1 - A_j(L+1)] \left[ Q_j^+(L) - \frac{C_{2,j}}{1 - A_j(L+1)} \left[ \sum_{m=s_j}^{a_j+L} \left[ \frac{1}{\lambda} - \frac{1}{\mu_j s_j} \sum_{l=m+1-s_j}^{\infty} (l - m + s_j) r_{j,l} \right] \pi_{j,m}^-(L+1) + \sum_{m=s_j}^{a_j+L+1} \left[ \frac{1}{\lambda} - \frac{1}{\mu_j s_j} \sum_{l=m+1-s_j}^{\infty} (l - m + s_j) r_{j,l} \right] \pi_{j,m}^+(L+1) \right] \right]. \quad (7)$$

Следовательно, выполняется и второе равенство в (5).

**Лемма 3.** *Функции  $G_j^-(L)$  и  $G_j^+(L)$  — невозрастающие функции.*

**Доказательство.** Обратим внимание на то, что сумма во внешней квадратной скобке в (6) выражает среднее время нахождения  $j$ -линии в состоянии с полностью занятыми приборами за период между соседними поступлениями заявок в системе с общедоступными местами в количестве  $L+1$  при условии, что поступившая извне заявка принадлежит  $j$ -потoku, а сумма во внешней квадратной скобке в (7) выражает то же самое для  $j$ -линии при

условии, что поступившая заявка не принадлежит  $j$ -потoku. Поэтому, очевидно, выражения в указанных скобках — возрастающие по  $L$  функции.

Рассмотрим разность

$$A_j(L+2) - A_j(L+1) = \frac{1 - P_{j,s_j-1}(L+2)}{1 - P_{j,s_j-2}(L+1)} - \frac{1 - P_{j,s_j-1}(L+1)}{1 - P_{j,s_j-2}(L)}. \quad (8)$$

Справедливы равенства:

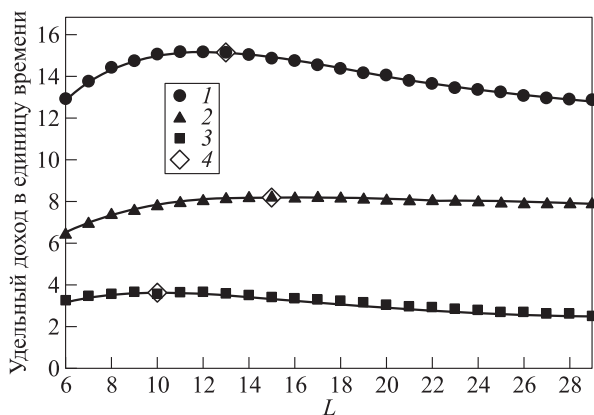
$$\begin{aligned} \frac{\pi_0(L+1)}{\pi_0(L+2)} [1 - P_{j,s_j-1}(L+2)] [1 - P_{j,s_j-2}(L)] &= \\ &= \frac{\pi_0(L)}{\pi_0(L+1)} [1 - P_{j,s_j-1}(L+1)]^2 = \\ &= \left[ \sum_{\bar{k} \in K_{j,s_j}^{L+1} \cup \bar{K}_{j,s_j}^{L+1}, l \neq j} \prod_{l=1}^n Z_j(k_j) + \sum_{m=s_j+1}^{a_j+L+1} \left( \frac{\rho_j}{s_j} \right)^{m-s_j} \sum_{\bar{k} \in K_{j,m}^{L+1} \cup \bar{K}_{j,m}^{L+1}, l \neq j} \prod_{l=1}^n Z_j(k_j) \right] \times \\ &\times \left[ \sum_{\bar{k} \in K_{j,s_j}^{L+2}, l=1, l \neq j} \prod_{l=1}^n Z_j(k_j) + \sum_{m=s_j+1}^{a_j+L+2} \left( \frac{\rho_j}{s_j} \right)^{m-s_j} \sum_{\bar{k} \in K_{j,m}^{L+2} \cup \bar{K}_{j,m}^{L+2}, l \neq j} \prod_{l=1}^n Z_j(k_j) \right]. \end{aligned}$$

Отсюда следует, что знак разности (8) совпадает со знаком разности  $\pi_0(L+2)/\pi_0(L+1) - \pi_0(L+1)/\pi_0(L)$ .

Для рассматриваемой системы в случае накопителя с  $(L+1)$  общедоступными местами величина  $\pi_0(L+1)/\pi_0(L)$  — вероятность того, что в накопителе есть хотя бы одно свободное место, и, очевидно, эта величина с увеличением  $L$  возрастает. Отсюда следует, что указанная выше разность для всех  $L \geq 0$  является положительной величиной. Следовательно,  $G_j^-(L)$  — невозрастающая функция. Точно так же доказывается лемма и в случае функции  $G_j^+(L)$ .

Из лемм 1 и 2 следует, что функции  $Q_j^-(L)$  и  $Q_j^+(L)$  удовлетворяют всем условиям теоремы в работе [19], откуда вытекает справедливость утверждений теоремы 1 об унимодальности функций  $Q_j^-(L)$  и  $Q_j^+(L)$ . Следовательно, теорема 1 доказана.

**Следствие 1.** Пусть  $L_j^-$  и  $L_j^+$  — точки максимума соответствующих функций  $Q_j^-(L)$  и  $Q_j^+(L)$ ,  $j = 1, \dots, n$ . Тогда значение  $L^*$  — точка глобального максимума функции  $D(L)$  — удовлетворяет условию:  $L_1^* \leq L^* \leq L_2^*$ , где  $L_1^* = \min\{L_j^-, L_j^+, j = 1, \dots, n\}$ ;  $L_2^* = \max\{L_j^-, L_j^+, j = 1, \dots, n\}$ .



Зависимость  $F_j(L)$  от  $L$ : 1 —  $Q(L)$ ; 2 —  $F_1(L)$ ; 3 —  $F_2(L) = F_3(L)$ ; 4 — максимальные значения функций  $F_j(L)$  и  $Q(L)$

В частности, при одинаковых для всех  $j = 1, \dots, n$  значениях величин  $\mu_j$ ,  $\lambda_j$ ,  $s_j$  и  $a_j$  функция  $Q(L)$  унимодальна.

На рисунке приведены графики функций  $Q(L)$  и  $F_j(L) = (1 - \lambda_j/\lambda)Q_j^+(L) + (\lambda_j/\lambda)Q_j^-(L)$ ,  $j = 1, 2, 3$ , демонстрирующие результаты теоремы 1, при следующих параметрах:  $\lambda_1 = 3$ ;  $\lambda_2 = \lambda_3 = 3/2$ ;  $\mu = 2$ ;  $s_1 = s_2 = s_3 = 2$ ;  $C_0 = 20$ ;  $C_{0j} = 20$ ;  $C_{1j} = 10$ ;  $C_{2j} = 3$ ;  $C_3 = 0$ ;  $C_4 = 0,01$ .

## 4 Заключение

Основной результат статьи — доказательство утверждения о том, что функция дохода рассмотренной СМО представляет собой линейную комбинацию унимодальных функций. Практическая важность результата заключается в том, что для такой функции границами области, где гарантированно лежит точка глобального максимума, служат минимальное и максимальное значения на множестве глобальных максимумов указанных унимодальных функций.

Для рассмотренной СМО аналогично теореме 1 доказывается также следующее утверждение: в случае схемы СР (схема SMA при  $L = 0$ ) если  $N$  — фиксированная величина, а  $a_j$  — переменные величины,  $j = 1, \dots, n$ , то функция дохода есть сумма  $\sum_{j=1}^n f_j(a_j)$  унимодальных на отрезке  $[1, N]$  функций  $f_j(a_j)$  таких, что  $f_j(a_j)$  — выпуклая по  $a_j$  на соответствующем отрезке  $[1, a_j^*]$ , где  $(a_1^*, \dots, a_n^*)$  — точка глобального максимума функции дохода на множестве наборов  $(a_1, \dots, a_n)$ ,  $\sum_{j=1}^n a_j \leq N$ ,  $a_j > 0$ ,  $j = 1, \dots, n$ . Из этого утверждения следует простое правило поиска точки глобального максимума функции дохода при схеме СР: пока  $\sum_{j=1}^n a_j < N$  последовательно на каждом шаге увеличиваем

на 1 одну из переменных  $a_j$  с индексом  $j^*$  таким, что приращение

$$f_{j^*}(a_{j^*} + 1) - f_{j^*}(a_{j^*}) = \max_j \{f_j(a_j + 1) - f_j(a_j) > 0, j = 1, \dots, n\}.$$

Результаты работы могут быть использованы при разработке и эксплуатации информационных и производственных потоковых систем с накопителями ограниченной емкости для повышения их эффективности.

## Литература

1. Клейнрок Л. Вычислительные системы с очередями / Пер. с англ. — М.: Мир, 1979. 600 с. (Kleinrock L. Queueing systems. — New York, NY, USA: Wiley, 1976. Vol. 2. 549 p.)
2. Irland M. Buffer management in a packet switch // IEEE T. Commun., 1978. Vol. 26. No. 3. P. 328–337. doi: 10.1109/TCOM.1978.1094076.
3. Kamoun F., Kleinrock L. Analysis of shared finite storage in a computer networks node environment under general traffic conditions // IEEE T. Commun., 1980. Vol. 28. No. 7. P. 992–1003. doi: 10.1109/TCOM.1980.1094756.
4. Köchel P. Finite queueing systems — structural investigations and optimal design // Int. J. Prod. Econ., 2004. Vol. 88. P. 157–171. doi: 10.1016/j.ijpe.2003.11.005.
5. Sonderman D. Comparing multi-server queues with finite waiting rooms, I: Same number of servers // Adv. Appl. Probab., 1979. Vol. 11. P. 439–447. doi: 10.2307/1426848.
6. Whitt W. Counterexamples for comparisons of queues with finite waiting rooms // Queueing Syst., 1992. Vol. 10. P. 271–278. doi: 10.1007/BF01159210.
7. Wei S.X., Coyle E.J., Hsiao M.-T.T. An optimal buffer management policy for high-performance packet switching // Global Telecommunications Conference “Countdown to the New Millennium” Proceedings. — Phoenix, AZ, USA: IEEE, 1991. Vol. 2. P. 924–928. doi: 10.1109/GLOCOM.1991.188515.
8. Cidon I., Georgiadis L., Guhrin R., Khamisy A. Optimal buffer sharing // IEEE J. Sel. Area. Comm., 1995. Vol. 13. No. 7. P. 1229–1240. doi: 10.1109/49.414642.
9. Choudhury A.K., Hahne E.L. Dynamic queue length thresholds for shared-memory packet switches // IEEE ACM T. Network., 1998. Vol. 6. No. 2. P. 130–140.
10. Ziya S. On the relationships among traffic load, capacity, and throughput for the  $M/M/1/m$ ,  $M/G/1/m$ -PS, and  $M/G/c/c$  queues // IEEE T. Automat. Contr., 2008. Vol. 53. P. 2696–2701. doi: 10.1109/TAC.2008.2007173.
11. Линец Г. И. Управление объемом буферной памяти и пропускной способностью каналов в мультисервисных сетях // Инфокоммуникационные технологии, 2008. Т. 6. № 2. С. 62–64.
12. Жерновский Ю. В. Решение задач оптимального синтеза для некоторых марковских моделей обслуживания //

- Информационные процессы, 2010. Т. 10. № 3. С. 257–274.
13. Михеев П. А. Анализ стратегий разделения конечной буферной памяти маршрутизатора между выходными каналами // Автоматика и телемеханика, 2014. № 10. С. 125–128.
  14. Агаларов Я. М. Оптимизация объема буферной памяти узла коммутации при схеме полного разделения памяти // Информатика и её применения, 2018. Т. 12. Вып. 4. С. 25–32. doi: 10.14357/19922264180404.
  15. Apostolaki M., Vanbever L., Ghobadi M. FAB: Toward flow-aware buffer sharing on programmable switches // ACM Workshop on Buffer Sizing, 2019. 6 p. doi: 10.1145/3375235.3375237. [https://people.csail.mit.edu/ghobadi/papers/fab\\_buffer\\_2019.pdf](https://people.csail.mit.edu/ghobadi/papers/fab_buffer_2019.pdf).
  16. Kim K. Numerical study of optimal buffer size and vacation length in  $M/G/1/K$  queues with multiple vacations // Int. J. Engineering Technologies Management Research, 2019. Vol. 6. No. 2. P. 1–13. doi: 10.29121/ijetmr.v6.i2.2019.350.
  17. Агаларов Я. М. Об оптимизации работы резервного прибора в многолинейной системе массового обслуживания // Информатика и её применения, 2023. Т. 17. Вып. 1. С. 89–95. doi: 10.14357/19922264230112.
  18. Башарин Г. П., Самуйлов К. Е. Об оптимальной структуре буферной памяти в сетях передачи данных с коммутацией пакетов. — М., 1982. Препринт АН СССР. 70 с.
  19. Агаларов Я. М. Признак унимодальности целочисленной функции одной переменной // Обзорение прикладной и промышленной математики, 2019. Т. 26. Вып. 1. С. 65–66.

Поступила в редакцию 01.06.23

## OPTIMIZATION OF THE BUFFER MEMORY ALLOCATION SCHEME OF THE PACKET SWITCHING NODE

Ya. M. Agalarov

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The buffer of the packet switching node shared by several output communication lines is considered. Sharing buffer memory by multiple users reduces the amount of memory needed to meet latency requirements and the likelihood of packet loss. However, there is a problem of allocating buffer memory between different users, since individual users, having occupied all the memory, can restrict (or close) access to communication lines to other users which can significantly reduce the performance of the switching node as a whole. There are many different buffer memory allocation schemes, one of which, called SMA (Sharing with Minimum Allocation), is being investigated in this paper in order to reduce the costs associated with packet rejection and delay and the operation of the drive and communication lines. A multithreaded queuing system with parallel devices of the  $M/M/s/K$  type with a buffer shared according to the SMA scheme with a fixed number of storage locations reserved for each device is used as a model of the switching node. The mathematical formulation of the problem of optimizing the SMA scheme in terms of the volume of publicly accessible buffer locations is formulated in order to minimize system losses arising from rejection of applications, delay of applications in the queue, and operation of the buffer and devices. The theorem on the boundaries of the domain containing the point of the global optimum is proved. A number of statements are also given which are the consequences of the theorem about the point of the global optimum of the objective function for other models of the switching node and special cases of SMA.

**Keywords:** switching node; buffer memory allocation; optimization; queuing system

**DOI:** 10.14357/19922264230306

**EDN:** QLXCKV

## References

1. Kleinrock, L. 1976. *Queueing systems*. New York, NY: Wiley. Vol. 2. 549 p.
2. Irland, M. 1978. Buffer management in a packet switch. *IEEE T. Commun.* 26(3):328–337. doi: 10.1109/TCOM.1978.1094076.
3. Kamoun, F., and L. Kleinrock. 1980. Analysis of shared finite storage in a computer networks node environment under general traffic conditions. *IEEE T. Commun.* 28(7):992–1003. doi: 10.1109/TCOM.1980.1094756.
4. Köchel, P. 2004. Finite queueing systems — structural investigations and optimal design. *Int. J. Prod. Econ.* 88(2):157–171. doi: 10.1016/j.ijpe.2003.11.005.
5. Sonderman, D. 1979. Comparing multi-server queues with finite waiting rooms, I: Same number of servers. *Adv. Appl. Probab.* 11:439–447. doi: 10.2307/1426848.

6. Whitt, W. 1992. Counterexamples for comparisons of queues with finite waiting rooms. *Queueing Syst.* 10:271–278. doi: 10.1007/BF01159210.
7. Wei, S. X., E. J. Coyle, and M.-T. T. Hsiao. 1991. An optimal buffer management policy for high-performance packet switching. *Global Telecommunications Conference “Countdown to the New Millennium” Proceedings*. Phoenix, AZ: IEEE. 2:924–928. doi: 10.1109/GLOCOM.1991.188515.
8. Cidon, I., L. Georgiadis, R. Gubrin, and A. Khamisy. 1995. Optimal buffer sharing. *IEEE J. Sel. Area. Comm.* 13(7):1229–1240. doi: 10.1109/49.414642.
9. Choudhury, A. K., and E. L. Hahne. 1998. Dynamic queue length thresholds for shared-memory packet switches. *IEEE ACM T. Network.* 6(2):130–140. doi: 10.1109/90.664262.
10. Ziya, S. 2008. On the relationships among traffic load, capacity, and throughput for the  $M/M/1/m$ ,  $M/G/1/m$ -PS, and  $M/G/c/c$  queues. *IEEE T. Automat. Contr.* 53(11):2696–2701. doi: 10.1109/TAC.2008.2007173.
11. Linets, G. I. 2010. Upravlenie ob’emom bufernoy pamyati i propusknoy sposobnost’yu kanalov v mul’tiservisnykh setyakh [Volume management of buffer memory and throughput of channels in multiservice networks]. *Informatsionnye tekhnologii* [Information Communication Technologies] 6(2):62–64.
12. Zhernovyy, Yu. V. 2010. Reshenie zadach optimal’nogo sinteza dlya nekotorykh markovskikh modeley obsluzhivaniya [Solution of optimum synthesis problem for some Markov models of service]. *Informatsionnye protsessy* [Information Processes] 10(3):257–274.
13. Mikheev, P. A. 2014. Analyzing sharing strategies for finite buffer memory in a router among outgoing channels. *Automat. Rem. Contr.* 75(10):1814–1825. doi: 10.1134/S0005117914100087.
14. Agalarov, Ya. M. 2018. Optimizatsiya ob’ema bufernoy pamyati uzla kommutatsii pri skheme polnogo razdeleniya pamyati [Optimization of buffer memory size of switching node in mode of full memory sharing]. *Informatika i ee Primeneniya — Inform. Appl.* 12(4):25–32. doi: 10.14357/19922264180404.
15. Apostolaki, M., L. Vanbever, and M. Ghobadi. 2019. FAB: Toward flow-aware buffer sharing on programmable switches. *ACM Workshop on Buffer Sizing*. 6 p. doi: 10.1145/3375235.3375237. Available at: [https://people.csail.mit.edu/ghobadi/papers/fab\\_buffer\\_2019.pdf](https://people.csail.mit.edu/ghobadi/papers/fab_buffer_2019.pdf) (accessed July 12, 2023).
16. Kim, K. 2019. Numerical study of optimal buffer size and vacation length in  $M/G/1/K$  queues with multiple vacations. *Int. J. Engineering Technologies Management Research* 6(2):1–13. doi: 10.29121/ijetmr.v6.i2.2019.350.
17. Agalarov, Ya. M. 2023. Ob optimizatsii raboty rezervnogo pribora v mnogolineynoy sisteme massovogo obsluzhivaniya [Optimization of a queue-length dependent additional server in the multiserver queue]. *Informatika i ee Primeneniya — Inform. Appl.* 17(1):89–95. doi: 10.14357/19922264230112.
18. Basharin, G. P., and K. E. Samuilov. 1982. *Ob optimal’noy strukture bufernoy pamyati v setyakh peredachi dannykh s kommutatsiyey paketov* [On the optimal structure of buffer memory in data transmission networks with packet commutation]. Moscow. USSR AS Preprint. 70 p.
19. Agalarov, Ya. M. 2019. Priznak unimodal’nosti tselochislennoy funktsii odnoy peremennoy [A sign of unimodality of an integer function of one variable]. *Obozrenie prikladnoy i promyshlennoy matematiki* [Surveys Applied and Industrial Mathematics] 26(1):65–66.

*Received June 1, 2023*

## Contributor

**Agalarov Yaver M.** (b. 1952) — Candidate of Science (PhD) in technology, associate professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; [aglar@yandex.ru](mailto:aglar@yandex.ru)



# О СКОРОСТИ СХОДИМОСТИ И ПРЕДЕЛЬНЫХ ХАРАКТЕРИСТИКАХ ДЛЯ ОДНОГО ОБОБЩЕННОГО ПРОЦЕССА РОЖДЕНИЯ И ГИБЕЛИ

И. А. Усов<sup>1</sup>, Я. А. Сатин<sup>2</sup>, А. И. Зейфман<sup>3</sup>

**Аннотация:** Рассмотрена система обслуживания с одним сервером и разными вариантами ремонта и отказов, число требований в которой описывается неоднородным обобщенным процессом рождения и гибели (ПРГ) (т.е. процессом, в котором интенсивности переходов не константы, а функции времени). Для обоснованного нахождения предельных вероятностных характеристик системы изучается скорость сходимости к ним (т.е. скорость, с которой «забываются» начальные условия системы). Для исследования скорости сходимости к предельному режиму применен недавно разработанный вариант подхода, основанного на понятии логарифмической нормы операторной функции, соответствующей оценке нормы матрицы Коши, а также модернизированного специального преобразования прямой системы Колмогорова. Рассмотрен численный пример, в котором детально показано получение оценок скорости сходимости и основанное на этих оценках построение некоторых предельных характеристик модели.

**Ключевые слова:** обобщенный процесс рождения и гибели; скорость сходимости; эргодичность; логарифмическая норма; системы массового обслуживания

**DOI:** 10.14357/19922264230307

**EDN:** RBKZJ1

## 1 Введение

Изучение обобщенных ПРГ играет важную роль в сложных телекоммуникационных системах, биологии и радиотехнике. В современной разработке программного обеспечения все чаще применяется микросервисная архитектура и облачная инфраструктура, где прием и обслуживание запросов имеют свои особенности и очереди.

Обобщенные ПРГ имеют различные приложения, например при моделировании систем сборки на заказ [1], производственных линий [2], беспроводных сетей [3]. Ранее многие исследователи изучали многосерверные системы с отказами и ремонтами [4–8], а проблемами теории надежности и исследованием некоторых систем массового обслуживания (СМО) активно занимались отечественные авторы (см., например, работы [9–11]).

В [12] была рассмотрена модель с различными вариантами ремонта и использован матрично-аналитический метод для получения стационарной вероятности и исследования различных показателей эффективности системы.

При исследовании СМО, число требований в которых описывается неоднородной марковской

цепью с непрерывным временем, очень важную роль играют предельные вероятностные характеристики (в однородном случае этому соответствует стационарное распределение). Для обоснованного их построения необходимо оценить скорость сходимости к предельному режиму. Основные методы, позволяющие получить такие оценки, описаны в [13]. Однако для обобщенных ПРГ эти методы применить не удастся.

В данной статье рассмотрен метод, основанный на понятии логарифмической нормы и использовании матрицы  $C(t)$  для перехода к редуцированной прямой системе Колмогорова в форме

$$\frac{dz(t)}{dt} = W(t)z(t)$$

при  $t \geq 0$  (подробнее см. в разд. 4). Данный метод без подробностей был описан в работе [14], а в работе [15] была впервые показана возможность его применения к обобщенному ПРГ, описывающему один класс суперкомпьютерных систем. Подробное описание данного подхода было сделано в [16]. В недавней работе [17] с помощью этого метода была исследована модель массового обслуживания РН/М/1 с одним сервером.

<sup>1</sup>Вологодский государственный университет, iusov35@yandex.ru

<sup>2</sup>Вологодский государственный университет, yasovi@mail.ru

<sup>3</sup>Вологодский государственный университет; Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Вологодский научный центр Российской академии наук; Московский центр фундаментальной и прикладной математики; Московский государственный университет имени М. В. Ломоносова, a\_zeifman@mail.ru

Модель, исследуемая в текущей работе, в однородном случае была введена и изучена в [18]. С помощью метода Лапласа–Стилтьеса были получены оценки для различных показателей надежности системы и исследованы частные случаи.

В настоящей работе исследован нестационарный случай (т.е. ситуация, в которой интенсивности процесса, описывающего число требований в системе, зависят от времени), получена оценка скорости сходимости к предельному режиму и рассмотрен численный пример.

## 2 Основные понятия

Обозначим через  $\|\cdot\|$   $l_1$ -норму вектора,  $\|\mathbf{x}\| = \sum |x_i|$ ,  $\|H\| = \sup_j \sum_{i \geq 0} |h_{ij}|$ , если  $H = (h_{ij})_{i,j=0}^\infty$ , и обозначим через  $\Omega$  множество всех векторов из  $l_1$  с неотрицательными координатами и единичной нормой. Пусть  $A(t) = (a_{ij}(t))_{i,j=0}^\infty$  — транспонированная матрица интенсивностей переходов некоторого марковского процесса, причем  $|\alpha_{ii}| < L < \infty$  при всех  $i$  и почти всех  $t \geq 0$ . Тогда имеем

$$\|A(t)\| = 2 \sup_i |a_{ii}(t)| \leq 2L$$

почти для всех  $t \geq 0$ .

Можем рассматривать прямую систему Колмогорова

$$\frac{d\mathbf{p}(t)}{dt} = A(t)\mathbf{p}(t)$$

как дифференциальное уравнение в пространстве последовательностей  $l_1$ , где  $A(t)$  — ограниченный почти для всех  $t \geq 0$  линейный оператор из  $l_1$  в себя.

Напомним, что марковская цепь  $X(t)$  называется *слабо эргодичной*, если

$$\lim_{t \rightarrow \infty} \|\mathbf{p}^1(t) - \mathbf{p}^2(t)\| = 0$$

для любых начальных условий

$$\mathbf{p}^1(0) = \mathbf{p}^1 \in \Omega, \quad \mathbf{p}^2(0) = \mathbf{p}^2 \in \Omega,$$

где  $\mathbf{p}^1(t)$  и  $\mathbf{p}^2(t)$  — это решения прямой системы Колмогорова, описывающей некоторый марковский процесс.

Напомним также, что логарифмическая норма операторной функции в  $l_1$  вычисляется по формуле:

$$\gamma(H(t))_1 = \sup_i \left( h_{ii}(t) + \sum_{j \neq i} |h_{ji}(t)| \right).$$

Оператор Коши соответствующего дифференциального уравнения

$$\frac{d\mathbf{x}(t)}{dt} = H(t)\mathbf{x}(t)$$

имеет вид:

$$U(t, s) = I + \int_s^t H(t_1) dt_1 + \int_s^t H(t_1) \int_s^{t_1} H(t_2) dt_2 dt_1 + \dots,$$

и справедлива следующая оценка (подробнее см. в [19]):

$$\|U(t, s)\| \leq e^{\int_s^t \gamma(H(\tau)) d\tau}.$$

## 3 Описание модели

Рассмотрим классическую СМО с конечным числом источников, т.е. из  $N$  приборов и одного сервера. Каждый прибор подвержен поломкам по независимому пуассоновскому процессу с интенсивностью  $\lambda(t)$ . Когда прибор ломается, заявка немедленно обслуживается сервером, если сервер доступен. В противном случае аварийные приборы должны ждать в очереди на обслуживание сервером.

Будем предполагать, что время обслуживания для аварийного прибора имеет экспоненциальное распределение с интенсивностью  $\mu(t)$ . Сервер может выйти из строя в любой момент, время до отказа имеет экспоненциальное распределение с разными вариантами отказов:  $\zeta_1(t)$  — в простое и  $\zeta_2(t)$  — во время занятости. Когда сервер выходит из строя, он отправляется в ремонт, а время ремонта имеет экспоненциальное распределение;  $\eta_1(t)$  — интенсивность ремонта, когда все приборы исправны, и  $\eta_2(t)$  — когда хотя бы один прибор вышел из строя. Все интенсивности предполагаются неотрицательными, локально интегрируемыми на  $[0, \infty)$ , детерминированными функциями времени.

Рассмотрим двумерный марковский процесс  $(X(t), Y(t))$ , где  $X(t)$  — число доступных серверов в момент времени  $t$ ,  $Y(t)$  — число аварийных приборов в момент времени  $t$ , его пространство состояний есть  $\{(i, j) | i = 0, 1; j = 0, 1, 2, \dots, N\}$ .

Обозначим через  $p_{i,j}(t)$  вероятность того, что система находится в состоянии  $(i, j)$  при  $t \geq 0$ . Тогда

$$\frac{dp_{0,0}(t)}{dt} = p_{1,0}(t)\zeta_1(t) - p_{0,0}(t)(\eta_1(t) + N\lambda(t));$$

$$\frac{dp_{1,0}(t)}{dt} = p_{0,0}(t)\eta_1(t) - p_{1,0}(t)(\zeta_1(t) + N\lambda(t)) + p_{1,1}(t)\mu(t);$$

$$\begin{aligned} \frac{dp_{0,j}(t)}{dt} &= p_{1,j}(t)\zeta_2(t) - p_{0,j}(t)(\eta_2(t) + \\ &+ (N-j)\lambda(t)) + p_{0,j-1}(t)(N-j+1)\lambda(t), \\ &0 < j < N; \end{aligned}$$

$$\begin{aligned} \frac{dp_{1,j}(t)}{dt} &= p_{0,j}(t)\eta_2(t) - p_{0,j}(t)(\zeta_2(t) + \\ &+ (N-j)\lambda(t)) + p_{1,j-1}(t)(N-j+1)\lambda(t) + \\ &+ p_{1,j+1}(t)\mu(t), \quad 0 < j < N; \\ \frac{dp_{0,N}(t)}{dt} &= p_{1,N}(t)\zeta_2(t) - p_{0,N}(t)\eta_2(t) + \\ &+ p_{0,N-1}(t)\lambda(t); \\ \frac{dp_{1,N}(t)}{dt} &= p_{0,N}(t)\eta_2(t) - p_{1,N}(t)(\zeta_2(t) + \mu(t)) + \\ &+ p_{0,N-1}(t)\lambda(t). \end{aligned}$$

Будем считать, что состояния имеют следующий порядок:  $\{(0, 0), (1, 0), (0, 1), (1, 1), \dots, (0, N), (1, N)\}$ , а соответствующий одномерный процесс обозначим  $\chi(t)$ . Тогда  $\mathbf{p}(t) = (p_{0,0}(t), p_{1,0}(t), p_{0,1}(t), p_{1,1}(t), \dots, p_{0,N}(t), p_{1,N}(t))^T$  — вектор распределения вероятностей состояний  $\chi(t)$  в момент времени  $t$ , а его матрица интенсивностей  $Q(t)$  представима в следующем виде:

$$Q = \begin{pmatrix} A_0(t) & C_0(t) & 0 & \dots & 0 \\ B_1(t) & A_1(t) & C_1(t) & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & B_{N-1}(t) & A_{N-1}(t) & C_{N-1}(t) \\ 0 & \dots & 0 & B_N(t) & A_N(t) \end{pmatrix},$$

где

$$B_1(t) = \dots = B_N(t) = \begin{pmatrix} 0 & 0 \\ 0 & \mu(t) \end{pmatrix};$$

$$C_j(t) = \begin{pmatrix} (N-j)\lambda(t) & 0 \\ 0 & (N-j)\lambda(t) \end{pmatrix}, \quad j = 0, 1, \dots, N-1;$$

$$A_0(t) = \begin{pmatrix} -\eta_1(t) - N\lambda(t) & \eta_1(t) \\ -\zeta_1(t) & -\zeta_1(t) - N\lambda(t) \end{pmatrix};$$

$$A_j(t) = \begin{pmatrix} -\eta_2(t) - (N-j)\lambda(t) & \eta_2(t) \\ -\zeta_2(t) & -\zeta_2(t) - \mu(t) - (N-j)\lambda(t) \end{pmatrix}, \quad j = 0, 1, \dots, N-1;$$

$$A_N(t) = \begin{pmatrix} -\eta_2(t) & \eta_2(t) \\ \zeta_2(t) & -\zeta_2(t) - \mu(t) \end{pmatrix}.$$

Процесс  $\chi(t)$  описывается прямой системой Колмогорова

$$\frac{d\mathbf{p}(t)}{dt} = A(t)\mathbf{p}(t), \quad (1)$$

где матрица  $A(t) = Q^T(t)$ .

## 4 Оценки скорости сходимости

Для получения оценки скорости сходимости к предельному режиму будем использовать подход из [16]. Пусть  $\mathbf{p}^*(t)$  и  $\mathbf{p}^{**}(t)$  — решения прямой системы Колмогорова (1) с соответствующими (различными) начальными условиями  $\mathbf{p}^*(0) \in \Omega$  и  $\mathbf{p}^{**}(0) \in \Omega$ . Тогда их разность  $\mathbf{z}(t) = \mathbf{p}^*(t) - \mathbf{p}^{**}(t)$  также будет решением системы (1), но при этом сумма всех координат вектора  $\mathbf{z}(t)$  равна нулю при всех  $t \geq 0$ ,  $\sum_{i \geq 0} z_i(t) = 0$ . Пусть  $c(t)$  — некоторая положительная при всех  $t \geq 0$  функция. Вычтем из правой части уравнения

$$z'_0(t) = \sum_{j \geq 0} a_{0j}(t)z_j(t)$$

выражение

$$c(t) \sum_{j \geq 0} z_j(t) = 0.$$

Запишем получившуюся при этом систему в виде:

$$\frac{d\mathbf{z}(t)}{dt} = W(t)\mathbf{z}(t), \quad t \geq 0,$$

где  $W(t) = A(t) - C(t)$  и у матрицы  $C(t)$  нулевая строка состоит из элементов  $c(t)$ , а все остальные строки состоят из нулей.

Рассмотрим последовательность положительных чисел  $d_k$  при  $0 \leq k \leq K$  и диагональную матрицу

$$D = \text{diag}(d_0, d_1, d_2, \dots, d_K),$$

где  $K = 2N + 1$ . Обозначим

$$\|\mathbf{z}(t)\|_{1D} = \|D\mathbf{z}(t)\|_1.$$

**Теорема 1.** Пусть существует последовательность положительных чисел  $d_k$ ,  $0 \leq k \leq K$ , и функция  $c(t)$  при  $t > 0$ , такая что  $c(t) \leq \zeta_1(t)$ . Тогда процесс, описывающий число требований в системе, слабо эргодичен и справедливо следующее неравенство:

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \leq \frac{2 \max_k d_k}{\min_k d_k} e^{-\int_0^t \beta_*(\tau) d\tau}$$

для любых начальных условий  $\mathbf{p}^*(0), \mathbf{p}^{**}(0)$  и  $\beta_*(t) = \min_k a_k(t)$ , если  $a_k(t) > 0$ ,  $0 \leq k \leq K$ .

**Доказательство.** Положим

$$\begin{aligned} \gamma(W(t))_{1D} &= \gamma(DW(t)D^{-1}) = \\ &= \min_i \left( |w_{ii}(t)| + \sum_{j \neq i} \frac{d_j}{d_i} |w_{ji}(t)| \right) = -\beta_*(t), \end{aligned}$$

где матрицу  $DW(t)D^{-1}$  можно записать в следующем виде:

$$DW(t)D^{-1} =$$

$$= \begin{pmatrix} -N\lambda(t) - \eta_1(t) - c(t) \frac{d_0}{d_1} (\zeta_1(t) - c(t)) & -\frac{d_0}{d_2} c(t) & & -\frac{d_0}{d_3} c(t) \\ \frac{d_1}{d_0} \eta_1(t) & -N\lambda(t) - \zeta_1(t) & 0 & \frac{d_1}{d_3} \mu(t) \\ \frac{d_2}{d_0} N\lambda(t) & 0 & -\lambda(t)(N-1) - \eta_2(t) & \frac{d_2}{d_3} \zeta_2(t) \\ 0 & \frac{d_3}{d_1} N\lambda(t) & \frac{d_3}{d_2} \eta_2(t) & -\mu(t) - \lambda(t)(N-1) - \zeta_2(t) \\ 0 & 0 & \frac{d_4}{d_2} \lambda(t)(N-1) & 0 \\ [10.0pt] 0 & 0 & 0 & \frac{d_5}{d_3} \lambda(t)(N-1) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \\ & -\frac{d_0}{d_4} c(t) & -\frac{d_0}{d_5} c(t) & \dots -\frac{d_0}{d_{K-1}} c(t) -\frac{d_0}{d_K} c(t) \\ & 0 & 0 & \dots 0 0 \\ & 0 & 0 & \dots 0 0 \\ & 0 & \frac{d_3}{d_5} \mu(t) & \dots 0 0 \\ -\lambda(t)(N-2) - \eta_2(t) & \frac{d_4}{d_5} \zeta_2(t) & & \dots 0 0 \\ \frac{d_5}{d_4} \eta_2(t) & -\mu(t) - \lambda(t)(N-2) - \zeta_2(t) & \dots & 0 0 \\ \frac{d_6}{d_4} \lambda(t)(N-2) & 0 & \dots & 0 0 \\ 0 & \frac{d_7}{d_5} \lambda(t)(N-2) & \dots & 0 0 \\ 0 & 0 & \dots & 0 \frac{d_{K-2}}{d_K} \mu \\ \vdots & \vdots & \ddots & -\eta_2 \frac{d_{K-1}}{d_K} \zeta_2(t) \\ 0 & 0 & \dots & \frac{d_K}{d_{K-1}} \eta_2 -\mu - \zeta_2(t) \end{pmatrix} ;$$

$$\beta_*(t) = \min_k \alpha_k, \text{ где } \alpha_k(t) = \begin{cases} \left(1 - \frac{d_2}{d_0}\right) N\lambda(t) + \left(1 - \frac{d_1}{d_0}\right) \eta_1(t) + c(t), & k = 0; \\ \left(1 - \frac{d_3}{d_1}\right) N\lambda(t) + \left(1 - \frac{d_0}{d_1}\right) \zeta_1(t) + c(t), & k = 1; \\ \left(1 - \frac{d_{k+2}}{d_k}\right) \left(N - \frac{k}{2}\right) \lambda(t) + \\ + \left(1 - \frac{d_{k+1}}{d_k}\right) \eta_2(t) - c(t) \frac{d_0}{d_k}, & k - \text{четное}, 2 \leq k \leq 2N; \\ \left(1 - \frac{d_{k+2}}{d_k}\right) \left(N - \frac{k-1}{2}\right) \lambda(t) + \\ + \left(1 - \frac{d_{k-2}}{d_k}\right) \mu(t) + \left(1 - \frac{d_{k-1}}{d_k}\right) \zeta_2(t) - \\ - c(t) \frac{d_0}{d_k}, & k - \text{нечетное}, 3 \leq k \leq 2N + 1. \end{cases} \quad (2)$$

Выберем положительные числа  $a$  и  $b$  такие, что  $a > b$ , и последовательность  $d_k$  при  $0 \leq k \leq K$  следующим образом:

$$d_k = \begin{cases} 1, & k = 0; \\ \frac{k}{2} + a, & k - \text{четное}, 2 \leq k \leq 2N, a > 0; \\ \frac{k-1}{2} + b, & k - \text{нечетное}, 1 \leq k \leq 2N + 1, \\ & b > 0. \end{cases}$$

Подставляем данную последовательность в (2) и получаем:

$$\alpha_k(t) = \begin{cases} \left(1 - (1+a)\right) N\lambda(t) + (1-b) \eta_1(t) + c(t), & k = 0; \\ \left(1 - \frac{1+b}{b}\right) N\lambda(t) + \left(1 - \frac{1}{b}\right) \zeta_1(t) + c(t), & k = 1; \\ \left(1 - \frac{k+2a+2}{k+2a}\right) \left(N - \frac{k}{2}\right) \lambda(t) + \\ + \left(1 - \frac{k+2b}{k+2a}\right) \eta_2(t) - c(t) \frac{2}{k+2a}, & k - \text{четное}, 2 \leq k \leq 2N; \\ \left(1 - \frac{k+2b+1}{k+2b-1}\right) \left(N - \frac{k-1}{2}\right) \lambda(t) + \\ + \left(1 - \frac{k+2b-3}{k+2b-1}\right) \mu(t) + \\ + \left(1 - \frac{k+2a-1}{k+2b-1}\right) \zeta_2(t) - c(t) \frac{2}{k+2b-1}, & k - \text{нечетное}, 3 \leq k \leq 2N + 1. \end{cases} \quad (3)$$

Заметим, что  $\|D\| = \max_k d_k$ ;  $\|D^{-1}\| = 1/(\min_k d_k)$ . Таким образом, получаем

$$\|\mathbf{z}(t)\|_{1D} \leq \max_k d_k \|\mathbf{z}(t)\|_1;$$

$$\|\mathbf{z}(t)\|_1 \leq \frac{1}{\min_k d_k} \|\mathbf{z}(t)\|_{1D};$$

$$\|\mathbf{z}^*(t) - \mathbf{z}^{**}(t)\|_{1D} \leq e^{-\int_0^t \beta_*(\tau) d\tau} \|\mathbf{z}^*(0) - \mathbf{z}^{**}(0)\|_{1D},$$

откуда и следует утверждение теоремы.

## 5 Численный пример

В качестве иллюстрации полученных результатов рассмотрим получение оценки скорости сходимости, позволяющей провести построение предельных вероятностных характеристик в случае периодических интенсивностей (с периодом 1).

В данном примере будем полагать, что  $a > b$  и  $a = 4, b = 3, N = 25, K = 51$ , а интенсивности имеют вид:

$$\begin{aligned} \lambda(t) &= 0,3 + 0,1 \sin(2\pi t); \\ \mu(t) &= 850 + 4,5 \cos(2\pi t); \\ \zeta_1(t) &= 75 + 0,1 \cos(2\pi t); \\ \zeta_2(t) &= 1 + 0,1 \cos(2\pi t); \\ \eta_1(t) &= 6 + \sin(2\pi t); \\ \eta_2(t) &= 750 + \cos(2\pi t); \\ c(t) &= \zeta_1(t). \end{aligned}$$

Последовательность  $d_k$  выберем следующим образом:

$$d_k(t) = \begin{cases} 1, & k = 0; \\ \frac{k}{2} + 4, & k - \text{четное}, 2 \leq k \leq 2N; \\ \frac{k-1}{2} + 3, & k - \text{нечетное}, 1 \leq k \leq 2N + 1. \end{cases}$$

Теперь подставим последовательность  $d_k$ , числа  $N, a, b$  и заданные функции интенсивностей в (3) и получим:

$$\alpha_k(t) = \begin{cases} -4N\lambda(t) - 2\eta_1(t) + c(t), & k = 0; \\ \left(1 - \frac{4}{3}\right) N\lambda(t) + \left(1 - \frac{1}{3}\right) \zeta_1(t) + c(t), & k = 1; \\ \left(1 - \frac{k+10}{k+8}\right) \left(N - \frac{k}{2}\right) \lambda(t) + \\ + \left(1 - \frac{k+6}{k+8}\right) \eta_2(t) - c(t) \frac{2}{k+8}, & k - \text{четное}, 2 \leq k \leq 2N; \\ \left(1 - \frac{k+7}{k+5}\right) \left(N - \frac{k-1}{2}\right) \lambda(t) + \\ + \left(1 - \frac{k+3}{k+5}\right) \mu(t) + \left(1 - \frac{k+7}{k+5}\right) \zeta_2(t) - \\ - c(t) \frac{2}{k+5}, & k - \text{нечетное}, \\ & 3 \leq k \leq 2N + 1, \end{cases} \geq$$

$$\geq \begin{cases} -4N\lambda(t) - 2\eta_1(t) + c(t), & k = 0; \\ -\frac{1}{3}N\lambda(t) + \frac{2}{3}\zeta_1(t) + c(t), & k = 1; \\ \frac{1}{5}(N-1)\lambda(t) + \left(1 - \frac{2N+6}{2N+8}\right)\eta_2(t) - \\ \quad - \frac{1}{5}c(t), & k - \text{четное}, 2 \leq k \leq 2N; \\ \frac{1}{4}(N-1)\lambda(t) + \left(1 - \frac{2N+4}{2N+6}\right)\mu(t) - \\ \quad - \frac{1}{4}\zeta_2(t) - \frac{1}{4}c(t), & k - \text{нечетное}, \\ & 3 \leq k \leq 2N+1. \end{cases}$$

Таким образом, окончательно получаем:

$$\alpha_k(t) = \begin{cases} 20,99, & k = 0; \\ 121,65, & k = 1; \\ 8,941, & k - \text{четное}, 2 \leq k \leq 2N; \\ 8,947, & k - \text{нечетное}, 2 \leq k \leq 2N+1; \end{cases}$$

$$\beta_*(t) = \min_k(\alpha_k) = 8,941.$$

Из теоремы 1 следует оценка скорости сходимости:

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \leq 58e^{-8,941t}. \quad (4)$$

Далее, применяя метод Рунге–Кутты четвертого порядка, численно решаем прямую систему Колмогорова, описывающую данный марковский процесс, с заданными функциями интенсивностей и строим следующие графики:

- на рис. 1, а представлен график вероятности пустой СМО при  $\chi(0) = 0$  (1) и 52 (2) и  $t \in [0, 4]$ . Заметим, что кривые на графике сходятся достаточно быстро;
- на рис. 1, б проиллюстрировано поведение кривых (вероятности пустой СМО) на  $t \in [4, 5]$ ;
- на рис. 2, а представлен график вероятности  $p_1(t)$  при  $\chi(0) = 0$  (1) и 52 (2) и  $t \in [0, 4]$ . Также заметим, что кривые на графике сходятся достаточно быстро;
- на рис. 2, б проиллюстрировано поведение кривых (вероятности  $p_1(t)$ ) на  $t \in [4, 5]$ .

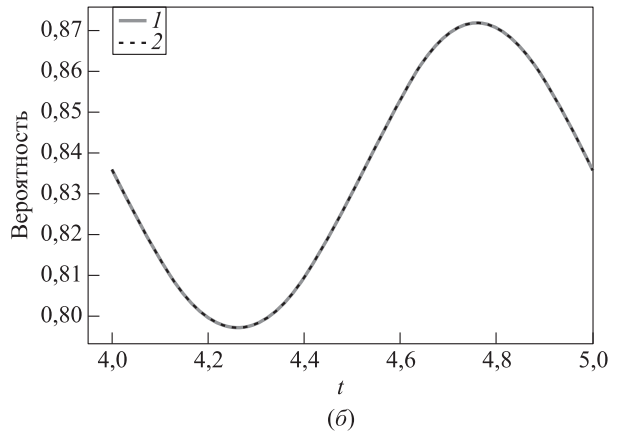
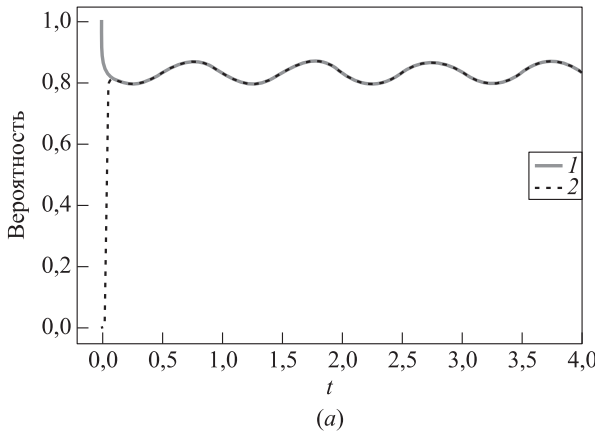


Рис. 1 Вероятность пустой СМО для  $t \in [0, 4]$  (а) и  $t \in [4, 5]$  (б):  $1 - \chi(0) = 0$ ;  $2 - \chi(0) = 52$

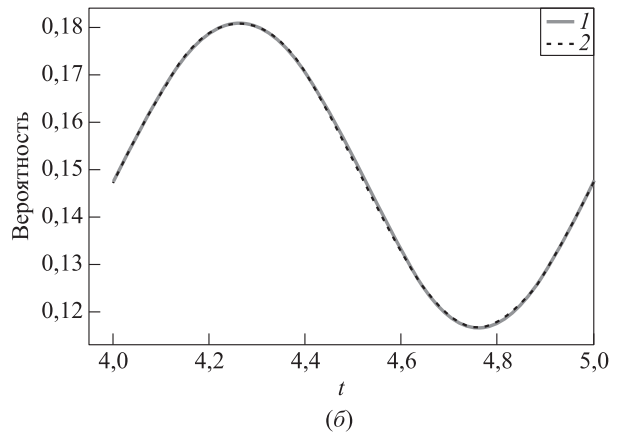
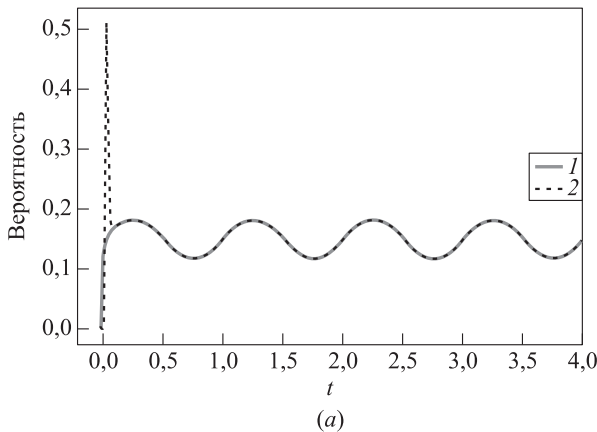


Рис. 2 Вероятность  $p_1(t)$  для  $t \in [0, 4]$  (а) и  $t \in [4, 5]$  (б):  $1 - \chi(0) = 0$ ;  $2 - \chi(0) = 52$

## 6 Заключение

В данной работе сформулирована и доказана теорема, позволяющая получать оценки скорости сходимости к предельному режиму неоднородного обобщенного ПРГ. Для исследования данной модели методы треугольного и диагональных преобразований не применимы, а метод дифференциальных неравенств трудоемок, поэтому был применен метод, основанный на понятии логарифмической нормы операторной функции. Также был рассмотрен численный пример, в котором была получена оценка скорости сходимости и построены соответствующие графики. Точность оценки зависит от выбора последовательности  $d_k$ . Данным методом, по-видимому, можно получать оценки предельных характеристик для моделей  $M/RH/c$  или  $RH/RH/c$ , что может стать предметом дальнейших исследований.

## Литература

1. *Irvani S. M. R., Luangkesorn K. L., Simchi-Levi D.* On assemble to order systems with flexible customers // *IIE Trans.*, 2010. Vol. 35. P. 389–403. doi: 10.1080/07408170390184107.
2. *Fadiloglu M. M., Yeralan S.* Models of production lines as quasi-birth-death processes // *Math. Comput. Model.*, 2002. Vol. 35. P. 913–930. doi: 10.1016/S0895-7177(02)00059-6.
3. *Kim Y. Y., Li S.* Performance evaluation of packet data services over cellular voice networks // *Wirel. Netw.*, 1999. Vol. 5. P. 211–219. doi: 10.1023/A:1019198927563.
4. *Wang K.* Profit analysis of the  $M/M/R$  machine repair problem with spares and server breakdowns // *J. Oper. Res. Soc.*, 1994. Vol. 45. P. 539–548. doi: 10.1057/jors.1994.81.
5. *Jain M., Ghimire R. P.* Machine repair queueing system with non-reliable server and heterogeneous service discipline // *J. M.A.C.T.*, 1997. Vol. 30. P. 105–115.
6. *Jain M., Maheshwari S.* Transient analysis of a redundant system with additional repairmen // *Am. J. Math.* — S., 2003. Vol. 23. P. 347–382. doi: 10.1080/01966324.2003.10737618.
7. *Jain M., Kulshrestha R., Maheshwari S.* N-policy for a machine repair system with spares and renegeing // *Appl. Math. Model.*, 2004. Vol. 28. P. 513–531. doi: 10.1016/j.apm.2003.10.013.
8. *Ke J. C., Wu C. H.* Multi-server machine repair model with standbys and synchronous multiple vacation // *Comput. Ind. Eng.*, 2012. Vol. 62. P. 296–305. doi: 10.1016/j.cie.2011.09.017.
9. *Дудин А. Н., Назаров А. А.* Система обслуживания ММАР/М/Р/0 с резервированием приборов, функционирующая в случайной среде // *Проблемы передачи информации*, 2015. Т. 51. Вып. 3. С. 93–104.
10. *Вишневецкий В. М., Дудин А. Н.* Системы массового обслуживания с коррелированными входными потоками и их применение для моделирования телекоммуникационных сетей // *Автоматика и телемеханика*, 2017. Вып. 8. С. 3–59.
11. *Вишневецкий В. М., Семенова О. В., Буй З. Т.* Исследование системы поллинга с адаптивным циклическим опросом и ее применение для проектирования широкополосных беспроводных сетей // *Проблемы управления*, 2020. Вып. 5. С. 50–55. doi: 10.25728/ru.2020.5.6.
12. *Wu C. H., Ke J. C.* Multi-server machine repair problems under a  $(V, R)$  synchronous single vacation policy // *Appl. Math. Model.*, 2014. Vol. 38. P. 2180–2189. doi: 10.1016/j.apm.2013.10.045.
13. *Zeifman A., Satin Y., Kryukova A., Razumchik R., Kiseleva K., Shilova G.* On three methods for bounding the rate of convergence for some continuous-time Markov chains // *Int. J. Appl. Math. Comp.*, 2020. Vol. 30. P. 251–266. doi: 10.34768/amcs-2020-0020.
14. *Сатин Я. А.* Исследование модели типа  $M_t/M_t/1$  с двумя различными классами требований // *Системы и средства информатики*, 2021. Т. 31. № 1. С. 17–27. doi: 10.14357/08696527210102.
15. *Razumchik R., Rumyantsev A.* Some ergodicity and truncation bounds for a small scale Markovian supercomputer model // *36th ECMS Conference (International) on Modelling and Simulation Proceedings*. — Saarbrücken-Dudweiler, Germany: Digitaldruck Pirrot GmbH, 2022. P. 324–330. doi: 10.7148/2022-0324.
16. *Ковалёв И. А., Сатин Я. А., Сеницина А. В., Зейфман А. И.* Об одном подходе к оцениванию скорости сходимости нестационарных марковских моделей систем обслуживания // *Информатика и её применения*, 2022. Т. 16. № 3. С. 75–82. doi: 10.14357/19922264220310.
17. *Usov I., Satin Y., Zeifman A.* Estimating the rate of convergence of the  $RH/M/1$  model by reducing to quasi-birth-death processes // *Mathematics*, 2023. Vol. 11. No. 6. Art. 1494. doi: 10.3390/math11061494.
18. *Lv S.* Multi-machine repairable system with one unreliable server and variable repair rate // *Mathematics*, 2021. Vol. 9. No. 11. Art. 1299. doi: 10.3390/math9111299.
19. *Зейфман А. И., Королев В. Ю., Коротышева А. В., Сатин Я. А.* Оценки для неоднородных марковских систем обслуживания с особенностями в нуле. — М.: ФИЦ ИУ РАН, 2016. 56 с.

Поступила в редакцию 14.04.23

## ON THE RATE OF CONVERGENCE AND LIMITING CHARACTERISTICS FOR ONE QUASI-BIRTH–DEATH PROCESS

I. A. Usov<sup>1</sup>, Y. A. Satin<sup>1</sup>, and A. I. Zeifman<sup>1,2,3,4</sup>

<sup>1</sup>Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation

<sup>2</sup>Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

<sup>3</sup>Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

<sup>4</sup>Vologda Research Center of the Russian Academy of Sciences, 56A Gorky Str., Vologda 160014, Russian Federation

**Abstract:** A queuing system with one server and different repair and failure options is considered, the number of requirements in which is described by a quasi-birth-death process. To reasonably find the limiting probabilistic characteristics of the system, the rate of convergence to them is studied (that is, the rate at which the initial conditions of the system are “forgotten”). To study the rate of convergence to the limiting regime, a recently developed version of the approach based on the concept of the logarithmic norm of the operator function corresponding to the estimate of the norm of the Cauchy matrix as well as a modernized special transformation of the forward Kolmogorov system was applied. A numerical example is considered for which the estimation of the rate of convergence is shown in detail as well as the construction of some limiting characteristics of the model based on these estimates.

**Keywords:** quasi-birth–death processes; rate of convergence; ergodicity bounds; logarithmic norm; queuing systems

**DOI:** 10.14357/19922264230307

**EDN:** RBKZJI

### References

1. Irvani, S. M. R., K. L. Luangkesorn, and D. Simchi-Levi. 2003. On assemble to order systems with flexible customers. *IIE Trans.* 35(5):389–403. doi: 10.1080/07408170390184107.
2. Fadiloglu, M. M., and S. Yeralan. 2002. Models of production lines as quasi-birth-death processes. *Math. Comput. Model.* 35(7-8):913–930. doi: 10.1016/S0895-7177(02)00059-6.
3. Kim, Y. Y., and S. Li. 1999. Performance evaluation of packet data services over cellular voice networks. *Wirel. Netw.* 5:211–219. doi: 10.1023/A:1019198927563.
4. Wang, K. 1994. Profit analysis of the  $M/M/R$  machine repair problem with spares and server breakdowns. *J. Oper. Res. Soc.* 45:539–548. doi: 10.1057/jors.1994.81.
5. Jain, M., and R. P. Ghimire. 1997. Machine repair queuing system with non-reliable server and heterogeneous service discipline. *J. M.A.C.T.* 30:105–115.
6. Jain, M., and S. Maheshwari. 2003. Transient analysis of a redundant system with additional repairmen. *Am. J. Math.* — S. 23(3-4):347–382. doi: 10.1080/01966324.2003.10737618.
7. Jain, M., R. Kulshrestha, and S. Maheshwari. 2004. N-policy for a machine repair system with spares and reneging. *Appl. Math. Model.* 28(6):513–531. doi: 10.1016/j.apm.2003.10.013.
8. Ke, J. C., and C. H. Wu. 2012. Multi-server machine repair model with standbys and synchronous multiple vacation. *Comput. Ind. Eng.* 62(1):296–305. doi: 10.1016/j.cie.2011.09.017.
9. Dudin, A. N., and A. A. Nazarov. 2015. The MMAP/M/R/0 queueing system with reservation of servers operating in a random environment. *Probl. Inf. Transm.* 51(3):289–298. doi: 10.1134/S0032946015030060.
10. Vishnevskii, V. M., and A. N. Dudin. 2017. Queueing systems with correlated arrival flows and their applications to modeling telecommunication networks. *Automat. Rem. Contr.* 78(8):1361–1403. doi: 10.1134/S000511791708001X.
11. Vishnevskiy, V. M., O. V. Semenova, and Z. T. Buy. 2020. Issledovanie sistemy pollinga s adaptivnym tsiklicheskim oprosom i ee primeneniye dlya proektirovaniya shirokopolosnykh besprovodnykh setey [Investigation of the stochastic polling system and its applications in broadband wireless networks]. *Control Sciences* 5:50–55. doi: 10.25728/pu.2020.5.6.
12. Wu, C. H., and J. C. Ke. 2014. Multi-server machine repair problems under a  $(V, R)$  synchronous single vacation policy. *Appl. Math. Model.* 38:2180–2189. doi: 10.1016/j.apm.2013.10.045.
13. Zeifman, A., Y. Satin, A. Kryukova, R. Razumchik, K. Kiseleva, and G. Shilova. 2020. On the three methods for bounding the rate of convergence for some continuous-time Markov chains. *Int. J. Appl. Math. Comp.* 30(2):251–266. doi: 10.34768/amcs-2020-0020.
14. Satin, Y. A. 2021. Issledovanie modeli tipa  $M_t/M_t/1$  s dvumya razlichnymi klassami trebovaniy [On the bounds



- of the rate of convergence for  $M_t/M_t/1$  model with two different requests]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(1):17–27. doi: 10.14357/08696527210102.
15. Razumchik, R., and A. Rumyantsev. 2022. Some ergodicity and truncation bounds for a small scale Markovian supercomputer model. *36th ECMS Conference (International) on Modelling and Simulation Proceedings*. Saarbrücken-Dudweiler, Germany: Digitaldruck Pirrot GmbH. 324–330. doi: 10.7148/2022-0324.
  16. Kovalev, I. A., Y. A. Satin, A. V. Sinitcina, and A. I. Zeifman. 2022. Ob odnom podkhode k otsenivaniyu skorosti skhodimosti nestatsionarnykh markovskikh modeley sistem obsluzhivaniya [On an approach for estimating the rate of convergence for nonstationary Markov models of queueing systems]. *Informatika i ee Primeneniya — Inform. Appl.* 16(3):75–82. doi: 10.14357/19922264220310.
  17. Usov, I., Y. Satin, and A. Zeifman. 2023. Estimating the rate of convergence of the PH/M/1 model by reducing to quasi-birth–death processes. *Mathematics* 11(6):1494. doi: 10.3390/math11061494.
  18. Lv, S. 2021. Multi-machine repairable system with one unreliable server and variable repair rate. *Mathematics* 9(11):1299. doi: 10.3390/math9111299.
  19. Zeifman, A. I., V. Yu. Korolev, A. V. Korotysheva, and Ya. A. Satin. 2016. *Otsenki dlya neodnorodnykh markovskikh sistem obsluzhivaniya s osobennostyami v nule* [Bounds for inhomogeneous Markov service systems with singularities at zero]. Moscow: FRC CSC RAS. 56 p.

*Received April 14, 2023*

## Contributors

**Usov Ilya A.** (b. 1996) — PhD student, Department of Applied Mathematics, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; iusov35@yandex.ru

**Satin Yacov A.** (b. 1978) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Applied Mathematics, Vologda State University, 15 Lenin Str., Vologda 160000; yacovi@mail.ru

**Zeifman Alexander I.** (b. 1954) — Doctor of Science in physics and mathematics, professor, head of department, Department of Applied Mathematics, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; principal scientist, Vologda Research Center of the Russian Academy of Sciences, 56A Gorky Str., Vologda 160014, Russian Federation; senior scientist, Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; a\_zeifman@mail.ru

# МЕТОД ОЦЕНИВАНИЯ ПАРАМЕТРОВ ГАММА-ЭКСПОНЕНЦИАЛЬНОГО РАСПРЕДЕЛЕНИЯ ПО ВЫБОРКЕ СО СЛАБО ЗАВИСИМЫМИ КОМПОНЕНТАМИ\*

А. А. Кудрявцев<sup>1</sup>, О. В. Шестаков<sup>2</sup>

**Аннотация:** Доказывается асимптотическая нормальность оценок параметров гамма-экспоненциального распределения, полученных при помощи модифицированного метода моментов, в случае слабой зависимости компонент выборки. Для оценок параметров изгиба и масштаба гамма-экспоненциального распределения при фиксированных параметрах формы и концентрации доказана центральная предельная теорема в случае, когда максимальный коэффициент корреляции между элементами выборки стремится к нулю. Метод доказательства основан на исследовании спектральной плотности выборки и результатах теории стационарных случайных последовательностей. Результаты статьи могут быть использованы для обоснования асимптотической нормальности оценок параметров дигамма-распределения, к частным видам которого относятся обобщенное гамма-распределение и обобщенное бета-распределение второго рода, возникающие при описании процессов, для моделирования которых используются распределения с неотрицательным неограниченным носителем.

**Ключевые слова:** слабая зависимость; оценивание параметров; гамма-экспоненциальное распределение; смешанные распределения; метод моментов; асимптотическая нормальность

DOI: 10.14357/19922264230308

EDN: PEXTVK

## 1 Введение

Подавляющее большинство моделей, используемых для описания случайных характеристик, имеющих непрерывные неотрицательные носители, оперируют распределениями из гамма- и бета-классов. При этом необходимо возникает задача статистического оценивания неизвестных параметров применяемого распределения. В статье рассматривается предложенное в [1] распределение, тесно связанное как с обобщенным гамма-распределением [2, 3], так и с обобщенным бета-распределением второго рода [4].

**Определение.** Будем говорить, что случайная величина  $\zeta$  имеет гамма-экспоненциальное распределение  $GE(r, \nu, s, t, \delta)$  с параметрами изгиба  $0 \leq r < 1$ , формы  $\nu \neq 0$ , концентрации  $s, t > 0$  и масштаба  $\delta > 0$ , если ее плотность при  $x > 0$  задается соотношением

$$g_E(x) = \frac{|\nu| x^{t\nu-1}}{\delta^{t\nu} \Gamma(s) \Gamma(t)} Ge_{r, tr+s} \left( - \left( \frac{x}{\delta} \right)^\nu \right), \quad (1)$$

где  $E = (r, \nu, s, t, \delta)$ ;  $Ge_{\alpha, \beta}(x)$  — гамма-экспоненциальная функция [5]:

$$Ge_{\alpha, \beta}(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \Gamma(\alpha k + \beta),$$

$$x \in \mathbb{R}, \quad 0 \leq \alpha < 1, \quad \beta > 0. \quad (2)$$

В работе [1] было показано, что распределение (1) адекватно описывает байесовские модели баланса [6]. Это прежде всего вызвано тем, что распределение с плотностью (1) может быть представлено как масштабная смесь двух случайных величин, имеющих обобщенное гамма-распределение [1].

Проблема оценивания параметров обобщенного гамма-распределения, его частных видов и смесей имеет богатую историю и до сих пор актуальна [7–10].

Ввиду представления плотности  $g_E(x)$  в терминах специальной гамма-экспоненциальной функции (2) метод максимального правдоподобия представляется затруднительным. То же можно сказать и о прямом методе моментов. По этой причине в работе [11] было предложено оценивать параметры гамма-экспоненциального распределения при помощи модифицированного метода, основанного на логарифмических моментах.

\* Работа выполнена при финансовой поддержке РФФИ (проект № 22-11-00212).

<sup>1</sup> Московский государственный университет имени М. В. Ломоносова, факультет вычислительной математики и кибернетики; Московский центр фундаментальной и прикладной математики, aakudryavtsev@cs.msu.ru

<sup>2</sup> Московский государственный университет имени М. В. Ломоносова, факультет вычислительной математики и кибернетики; Московский центр фундаментальной и прикладной математики; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, oshestakov@cs.msu.ru

Кроме того, в ряде задач приходится рассматривать выборку из зависимых компонент. Такие ситуации возникают, например, при анализе различных геофизических процессов и анализе помех в коммуникационных каналах.

Пусть  $\{X_i\}_{i \in \mathbb{Z}}$  — стационарная в строгом смысле последовательность случайных величин, определенных на одном вероятностном пространстве  $(\Omega, \mathfrak{F}, P)$ , и  $\mathfrak{F}_k^m$  —  $\sigma$ -алгебра, порожденная случайными величинами  $X_i$ ,  $k \leq i \leq m$ . Обозначим через  $\mathcal{L}^2(\mathfrak{F}_k^m)$  совокупность всех  $\mathfrak{F}_k^m$ -измеримых случайных величин с конечными вторыми моментами. Определим максимальный коэффициент корреляции:

$$\rho(n) = \sup_{f, g} |\text{corr}(f, g)|, \\ f \in \mathcal{L}^2(\mathfrak{F}_{-\infty}^0), g \in \mathcal{L}^2(\mathfrak{F}_n^\infty), n > 0.$$

Условие  $\rho(n) \rightarrow 0$  при  $n \rightarrow \infty$  означает ослабление зависимости в последовательности  $\{X_i\}_{i \in \mathbb{Z}}$  между «прошлым» и «будущим».

В статье показывается асимптотическая нормальность оценок параметров гамма-экспоненциального распределения, полученных при помощи модифицированного метода моментов. При этом предполагается, что компоненты выборки, по которой строятся оценки, слабо зависимы.

## 2 Структурные характеристики распределения выборки

В дальнейших рассуждениях понадобятся следующие моментные характеристики гамма-экспоненциального распределения (1), полученные в работе [11]:

$$\left. \begin{aligned} \mu_1 &\equiv E \ln \zeta = \frac{\nu \ln \delta + \psi(t) - r\psi(s)}{\nu}; \\ \mu_2 &\equiv E \ln^2 \zeta = \frac{[\nu \ln \delta + \psi(t) - r\psi(s)]^2}{\nu^2} + \\ &\quad + \frac{\psi'(t) + r^2\psi'(s)}{\nu^2}. \end{aligned} \right\} (3)$$

Обозначим

$$R_{k,l}(m) = \text{cov} \left( \ln^k X_n, \ln^l X_{n+m} \right) = \\ = E \ln^k X_n \ln^l X_{n+m} - \mu_k \mu_l, \quad k, l = 1, 2.$$

Всюду далее предполагается, что ряд  $\sum_{n=1}^{\infty} \rho(n)$  сходится. В этом случае существует положительно полуопределенная спектральная матрица [12, 13]

$$\Sigma_\lambda = 2\pi \begin{pmatrix} f_{1,1}(\lambda) & f_{1,2}(\lambda) \\ f_{1,2}(\lambda) & f_{2,2}(\lambda) \end{pmatrix}, \quad (4)$$

где

$$f_{k,l}(\lambda) = \frac{1}{2\pi} \sum_m e^{-im\lambda} R_{k,l}(m), \quad k, l = 1, 2,$$

непрерывны и ограничены. Ввиду четности  $R_{k,l}(m)$

$$f_{k,l}(\lambda) = \frac{1}{2\pi} \sum_m e^{-im\lambda} R_{k,l}(m) = \\ = \frac{1}{2\pi} \sum_m e^{im\lambda} R_{k,l}(m) = \frac{1}{2\pi} \sum_m \cos(m\lambda) R_{k,l}(m) = \\ = \frac{1}{2\pi} \left( R_{k,l}(0) + 2 \sum_{m=1}^{\infty} \cos(m\lambda) R_{k,l}(m) \right).$$

## 3 Оценки параметров гамма-экспоненциального распределения

Введем обозначение для выборочных логарифмических моментов случайной величины  $\zeta$ :

$$L_k(\mathbb{X}) = \frac{1}{n} \sum_{i=1}^n \ln^k X_i,$$

где  $\mathbb{X} = (X_1, \dots, X_n)$  — выборка из распределения  $\zeta \sim \text{GE}(r, \nu, s, t, \delta)$ .

Метод построения рассматриваемых оценок основан на решении системы уравнений

$$\mu_k = L_k(\mathbb{X}), \quad k = 1, 2,$$

причем предполагается, что параметры формы  $\nu$  и концентрации  $s$  и  $t$  фиксированы.

Приведем известные оценки параметров изгиба  $r$  и масштаба  $\delta$  гамма-экспоненциального распределения. Для этого введем в рассмотрение дигамма-функцию  $\psi(z) = \Gamma'(z)/\Gamma(z)$  и функции

$$\left. \begin{aligned} R(l_1, l_2) &= \sqrt{\frac{\nu^2(l_2 - l_1^2) - \psi'(t)}{\psi'(s)}}; \\ D(l_1, l_2) &= \exp \left\{ l_1 - \frac{\psi(t)}{\nu} + \right. \\ &\quad \left. + \frac{\psi(s)}{\nu} \sqrt{\frac{\nu^2(l_2 - l_1^2) - \psi'(t)}{\psi'(s)}} \right\}. \end{aligned} \right\} (5)$$

В силу состоятельности выборочных логарифмических моментов для оценок параметров  $r$  и  $\delta$ , полученных в работе [11], справедливо следующее утверждение.

**Лемма 1.** При фиксированных параметрах  $\nu, s$  и  $t$  распределения  $\text{GE}(r, \nu, s, t, \delta)$  оценки

$$\hat{r}(\mathbb{X}) = R(L_1(\mathbb{X}), L_2(\mathbb{X})); \quad (6)$$

$$\hat{\delta}(\mathbb{X}) = D(L_1(\mathbb{X}), L_2(\mathbb{X})) \quad (7)$$

параметров  $r$  и  $\delta$  обладают свойством состоятельности.

#### 4 Асимптотическая нормальность оценок параметров гамма-экспоненциального распределения

Утверждения об асимптотической нормальности оценок (6) и (7) базируются на следующих леммах [14].

**Лемма 2.** В  $\mathbb{R}^n$  случайный вектор  $X_n$  сходится по распределению к случайному вектору  $X$ , если и только если каждая линейная комбинация компонент  $X_n$  сходится по распределению к такой же линейной комбинации компонент  $X$ .

**Лемма 3.** Предположим, что в  $\mathbb{R}^k$  при  $n \rightarrow \infty$

$$\sqrt{n}(T_{n1}, \dots, T_{nk}) \implies N(\mu, \Sigma),$$

где  $\Sigma$  — ковариационная матрица. Пусть действительная функция  $g(t) = g(t_1, \dots, t_k)$  имеет отличный от нуля дифференциал в точке  $t = \mu$ . Положим

$$d = \left( \left. \frac{\partial g}{\partial t_1} \right|_{t=\mu}, \dots, \left. \frac{\partial g}{\partial t_k} \right|_{t=\mu} \right).$$

Тогда  $\sqrt{n}g(T_{n1}, \dots, T_{nk}) \implies N(g(\mu), d\Sigma d^T)$ .

Ввиду леммы 3 понадобятся производные функций (5):

$$\frac{\partial R}{\partial l_1}(l_1, l_2) = \frac{\nu^2}{2\psi'(s)R(l_1, l_2)};$$

$$\frac{\partial R}{\partial l_2}(l_1, l_2) = -\frac{l_2\nu^2}{\psi'(s)R(l_1, l_2)};$$

$$\frac{\partial D}{\partial l_1}(l_1, l_2) = \left( \frac{\nu\psi(s)}{2\psi'(s)R(l_1, l_2)} + 1 \right) D(l_1, l_2);$$

$$\frac{\partial D}{\partial l_2}(l_1, l_2) = -\frac{\nu\psi(s)l_2D(l_1, l_2)}{\psi'(s)R(l_1, l_2)}.$$

Обозначим

$$d_R = \left( \left. \frac{\partial R}{\partial l_1}(l_1, l_2) \right|_{(l_1, l_2)=(\mu_1, \mu_2)}, \left. \frac{\partial R}{\partial l_2}(l_1, l_2) \right|_{(l_1, l_2)=(\mu_1, \mu_2)} \right); \quad (8)$$

$$d_D = \left( \left. \frac{\partial D}{\partial l_1}(l_1, l_2) \right|_{(l_1, l_2)=(\mu_1, \mu_2)}, \left. \frac{\partial D}{\partial l_2}(l_1, l_2) \right|_{(l_1, l_2)=(\mu_1, \mu_2)} \right), \quad (9)$$

где математические ожидания  $\mu_k$  определены в (3).

Сформулируем утверждения об асимптотической нормальности оценок (6) и (7) при фиксированных параметрах формы  $\nu$  и концентрации  $s$  и  $t$ .

**Теорема 1.** Предположим, что спектральная матрица  $\Sigma_0$  в (4) положительно определена. При  $n \rightarrow \infty$  оценка (6) неизвестного параметра  $r$  асимптотически нормальна:

$$\sqrt{n} \frac{\hat{r}(\mathbb{X}) - r}{\sqrt{d_R \Sigma_0 d_R^T}} \implies N(0, 1),$$

где матрица  $\Sigma_0$  и вектор  $d_R$  задаются соотношениями (4) и (8).

Доказательство. Покажем асимптотическую нормальность вектора из выборочных логарифмических моментов  $(L_1(\mathbb{X}), L_2(\mathbb{X}))$ . Для этого введем в рассмотрение линейную комбинацию

$$Z_i(\alpha_1, \alpha_2) = \alpha_1 \ln X_i + \alpha_2 \ln^2 X_i, \quad \alpha_1^2 + \alpha_2^2 > 0.$$

Найдем дисперсию нормированной суммы таких линейных комбинаций. Имеем

$$\begin{aligned} D \frac{1}{n} \sum_{i=1}^n Z_i(\alpha_1, \alpha_2) &= D(\alpha_1 L_1(\mathbb{X}) + \alpha_2 L_2(\mathbb{X})) = \\ &= (\alpha_1, \alpha_2) \Sigma'(\alpha_1, \alpha_2)^T, \end{aligned}$$

где

$$\begin{aligned} \Sigma' &= \begin{pmatrix} \sigma_{1,1}(n) & \sigma_{1,2}(n) \\ \sigma_{1,2}(n) & \sigma_{2,2}(n) \end{pmatrix} \equiv \\ &\equiv \begin{pmatrix} DL_1(\mathbb{X}) & \text{cov}(L_1(\mathbb{X}), L_2(\mathbb{X})) \\ \text{cov}(L_1(\mathbb{X}), L_2(\mathbb{X})) & DL_2(\mathbb{X}) \end{pmatrix}. \end{aligned}$$

При  $n \rightarrow \infty$  [13]

$$n\sigma_{k,l}(n) \longrightarrow 2\pi f_{k,l}(0), \quad k, l = 1, 2.$$

Таким образом,

$$nD \frac{1}{n} \sum_{i=1}^n Z_i(\alpha_1, \alpha_2) \longrightarrow (\alpha_1, \alpha_2) \Sigma_0(\alpha_1, \alpha_2)^T.$$

Поскольку матрица  $\Sigma_0$  положительно определена и ряд  $\sum_{n=1}^{\infty} \rho(n)$  сходится, имеем [13]

$$\begin{aligned} &\left( \frac{1}{n} \sum_{i=1}^n Z_i(\alpha_1, \alpha_2) - \mathbf{E} \frac{1}{n} \sum_{i=1}^n Z_i(\alpha_1, \alpha_2) \right) \times \\ &\times \left( D \frac{1}{n} \sum_{i=1}^n Z_i(\alpha_1, \alpha_2) \right)^{-1/2} \implies N(0, 1). \end{aligned}$$

Следовательно, по лемме 2 получаем

$$\sqrt{n} (L_1(\mathbb{X}), L_2(\mathbb{X})) \implies N(\mu, \Sigma_0), \quad n \rightarrow \infty.$$

Кроме того, в условиях теоремы компоненты вектора  $d_R$ , определенного в (8), конечны и функция  $R(l_1, l_2)$  имеет в точке  $(\mu_1, \mu_2)$  отличный от нуля дифференциал.

Таким образом, выполнены все условия леммы 3. Следовательно,

$$\sqrt{n}R(L_1(\mathbb{X}), L_2(\mathbb{X})) \implies N(R(\mu_1, \mu_2), d_R \Sigma_0 d_R^T).$$

Теорема доказана.

Полностью аналогично доказывается теорема об асимптотической нормальности оценки  $\hat{\delta}(\mathbb{X})$  параметра масштаба  $\delta$ .

**Теорема 2.** *Предположим, что спектральная матрица  $\Sigma_0$  в (4) положительно определена. При  $n \rightarrow \infty$  оценка (7) неизвестного параметра  $\delta$  асимптотически нормальна:*

$$\sqrt{n} \frac{\hat{\delta}(\mathbb{X}) - \delta}{\sqrt{d_D \Sigma_0 d_D^T}} \implies N(0, 1),$$

где матрица  $\Sigma_0$  и вектор  $d_D$  задаются соотношениями (4) и (9).

## 5 Заключение

В работе показана асимптотическая нормальность построенных по выборке со слабо зависимыми компонентами оценок параметров изгиба и масштаба гамма-экспоненциального распределения при фиксированных параметрах формы и концентрации. Результаты работы автоматически переносятся на пару оценок параметров формы и масштаба при фиксированных параметрах изгиба и концентрации. Предложенный метод доказательства асимптотической нормальности применим в более широком классе задач, в частности при доказательстве центральной предельной теоремы для оценок параметров дигамма-распределения, обобщающего гамма-экспоненциальное распределение. Решение подобных задач представляет собой направление дальнейших исследований авторов.

## Литература

1. Кудрявцев А. А. О представлении гамма-экспоненциального и обобщенного отрицательного биномиального

распределений // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 76–80. doi: 10.14357/19922264190412.

2. Крицкий С. Н., Менкель М. Ф. О приемах исследования случайных колебаний речного стока // Труды НИУ ГУГМС. Сер. IV, 1946. Вып. 29. С. 3–32.
3. Крицкий С. Н., Менкель М. Ф. Выбор кривых распределения вероятностей для расчетов речного стока // Известия АН СССР. Отд. техн. наук, 1948. № 6. С. 15–21.
4. McDonald J. B. Some generalized functions for the size distribution of income // *Econometrica*, 1984. Vol. 52. No. 3. P. 647–665.
5. Кудрявцев А. А., Титова А. И. Гамма-экспоненциальная функция в байесовских моделях массового обслуживания // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 104–108. doi: 10.14357/19922264170413.
6. Кудрявцев А. А. Байесовские модели баланса // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 18–27. doi: 10.14357/19922264180303.
7. Iriarte Y. A., Varela H., Gómez H. J., Gómez H. W. A gamma-type distribution with applications // *Symmetry*, 2020. Vol. 12. Iss. 5. Art. 870. 15 p. doi: 10.3390/sym12050870.
8. Rivera P. A., Barranco-Chamorro I., Gallardo D. I., Gómez H. W. Scale mixture of Rayleigh distribution // *Mathematics*, 2020. Vol. 8. Iss. 10. Art. 1842. 22 p. doi: 10.3390/math8101842.
9. Combes C., Ng H. K. T. On parameter estimation for Amoroso family of distributions // *Math. Comput. Simulat.*, 2021. Vol. 191. P. 309–327.
10. Santoro K. I., Gómez H. J., Barranco-Chamorro I., Gómez H. W. Extended half-power exponential distribution with applications to COVID-19 data // *Mathematics*, 2022. Vol. 10. Iss. 6. Art. 942. 16 p. doi: 10.3390/math10060942.
11. Кудрявцев А. А., Шестаков О. В. Метод логарифмических моментов для оценивания параметров гамма-экспоненциального распределения // Информатика и её применения, 2020. Т. 14. Вып. 3. С. 49–54. doi: 10.14357/19922264200307.
12. Дженкинс Г., Ваттс Д. Спектральный анализ и его приложения / Пер. с англ. — М.: Мир, 1972. 285 с. (Jenkins G. M., Watts D. G. *Spectral analysis and its applications*. — San Francisco, CA, USA: Holden-Day, 1968. 552 p.)
13. Ибрагимов И. А. Замечание о центральной предельной теореме для зависимых случайных величин // Теория вероятностей и её применения, 1975. Т. 20. Вып. 1. С. 134–140.
14. Serfling R. J. *Approximation theorems of mathematical statistics*. — New York, NY, USA: John Wiley & Sons, 2002. 371 p.

Поступила в редакцию 03.07.23

# A METHOD FOR ESTIMATING PARAMETERS OF THE GAMMA-EXPONENTIAL DISTRIBUTION FROM A SAMPLE WITH WEAKLY DEPENDENT COMPONENTS

A. A. Kudryavtsev<sup>1,2</sup> and O. V. Shestakov<sup>1,2,3</sup>

<sup>1</sup>Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

<sup>2</sup>Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

<sup>3</sup>Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

**Abstract:** The article proves the asymptotic normality of the estimators for the gamma-exponential distribution parameters obtained using the modified method of moments in the case of a weak dependence of the sample components. For the estimators of the bent and scale parameters of the gamma-exponential distribution with fixed shape and concentration parameters, the central limit theorem is proved in the case when the maximum correlation coefficient between the sample elements tends to zero. The proof is based on the study of the sample spectral density and the results of the theory of stationary random sequences. The results of the article can be used to substantiate the asymptotic normality of the estimators for the parameters of the digamma distribution, the particular types of which include the generalized gamma distribution and the generalized beta distribution of the second kind that arise when describing processes modeled with distributions having a nonnegative unbounded support.

**Keywords:** weak dependence; parameter estimation; gamma-exponential distribution; mixed distributions; method of moments; asymptotic normality

**DOI:** 10.14357/19922264230308

**EDN:** PEXTVK

## Acknowledgments

The research was supported by the Russian Science Foundation, project No. 22-11-00212.

## References

1. Kudryavtsev, A. A. 2019. O predstavlenii gamma-eksponentsial'nogo i obobshchennogo otritsatel'nogo binomial'nogo raspredeleniy [On the representation of gamma-exponential and generalized negative binomial distributions]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):76–80. doi: 10.14357/19922264190412.
2. Kritsky, S. N., and M. F. Menkel. 1946. O priemakh issledovaniya sluchaynykh kolebaniy rechnogo stoka [Methods of investigation of random fluctuations of river flow]. *Trudy NIU GUGMS. Ser. IV* [Proceedings of GUGMS research institutions. Ser. IV] 29:3–32.
3. Kritsky, S. N., and M. F. Menkel. 1948. Vybor krivykh raspredeleniya veroyatnostey dlya raschetov rechnogo stoka [Selection of probability distribution curves for river flow calculations]. *Izvestiya AN SSSR. Otd. tekhn. nauk* [Herald of the Russian Academy of Sciences. Technical Sciences] 6:15–21.
4. McDonald, J. B. 1984. Some generalized functions for the size distribution of income. *Econometrica* 52(3):647–665.
5. Kudryavtsev, A. A., and A. I. Titova. 2017. Gamma-eksponentsial'naya funktsiya v bayesovskikh modelyakh massovogo obsluzhivaniya [Gamma-exponential function in Bayesian queuing models]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):104–108. doi: 10.14357/19922264170413.
6. Kudryavtsev, A. A. 2018. Bayesovskie modeli balansa [Bayesian balance models]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):18–27. doi: 10.14357/19922264180303.
7. Iriarte, Y. A., H. Varela, H. J. Gómez, and H. W. Gómez. 2020. A gamma-type distribution with applications. *Symmetry* 12(5):870. 15 p. doi: 10.3390/sym12050870.
8. Rivera, P. A., I. Barranco-Chamorro, D. I. Gallardo, and H. W. Gómez. 2020. Scale mixture of Rayleigh distribution. *Mathematics* 8(10):1842. 22 p. doi: 10.3390/math8101842.
9. Combes, C., and H. K. T. Ng. 2021. On parameter estimation for Amoroso family of distributions. *Math. Comput. Simulat.* 191:309–327. doi: 10.1016/j.matcom.2021.07.004
10. Santoro, K. I., H. J. Gómez, I. Barranco-Chamorro, and H. W. Gómez. 2022. Extended half-power exponential

- distribution with applications to COVID-19 data. *Mathematics* 10(6):942. 16 p. doi: 10.3390/math10060942.
11. Kudryavtsev, A. A., and O. V. Shestakov. 2020. Metod logarifmicheskikh momentov dlya otsenivaniya parametrov gamma-eksponentsialnogo raspredeleniya [Method of logarithmic moments for estimating the gamma-exponential distribution parameters]. *Informatika i ee Primeneniya — Inform. Appl.* 14(3):49–54. doi: 10.14357/19922264200307.
  12. Jenkins, G. M., and D. G. Watts. 1968. *Spectral analysis and its applications*. San Francisco, CA: Holden-Day. 552 p.
  13. Ibragimov, I. A. 1975. A note on the central limit theorems for dependent random variables. *Theor. Probab. Appl.* 20(1):135–141. doi: 10.1137/1120011.
  14. Serfling, R. J. 2002. *Approximation theorems of mathematical statistics*. New York, NY: John Wiley & Sons. 371 p.

Received July 3, 2023

## Contributors

**Kudryavtsev Alexey A.** (b. 1978) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; senior scientist, Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; aakudryavtsev@cs.msu.ru

**Shestakov Oleg V.** (b. 1976) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; leading scientist, Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; oshestakov@cs.msu.ru

# ЛОГИЧЕСКИЕ МЕТОДЫ КОРРЕКТНОЙ КЛАССИФИКАЦИИ ДАННЫХ

Е. В. Дюкова<sup>1</sup>, Г. О. Масляков<sup>2</sup>, А. П. Дюкова<sup>3</sup>

**Аннотация:** Работа посвящена вопросам применения дискретного аппарата (логических методов анализа целочисленных данных) для задачи классификации по прецедентам. Рассматриваются три направления логической классификации: Correct Voting Procedures (CVP), Logical Analysis of Data (LAD) и Formal Concept Analysis (FCA). С использованием терминологии направления CVP приводятся основные понятия, используемые в LAD и FCA. Описывается общая схема работы логического классификатора, согласно которой каждый логический классификатор на этапе обучения задает некоторый частичный порядок на специальном множестве фрагментов описаний прецедентов и ищет максимальные относительно заданного порядка элементы этого множества. Подобные исследования важны для создания общей теории корректной классификации по прецедентам на основе применения дискретного аппарата.

**Ключевые слова:** классификация на основе прецедентов; логический классификатор; процедуры корректного голосования; логический анализ данных; анализ формальных понятий; тупиковый представительный элементарный классификатор; сильная логическая закономерность; ДСМ-гипотеза; частичный порядок

DOI: 10.14357/19922264230309

EDN: OZHXOX

## 1 Введение

Задача классификации на основе прецедентов рассматривается в следующей постановке.

Исследуется множество объектов  $M$ . Каждый объект из  $M$  может быть представлен в виде числового набора, полученного на основе наблюдения или измерения ряда его характеристик. Такие характеристики называют признаками. Предполагается, что каждый признак имеет ограниченное множество допустимых значений, которые кодируются целыми числами. Известно, что  $M$  представимо в виде объединения непересекающихся подмножеств, называемых классами. Имеется конечный набор объектов множества  $M$ , про каждый из которых известно, какому классу он принадлежит. Это прецеденты, или обучающие объекты. Требуется по предъявленному набору значений признаков, описывающему некоторый объект из  $M$ , о котором, вообще говоря, неизвестно, какому классу он принадлежит, определить (распознать) этот класс.

Логический, или дискретный, подход к задаче классификации возник в связи с необходимостью прогнозировать редкие события, для которых нет достаточного статистического материала. Обучение классификатора сводится к поиску в исходных данных определенных закономерностей. Найденные закономерности позволяют различать объекты из разных классов и, как правило, имеют содержа-

тельное описание в терминах той прикладной области, в которой решается задача. По их наличию или, наоборот, отсутствию в описании распознаваемого объекта решается вопрос о его классификации. При этом большое внимание уделяется корректному обучению, позволяющему безошибочно классифицировать материал обучения. Особенно эффективен логический подход в случае целочисленной информации низкой значности, например бинарной. Однако в случае большого числа признаков при поиске закономерностей в описаниях прецедентов необходимо решать сложные дискретные задачи.

Фундаментальную роль в создании отечественных методов логической классификации сыграли работы члена-корреспондента РАН С. В. Яблонского, в которых введено хорошо известное в дискретной математике понятие теста, и работы академика РАН Ю. И. Журавлёва, опубликованные в 1970-х и 1980-х гг. Понятие теста (введено в [1]), первоначально применявшееся в задачах контроля управляющих систем, стало основным для конструирования одной из первых моделей логических классификаторов направления, именуемого далее CVP. Основы проблематики были заложены также в статьях российских ученых М. М. Бонгарда (1967 г.) и М. Н. Вайнцвайга (1973 г.). В дальнейшем направление CVP развивалось в работах отечественных и зарубежных авторов и наиболее

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, edjukova@mail.ru

<sup>2</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, gleb-mas@mail.ru

<sup>3</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, anastasia.d.95@gmail.com



существенное развитие получило в статьях [2–8]. В частности, в [6, 7] впервые построены логические классификаторы, ориентированные на задание частичных порядков на множествах значений признаков.

Зарубежные исследования в области логической классификации представлены методами направлений LAD и FCA. основополагающие идеи LAD и FCA принадлежат соответственно П. Хаммеру [9] и Р. Вилле [10].

В России методы LAD предложены практически параллельно с зарубежными авторами и, в основном, получили развитие в работах школы Ю. И. Журавлёва (см., например, [11]). Следует отметить, что в публикациях И. С. Масича развиваются идеи зарубежных представителей LAD [12]. Методы FCA представлены работами В. К. Финна, С. О. Кузнецова, М. И. Забейайло, Д. И. Игнатова и Д. В. Виноградова [13–17]. В [13] предложен так называемый метод автоматического порождения гипотез (или ДСМ-метод), который позднее в 1990-х гг. был адаптирован В. К. Финном и его учениками для задач машинного обучения.

Все три названных направления имеют много общего. С другой стороны, каждый из них использует свою терминологию и демонстрирует некоторую оригинальность. В разд. 2–4 настоящей работы приводятся основные понятия, используемые при описании классификаторов направлений CVP, LAD и FCA, а также принципиальные схемы работы этих классификаторов. В разд. 5 предлагается общая схема логической классификации целочисленных данных с использованием понятий направления CVP.

## 2 Направление CVP

Введем основные понятия, используемые при описании классификаторов направления CVP.

Пусть  $H = \{x_{j_1}, \dots, x_{j_r}\}$  — набор из  $r$  различных признаков,  $\sigma = (\sigma_1, \dots, \sigma_r)$  — набор, в котором  $\sigma_i$  — допустимое значение признака  $x_{j_i}$ ,  $i = 1, 2, \dots, r$ . Пара  $(\sigma, H)$  называется элементарным классификатором (ЭК).

Пусть  $S \in M$ ,  $S = (a_1, \dots, a_n)$ , где  $a_j$ ,  $j \in \{1, 2, \dots, n\}$ , — значение признака  $x_j$  для объекта  $S$ . Близость объекта  $S$  и ЭК  $(\sigma, H)$  оценивается величиной  $B(S, \sigma, H)$ :

$$B(S, \sigma, H) = \begin{cases} 1, & \text{если } a_{j_t} = \sigma_t \text{ при } t = 1, 2, \dots, r; \\ 0 & \text{в противном случае.} \end{cases}$$

Объект  $S$  содержит ЭК  $(\sigma, H)$ , если  $B(S, \sigma, H) = 1$ .

Элементарный классификатор  $(\sigma, H)$  называется корректным для класса  $K$ , если для любой пары прецедентов  $S \in K$  и  $S' \notin K$  не выполняется  $B(S, \sigma, H) = B(S', \sigma, H) = 1$ . Корректный ЭК  $(\sigma, H)$  класса  $K$  называется тупиковым, если любой ЭК  $(\sigma', H')$  такой, что  $\sigma' \subset \sigma$ ,  $H' \subset H$ , не является корректным для  $K$ .

Множество прецедентов класса  $K$  обозначается через  $R(K)$ . Элементарный классификатор  $(\sigma, H)$  — (тупиковый) представительный для класса  $K$ , если  $(\sigma, H)$  — (тупиковый) корректный ЭК для  $K$  и хотя бы один объект из  $R(K)$  содержит  $(\sigma, H)$ .

Положим  $R_K(\sigma, H) = \{S \in R(K) : B(S, \sigma, H) = 1\}$ ,  $|R_K(\sigma, H)|$  — мощность множества  $R_K(\sigma, H)$ .

На этапе обучения классифицирующий алгоритм строит для каждого класса  $K$  некоторое множество  $P_1(K)$  тупиковых представительных ЭК. Оценка за принадлежность объекта  $S$  классу  $K$  вычисляется на основе суммирования величин  $|R_K(\sigma, H)| \times B(S, \sigma, H)$ ,  $(\sigma, H) \in P_1(K)$ . Объект  $S$  относится к классу с наибольшей оценкой. Если максимальную оценку получают несколько классов, то происходит отказ от классификации. При построении  $P_1(K)$  возникают труднорешаемые дискретные перечислительные задачи, среди которых центральное место принадлежит задаче монотонной дуализации [6].

## 3 Направление LAD

При описании методов направления LAD, как правило, рассматриваются задачи с бинарными признаками. Переход от целочисленных признаков к бинарным может быть осуществлен посредством описанного ниже *one-hot* преобразования (см. замечание 1).

Отметим, что в случае бинарных признаков ЭК  $(\sigma, H)$ ,  $H = \{x_{j_1}, \dots, x_{j_r}\}$ ,  $\sigma = (\sigma_1, \dots, \sigma_r)$ ,  $\sigma_i \in \{0, 1\}$ ,  $i = 1, 2, \dots, r$ , — это элементарная конъюнкция  $B$  над переменными  $x_1, \dots, x_n$  вида  $x_{j_1}^{\sigma_1} \wedge \dots \wedge x_{j_r}^{\sigma_r}$ , которая обращается в 1 на описании объекта  $S$ , если объект  $S$  содержит ЭК  $(\sigma, H)$ , т. е. если описание объекта  $S$  принадлежит интервалу истинности  $N_B$  конъюнкции  $B$ .

Пусть  $f_K(x_1, \dots, x_n)$  — частичная булева функция, определенная на бинарных описаниях объектов обучающей выборки и принимающая значения 1 и 0 соответственно на описаниях прецедентов из класса  $K$  и не из  $K$ . Через  $N_{f_K}$  и  $N_{\bar{f}_K}$  обозначим соответственно множество описаний прецедентов из класса  $K$  и множество описаний прецедентов, не принадлежащих классу  $K$ .

Элементарная конъюнкция  $B$  называется *допустимой* для  $f_K$ , если  $N_B \cup N_{\bar{f}_K} = \emptyset$  и  $N_B \cup N_{f_K} \neq \emptyset$ . Допустимая для  $f_K$  конъюнкция  $B$  называется *максимальной* для  $f_K$ , если не существует допустимой для  $f_K$  конъюнкции  $B'$  такой, что  $N_{B'} \supset N_B$ . Нетрудно видеть, что в терминах теории булевых функций представительный ЭК класса  $K$  — это допустимая конъюнкция для функции  $f_K$ , а тупиковый представительный ЭК класса  $K$  — это максимальная конъюнкция для  $f_K$ . В LAD допустимая конъюнкция для  $f_K$  называется *логической закономерностью* (ЛЗ) класса  $K$ . Таким образом, определение ЛЗ эквивалентно определению представительного ЭК.

Пусть ЭК  $(\sigma, H)$  — представительный ЭК для класса  $K$ . Элементарный классификатор  $(\sigma, H)$  называется *наибольшей ЛЗ* (*maximum pattern*), если  $|R_K(\sigma, H)| \geq |R_K(\sigma', H')|$  для любого другого представительного ЭК  $(\sigma', H')$  класса  $K$ . Элементарный классификатор  $(\sigma, H)$  называется *сильной ЛЗ* (*strong pattern*), если не существует другого представительного ЭК  $(\sigma', H')$  класса  $K$  такого, что  $R_K(\sigma, H) \subset R_K(\sigma', H')$  [18].

В LAD схема работы классифицирующего алгоритма полностью аналогична схеме работы алгоритма CVP. На этапе обучения для каждого класса  $K$  строится некоторое множество  $P_2(K)$  ЛЗ, например строятся сильные ЛЗ. При этом решаются сложные в вычислительном плане оптимизационные задачи линейного программирования. На следующем этапе каждый элемент множества  $P_2(K)$  «голосует» за отнесение объекта  $S$  классу  $K$ . Для оценки принадлежности распознаваемого объекта  $S$  к классу  $K$  суммируются величины  $B(S, \sigma, H)$ ,  $(\sigma, H) \in P_2(K)$ .

**Замечание 1.** Опишем процедуру бинаризации исходных целочисленных данных. Рассмотрим признак  $x_j$ ,  $j \in \{1, 2, \dots, n\}$ , и множество  $N_j = \{a_{j1}, \dots, a_{jk_j}\}$  допустимых значений этого признака. Пусть  $a, b \in N_j$ . Положим

$$\delta(a, b) = \begin{cases} 1, & \text{если } a = b; \\ 0 & \text{иначе.} \end{cases}$$

Тогда бинарное описание объекта  $S = (a_1, \dots, a_n)$ ,  $a_j \in N_j$ ,  $j = 1, 2, \dots, n$ , полученное из исходного на основе *one-hot* преобразования, будет иметь вид:  $(\delta(a_{11}, a_1), \dots, \delta(a_{1k_1}, a_1), \dots, \delta(a_{n1}, a_n), \dots, \delta(a_{nk_n}, a_n))$ .

Таким образом, каждый объект может быть описан при помощи  $k_1 + \dots + k_n$  бинарных признаков. Нетрудно заметить, что признаку  $x_j$ ,  $j = 1, 2, \dots, n$ , и допустимому значению  $a_{ji} \in N_j$ ,  $i = 1, 2, \dots, k_j$ , признака  $x_j$  соответствует бинарный признак  $\tilde{x}_t$ ,  $t = k_1 + \dots + k_{j-1} + i$ , причем  $\tilde{x}_t$  принимает значение 1 тогда и только тогда, когда признак  $x_j$  при-

нимает значение  $a_{ji}$ . Бинарный признак  $\tilde{x}_t$  и пару  $(a_{ji}, x_j)$  назовем *родственными*.

**Замечание 2.** В задачах с вещественными признаками, как правило, признаки сначала преобразуют в целочисленные признаки (см., например, [19]), а затем проводится процесс бинаризации.

## 4 Направление FCA

Опишем основные понятия из FCA. Отметим, что обычно методы FCA, как и методы LAD, работают с бинарной информацией. Однако *one-hot* преобразование (см. замечание 1 в разд. 3) применяется в FCA и в этом случае.

Пусть  $\tilde{X}$  — множество бинарных признаков мощности  $n'$ , полученное в результате *one-hot* преобразования множества  $X$ . Говорят, что объект  $S = (a_1, \dots, a_{n'})$  *обладает признаком*  $\tilde{x}_j \in \tilde{X}$ ,  $j \in \{1, 2, \dots, n'\}$ , если  $a_j = 1$ .

*Формальным контекстом* называется тройка вида  $C = (R, \tilde{X}, I)$ , где  $R$  — некоторое множество объектов;  $I$  — бинарное отношение между множествами  $R$  и  $\tilde{X}$ . Иначе говоря, формальный контекст представляет собой булеву матрицу  $L$ , строками которой служат описания объектов из  $R$  посредством признаков из  $\tilde{X}$ . Далее запись  $(S, \tilde{x}) \in I$  означает, что объект  $S \in R$  обладает признаком  $\tilde{x} \in \tilde{X}$ .

*Операторами вывода* для формального контекста  $C = (R, \tilde{X}, I)$  и множеств  $A \subseteq R$ ,  $B \subseteq \tilde{X}$  называются соответственно множества  $A^* = \{\tilde{x} \in \tilde{X} | (S, \tilde{x}) \in I \forall S \in A\}$  и  $B^* = \{S \in R | (S, \tilde{x}) \in I \forall \tilde{x} \in B\}$ .

Таким образом,  $A^*$  — это оператор, возвращающий столбцы матрицы  $L$ , которые в пересечении с заданными строками образуют «максимальную» подматрицу, имеющую  $|A|$  строк и не имеющую элементов, равных 0. Аналогично,  $B^*$  — это оператор, возвращающий строки матрицы  $L$ , которые в пересечении с заданными столбцами образуют «максимальную» подматрицу, имеющую  $|B|$  столбцов и не имеющую элементов, равных 0.

*Формальным понятием* формального контекста  $C = (R, \tilde{X}, I)$  называется пара вида  $(A, B)$ ,  $A \subseteq R$ ,  $B \subseteq \tilde{X}$ , такая что  $A^* = B$ ,  $B^* = A$ . Заметим, что паре  $(A, B)$  соответствует «максимальная» подматрица матрицы  $L$ , состоящая из единичных элементов.

Рассмотрим формальные контексты  $C_K = (R(K), \tilde{X}, I)$  и  $C_{\bar{K}} = (R(\bar{K}), \tilde{X}, I)$ . Введем обозначения:  $H^+$  — оператор вывода для формального контекста  $C_K$  и множества признаков  $H$ ;  $H^-$  — оператор вывода для формального контекста  $C_{\bar{K}}$  и множества признаков  $H$ .

Набор признаков  $H \subseteq \tilde{X}$  называется *положительной ДСМ-гипотезой* для класса  $K$ , если пара

$(H^+, H)$  является формальным понятием формального контекста  $C_K$  и  $H^- = \emptyset$ .

Положительной ДСМ-гипотезе  $H = \{\tilde{x}_{j_1}, \dots, \tilde{x}_{j_r}\}$  класса  $K$  поставим в соответствие ЭК  $(\sigma, H')$ ,  $\sigma = (\sigma_1, \dots, \sigma_r)$ ,  $H' = \{x_{i_1}, \dots, x_{i_r}\}$ ,  $H' \subseteq \subseteq X$ , такой что при любом  $i \in \{1, 2, \dots, r\}$  пара  $(\sigma_i, x_{i_i})$  и признак  $\tilde{x}_{j_i}$  являются родственными (см. замечание 1 в разд. 3). Будем говорить, что  $H$  порождает ЭК  $(\sigma, H')$ . Нетрудно видеть, что  $(\sigma, H')$  — представительный ЭК класса  $K$ .

**Пример.** Рассмотрим задачу классификации с тремя прецедентами  $S_1 = (1, 1, 1)$ ,  $S_2 = (0, 1, 1)$  и  $S_3 = (1, 0, 0)$ . При этом  $S_1, S_2 \in K_1$ ,  $S_3 \in K_2$ . После one-hot преобразования прецеденты будут иметь соответственно описания  $(0, 1, 0, 1, 0, 1)$ ,  $(1, 0, 0, 1, 0, 1)$ ,  $(0, 1, 1, 0, 1, 0)$ . В данном примере ЭК  $((1, 1), \{x_2, x_3\})$  порожден положительной для класса  $K_1$  ДСМ-гипотезой  $H = (\tilde{x}_4, \tilde{x}_6)$ . Этот ЭК не является тупиковым представительным ЭК класса  $K_1$ , поскольку  $((1), \{x_2\})$  — представительный ЭК для  $K_1$ . Отметим, что тупиковый представительный ЭК  $((1), \{x_2\})$  класса  $K_1$  не порождается ни одной положительной ДСМ-гипотезой класса  $K_1$ .

ДСМ-классификатор действует более строго по сравнению с классификаторами из CVP и LAD. На первом этапе для каждого класса  $K$  строится некоторое множество  $P_3(K)$  положительных для класса  $K$  гипотез. Объект  $S$  относится к классу  $K$ , если  $S$  содержит хотя бы один ЭК, порождаемый гипотезой из  $P_3(K)$ , и не содержит ни одного ЭК, порождаемого гипотезами из  $P_3(K')$ ,  $K' \neq K$ . В противном случае происходит отказ от классификации. При построении  $P_3(K)$  возникают дискретные перечислительные задачи, которые алгоритмически менее сложны, чем задача монотонной дуализации [20].

## 5 Схема обучения логического классификатора

Предлагаемое описание общей схемы обучения алгоритмов логической классификации основано на приводимых ниже утверждениях 1–4.

Пусть  $\mathcal{P}(K)$  — множество всех представительных ЭК класса  $K$ . Бинарное отношение  $\preceq$  на  $\mathcal{P}(K)$  называется частичным предпорядком, если это отношение рефлексивно  $((\sigma, H) \preceq (\sigma, H))$  для любого  $(\sigma, H) \in \mathcal{P}(K)$  и транзитивно  $((\sigma, H) \preceq (\sigma', H')$  и  $(\sigma', H') \preceq (\sigma'', H'')$  влечет  $(\sigma, H) \preceq (\sigma'', H'')$ ). Частичный предпорядок называется *частичным порядком*, если он антисимметричен  $((\sigma, H) \preceq (\sigma', H')$  и  $(\sigma', H') \preceq (\sigma, H)$  влечет  $(\sigma, H) = (\sigma', H')$ ). Запись  $(\sigma, H) \prec (\sigma', H')$  означает, что  $(\sigma, H) \preceq (\sigma', H')$

и  $(\sigma', H') \not\preceq (\sigma, H)$ . Элементарный классификатор  $(\sigma, H) \in \mathcal{P}(K)$  называется *максимальным элементом* в  $\mathcal{P}(K)$  относительно частичного (пред)порядка  $\preceq$ , если не существует ЭК  $(\sigma', H') \in \mathcal{P}(K)$  такого, что  $(\sigma, H) \prec (\sigma', H')$ .

Зададим частичный порядок  $\preceq_1$  на множестве  $\mathcal{P}(K)$ . Будем считать, что  $(\sigma_2, H_2) \in \mathcal{P}(K)$  следует за  $(\sigma_1, H_1) \in \mathcal{P}(K)$ , если  $H_2 \subseteq H_1$  и  $\sigma_2 \subseteq \sigma_1$ . Справедливо

**Утверждение 1.** *Элементарный классификатор  $(\sigma, H)$  будет тупиковым представительным для класса  $K$  тогда и только тогда, когда  $(\sigma, H)$  — максимальный относительно частичного порядка  $\preceq_1$  элемент множества  $\mathcal{P}(K)$ .*

Справедливость утверждения 1 следует непосредственно из определения понятия тупикового представительного ЭК, приведенного в разд. 2.

Зададим на множестве  $\mathcal{P}(K)$  отношение частичного предпорядка  $\preceq_2$ . Будем считать, что  $(\sigma_2, H_2) \in \mathcal{P}(K)$  следует за  $(\sigma_1, H_1) \in \mathcal{P}(K)$ , если  $|R_K(\sigma_1, H_1)| \leq |R_K(\sigma_2, H_2)|$ . Справедливо

**Утверждение 2.** *Элементарный классификатор  $(\sigma, H)$  будет наибольшей ЛЗ класса  $K$  тогда и только тогда, когда  $(\sigma, H)$  — максимальный относительно частичного предпорядка  $\preceq_2$  элемент множества  $\mathcal{P}(K)$ .*

Справедливость утверждения 2 следует непосредственно из определения понятия наибольшей ЛЗ, приведенного в разд. 3.

Зададим на множестве  $\mathcal{P}(K)$  отношение частичного предпорядка  $\preceq_3$ , предложенного в [21]. Будем считать, что  $(\sigma_2, H_2) \in \mathcal{P}(K)$  следует за  $(\sigma_1, H_1) \in \mathcal{P}(K)$ , если  $R_K(\sigma_1, H_1) \subseteq R_K(\sigma_2, H_2)$ . Справедливо

**Утверждение 3.** *Элементарный классификатор  $(\sigma, H)$  будет сильной ЛЗ класса  $K$  тогда и только тогда, когда  $(\sigma, H)$  — максимальный относительно частичного предпорядка  $\preceq_3$  элемент множества  $\mathcal{P}(K)$ .*

Справедливость утверждения 3 следует непосредственно из определения понятия сильной ЛЗ, приведенного в разд. 3.

Зададим на множестве  $\mathcal{P}(K)$  отношение частичного порядка  $\preceq_4$ . Будем считать, что ЭК  $(\sigma_2, H_2) \in \mathcal{P}(K)$  следует за  $(\sigma_1, H_1) \in \mathcal{P}(K)$ , если  $R_K(\sigma_1, H_1) \subseteq R_K(\sigma_2, H_2)$  и  $H_1 \subseteq H_2$ . Справедливо

**Утверждение 4.** *Элементарный классификатор  $(\sigma, H)$  порождается положительной ДСМ-гипотезой для класса  $K$  тогда и только тогда, когда  $(\sigma, H)$  — максимальный относительно частичного порядка  $\preceq_4$  элемент множества  $\mathcal{P}(K)$ .*

**Доказательство. Достаточность.** Пусть ЭК  $(\sigma, H) \in \mathcal{P}(K)$  — максимальный элемент в  $\mathcal{P}(K)$  относительно порядка  $\preceq_4$ . Рассмотрим формальный контекст  $C_K = (R(K), \tilde{X}, I)$ , а также множе-

ство признаков  $\tilde{H}$  из  $\tilde{X}$ , порождающее ЭК  $(\sigma, H)$ . По определению  $\tilde{H}^+ = R_K(\sigma, H)$ . Покажем, что  $R_K(\sigma, H)^+ = \tilde{H}$ .

Предположим противное. Пусть  $R_K(\sigma, H)^+ = H'$ ,  $\tilde{H} \subset H'$ . Это означает, что существуют признак  $x_i \in X$ ,  $x_i \notin H$ , и его допустимое значение  $\sigma'$ , такие что признак  $x_i$  принимает значение  $\sigma'$  на всех объектах из  $R_K(\sigma, H)$ . Заметим, что ЭК  $((\sigma_1, \dots, \sigma_r, \sigma'), H \cup x_i)$  — корректный для класса  $K$ , так как  $(\sigma, H)$  не содержится в прецедентах не из класса  $K$ . Рассматриваемый ЭК содержится в прецедентах из класса  $K$ , поскольку он содержится в объектах из множества  $R_K(\sigma, H)$ . Следовательно,  $((\sigma_1, \dots, \sigma_r, \sigma'), H \cup x_i)$  — представительный ЭК для класса  $K$ .

Нетрудно видеть, что  $R_K((\sigma_1, \dots, \sigma_r, \sigma'), H \cup x_i) = R_K(\sigma, H)$ . Это значит, что  $(\sigma, H) \prec_4 \prec_4 ((\sigma_1, \dots, \sigma_r, \sigma'), H \cup x_i)$ . Противоречие.

Необходимость очевидна. Утверждение доказано.

## 6 Заключение

В работе приведена общая схема работы алгоритмов логической классификации. В рамках данной схемы в единой терминологии дано оригинальное описание основных моделей логических классификаторов направлений CVP, LAD и FCA. Приведены утверждения, показывающие, что каждое из рассмотренных направлений логической классификации на этапе обучения строит для каждого класса  $K$  специальное подмножество множества  $\mathcal{P}(K)$  представительных ЭК этого класса. Данное подмножество определяется заданием частичного порядка или предпорядка на множестве  $\mathcal{P}(K)$  и образуется максимальными относительно заданного порядка элементами. При этом каждое из направлений ориентируется на «свой» порядок. Результаты статьи имеют значение для создания общей теории логической классификации и разработки более совершенных моделей классификаторов. Например, идеи, лежащие в основе методов анализа частично упорядоченных данных, разработанных в последнее время в направлении CVP, могут быть успешно применены и для построения классификаторов более общего вида в двух других направлениях логической классификации.

## Литература

1. Чегис И. А., Яблонский С. В. Логические способы контроля электрических схем // Труды Математического института имени В. А. Стеклова АН СССР, 1958. Т. 51. С. 270–360.

2. Баскакова Л. В., Журавлёв Ю. И. Модель распознающих алгоритмов с представительными наборами и системами опорных множеств // Ж. вычисл. матем. и матем. физ., 1981. Т. 21. № 5. С. 1264–1275.
3. Дюкова Е. В., Журавлёв Ю. И., Рудаков К. В. Об алгебраическом синтезе корректирующих процедур распознавания на базе элементарных алгоритмов // Ж. вычисл. матем. и матем. физ., 1996. Т. 36. № 8. С. 215–223.
4. Дюкова Е. В., Журавлёв Ю. И. Дискретный анализ признаков описаний в задачах распознавания большой размерности // Ж. вычисл. матем. и матем. физ., 2000. Т. 40. № 8. С. 1264–1278.
5. Дюкова Е. В., Песков Н. В. Поиск информативных фрагментов описаний объектов в дискретных процедурах распознавания // Ж. вычисл. матем. и матем. физ., 2002. Т. 42. № 5. С. 741–753.
6. Дюкова Е. В., Масляков Г. О., Прокофьев П. А. О логическом анализе данных с частичными порядками в задаче классификации по прецедентам // Ж. вычисл. матем. и матем. физ., 2019. Т. 59. № 9. С. 1605–1616.
7. Дюкова Е. В., Масляков Г. О. О выборе частичных порядков на множествах значений признаков в задаче классификации // Информатика и её применения, 2021. Т. 15. Вып. 4. С. 72–78. doi: 10.14357/19922264210410.
8. Dragunov N. A., Djukova E. V., Djukova A. P. Supervised classification and finding frequent elements in data // 8th Conference (International) on Information Technology and Nanotechnology Proceedings. — Piscataway, NJ, USA: IEEE, 2022. Art. 9848521. 5 p.
9. Crata, Y., Hammer P. L., Ibaraki T. Cause-effect relationships and partially defined Boolean functions // Ann. Oper. Res., 1988. Vol. 16. Iss. 1. P. 299–325. doi: 10.1007/BF02283750.
10. Wille R. Restructuring lattice theory: An approach based on hierarchies of concepts // Ordered sets / Ed. I. Rival. — NATO Advanced Study Institutes ser. — Netherlands: Springer, 1981. Vol. 83. P. 445–470. doi: 10.1007/978-94-009-7798-3\_15.
11. Журавлёв Ю. И., Рязанов В. В., Сенько О. В. Распознавание. Математические методы. Программная система. Практические применения. — М.: ФАЗИС, 2006. 159 с.
12. Масич И. С. Метод оптимальных логических решающих правил для задач распознавания и прогнозирования // Системы управления и информационные технологии, 2019. Т. 75. № 1. С. 31–37.
13. Финн В. К. О возможности формализации правдоподобных рассуждений средствами многозначных логик // Всесоюз. симпозиум по логике и методологии науки. — Киев: Наукова думка, 1976. С. 82–83.
14. Kuznetsov S. O. Mathematical aspects of concept analysis // J. Mathematical Sciences, 1996. Vol. 80. Iss. 2. P. 1654–1698. doi: 10.1007/BF02362847.
15. Gnatyshak D. V., Ignatov D. I., Kuznetsov S. O. Triadic formal concept analysis and triclustering: Searching for opti-

- mal patterns // *Mach. Learn.*, 2015. Vol. 101. P. 271–302. doi: 10.1007/s10994-015-5487-y.
16. *Забезжайло М. И.* О некоторых оценках сложности вычислений в ДСМ-рассуждениях // *Искусственный интеллект и принятие решений*, 2015. Часть I: № 1. С. 3–17. Часть II: № 2. С. 3–17.
  17. *Виноградов Д. В.* О представлении объектов битовыми строками для ВКФ-метода // *Научная и техническая информация*. Сер. 2, 2018. № 5. С. 1–4.
  18. *Bonates T. O., Hammer P. L., Kogan A.* Maximum patterns in datasets // *Discrete Appl. Math.*, 2008. Vol. 156. Iss. 6. P. 846–861. doi: 10.1016/j.dam.2007.06.004.
  19. *Eckstein J., Hammer P. L., Liu Y., Nediak M., Simeone B.* The maximum box problem and its application to data analysis // *Comput. Optim. Appl.*, 2002. Vol. 23. P. 285–298. doi: 10.1023/A:1020546910706.
  20. *Kuznetsov S. O., Obiedkov S. A.* Comparing performance of algorithms for generating concept lattice // *J. Exp. Theor. Artif. In.*, 2002. Vol. 14. Iss. 2-3. P. 189–216. doi: 10.1080/09528130210164170.
  21. *Hammer P. L., Kogan A., Simeone A., Szedmak B.* Pareto-optimal patterns in logical analysis of data // *Discrete Appl. Math.*, 2004. Vol. 144. Iss. 1-2. P. 79–102. doi: 10.1016/j.dam.2003.08.013.

*Поступила в редакцию 11.01.23*

---



---

## LOGICAL METHODS OF CORRECT DATA CLASSIFICATION

E. V. Djukova, G. O. Masliakov, and A. P. Djukova

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

**Abstract:** The work is devoted to issues of the application of discrete apparatus (logical methods of integer data analysis) for the supervised classification problem. Three main directions of logical classification are considered: Correct Voting Procedures (CVP), Logical Analysis of Data (LAD), and Formal Concept Analysis (FCA). Using the terminology of the CVP direction, the basic concepts applied in LAD and FCA are presented. The general scheme of the logical classifier is described, according to which each logical classifier at the training stage sets some partial order on a special set of fragments of precedents descriptions and searches for the maximum elements of this set relative to the given order. Such studies are important for creating a general theory of correct classification according to precedents based on the use of a discrete apparatus.

**Keywords:** supervised classification; logical classifier; Correct Voting Procedures; Logical Analysis of Data; Formal Concept Analysis; irredundant representative elementary classifier; strong pattern; JSM-hypothesis; partial order

**DOI:** 10.14357/19922264230309

**EDN:** OZHXOX

### References

1. Chegis, I. A., and S. V. Yablonskiy. 1958. Logicheskie sposoby kontrolya elektricheskikh skhem [Logical methods of control of work of electric schemes]. *Trudy Matematicheskogo instituta imeni V. A. Steklova AN SSSR* [Proceedings of the Steklov Institute of Mathematics of the USSR Academy of Sciences] 51:270–360.
2. Baskakova, L., and Yu. Zhuravlev. 1981. A model of recognition algorithms with representative samples and systems of supporting sets. *USSR Comp. Math. Math.* 21(5):189–199. doi: 10.1016/0041-5553(81)90109-9.
3. Djukova, E. V., Yu. I. Zhuravlev, and K. V. Rudakov. 1996. Algebraic-logic synthesis of correct recognition procedures based on elementary algorithms. *Comput. Math. Math. Phys.* 36(8):1161–1167.
4. Djukova, E., and Yu. Zhuravlev. 2000. Discrete analysis of feature descriptions in recognition problems of high dimensionality. *Comp. Math. Math. Phys.* 40(8):1214–1227.
5. Djukova, E., and N. Peskov. 2002. Search for informative fragments of object descriptions in discrete recognition procedures. *Comp. Math. Math. Phys.* 42(5):711–723.
6. Djukova, E. V., G. O. Maslyakov, and P. A. Prokofyev. 2019. On the logical analysis of partially ordered data in the supervised classification problem. *Comput. Math. Math. Phys.* 59(9):1542–1552.
7. Djukova, E. V., and G. O. Maslyakov. 2021. O vybore chas-tichnykh poryadkov na mnozhestvakh znacheniy priznakov v zadache klassifikatsii [On the choice of partial orders on feature values sets in the supervised classification problem]. *Informatika i ee Primeneniya — Inform. Appl.* 15(4):72–78. doi: 10.14357/19922264210410.
8. Dragunov, N. A., E. V. Djukova, and A. P. Djukova. 2022. Supervised classification and finding frequent elements in data. *8th Conference (International) on Information Technology and Nanotechnology Proceedings*. 9848521. 5 p.
9. Crama, Y., P. L. Hammer, and T. Ibaraki. 1988. Cause-effect relationships and partially defined Boolean functions. *Ann. Oper. Res.* 16(1):299–325. doi: 10.1007/BF02283750.
10. Wille, R. 1982. Restructuring lattice theory: An approach based on hierarchies of concepts. *Ordered sets*. Ed. I. Rival. NATO Advanced Study Institutes ser. Netherlands: Springer. 83:445–470. doi: 10.1007/978-94-009-7798-3\_15.

11. Zhuravlev, Yu. I., V. V. Ryazanov, and O. V. Senko. 2006. *Raspoznavanie. Matematicheskie metody. Programmaya sistema. Prakticheskie primeneniya* [Recognition. Mathematical methods. Software system. Practical applications]. Moscow: FAZIS. 159 p.
12. Masich, I. S. 2019. Metod optimal'nykh logicheskikh reshayushchikh pravil dlya zadach raspoznavaniya i prognozirovaniya [The method of optimal logical decision rules for recognition and prediction problems]. *Sistemy upravleniya i informatsionnye tekhnologii* [Control Systems and Information Technologies] 75(1):31–37.
13. Finn, V. K. 1976. O vozmozhnosti formalizatsii pravdopodobnykh rassuzhdeniy sredstvami mnogoznachnykh logik [On the possibility of formalizing plausible reasoning by means of multivalued logics]. *Vsesoyuzn. simpozium po logike i metodologii nauki* [All-Union Symposium on the Logic and Methodology of Science]. Kiev: Naukova dumka. 82–83.
14. Kuznetsov, S. O. 1996. Mathematical aspects of concept analysis. *J. Mathematical Sciences* 80(2):1654–1698. doi: 10.1007/BF02362847.
15. Ignatov, D. I., D. V. Gnatyshak, S. O. Kuznetsov, and B. G. Mirkin. 2015. Triadic formal concept analysis and triclustering: Searching for optimal patterns. *Mach. Learn.* 101:271–302. doi: 10.1007/s10994-015-5487-y.
16. Zabezhailo, M. I. 2015. Some capabilities of enumeration control in the DSM method. *Scientific and Technical Information Processing. Part One*: 41(6):335–347. doi: 10.3103/S0147688214060082. *Part Two*: 41(6):348–361. doi: 10.3103/S0147688214060094.
17. Vinogradov, D. V. 2018. On object representation by bit strings for the VKF-method. *Automatic Documentation Mathematical Linguistics* 52(3):113–116. doi: 10.3103/S0005105518030044.
18. Bonates, T. O., P. L. Hammer, and A. Kogan. 2008. Maximum patterns in datasets. *Discrete Appl. Math.* 156(6):846–861. doi: 10.1016/j.dam.2007.06.004.
19. Eckstein, J., P. L. Hammer, Y. Liu, M. Nediak, and B. Simeone. 2002. The maximum box problem and its application to data analysis. *Comput. Optim. Appl.* 23:285–298. doi: 10.1023/A:1020546910706.
20. Kuznetsov, S. O., and S. A. Obiedkov. 2002. Comparing performance of algorithms for generating concept lattice. *J. Exp. Theor. Artif. In.* 14(2-3):189–216. doi: 10.1080/09528130210164170.
21. Hammer, P. L., A. Kogan, A. Simeone, and B. Szedmák. 2004. Pareto-optimal patterns in logical analysis of data. *Discrete Appl. Math.* 144(1-2):79–102. doi: 10.1016/j.dam.2003.08.013.

Received January 11, 2023

## Contributors

**Djukova Elena V.** (b. 1945) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119333, Russian Federation; edjukova@mail.ru

**Masliakov Gleb O.** (b. 1996) — PhD student, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; gleb-mas@mail.ru

**Djukova Anastasia P.** (b. 1995) — PhD student, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; anastasia.d.95@gmail.com

# КЛАССИФИКАЦИЯ С ПОМОЩЬЮ ПРИЧИННО-СЛЕДСТВЕННЫХ СВЯЗЕЙ

А. А. Грушо<sup>1</sup>, Н. А. Грушо<sup>2</sup>, М. И. Забейло<sup>3</sup>, Д. В. Смирнов<sup>4</sup>, Е. Е. Тимонина<sup>5</sup>

**Аннотация:** По определению свойство  $A$  в объекте  $O$  служит причиной появления следствия  $B$ , которое доступно наблюдению в информационном пространстве  $I$ , если характеристики  $A$  могут породить объект в пространстве  $I$ , содержащий следствие  $B$ , и в этом случае при появлении  $A$  детерминированно появляется  $B$ . Поэтому можно рассматривать задачу классификации как вычисление следствий характеристик объекта, где в качестве следствий выступают характеристики класса. В этом случае характеристики объекта классификации можно рассматривать как причину, которая детерминированно (классификация как отображение) порождает следствия (характеристики класса). В рассматриваемом случае каждое из свойств  $A_i$ ,  $i = 1, \dots, k$ , служит причиной детерминированного появления непустого множества своих следствий. Если число классов велико, так же как множества следствий каждого  $A_i$ , то задача классификации может оказаться трудно вычислимой из-за того, что возможны повторения следствий в множествах следствий. Поэтому целесообразно искать упрощенные схемы классификации объектов по находящимся в них причинам следствий. Для этого может быть использован аппарат систем различных представителей (СРП). В условиях задачи классификации причин по следствиям нельзя непосредственно воспользоваться теоремой Ф. Холла об СРП, так как нельзя разрывать элементы цепочек причинно-следственных связей. В статье показано, что преобразование каждой из одинаковых цепочек причинно-следственных связей в один общий новый элемент в множествах следствий формирует возможности применения условий теоремы Ф. Холла.

**Ключевые слова:** причинно-следственные связи; конечная классификация; поиск свойств в ненаблюдаемых данных

DOI: 10.14357/19922264230310

EDN: AKWBZD

## 1 Введение

Задача конечной классификации обычно рассматривается как отображение множества объектов  $D$  в конечное множество классов  $K$ . Такое отображение порождает разбиение множества  $D$  на конечное число классов. Существуют ситуации, когда требуется другое определение классов вплоть до возможности пересечения классов. Если не оговорено иное, то под классификацией будем понимать отображение объектов в множество классов. Решение задачи классификации основано на использовании структуры объектов, которая позволяет вычислить (вычленив) такие характеристики объекта, которые соответствуют описанию соответствующего класса. Поэтому можно рассматривать задачу классификации как вычисление следствий характеристик объекта, где в качестве следствий выступают характеристики класса. В этом случае характеристики объекта классификации можно рассматривать как причину, которая детерминиро-

ванно (классификация как отображение) порождает следствия (характеристики класса).

Применение причинно-следственных связей в задачах классификации исследовалось во многих научных работах. Сюда надо прежде всего отнести прикладные исследования и теоретические обоснования использования причинно-следственных связей в задачах медицинской диагностики. В работах [1–3] построены модели каузального вывода в медицинской диагностике. В [4] построена модель использования причинно-следственных связей в задачах классификации изображений с применением байесовских сетей и нейронных сетей.

В данной работе предмет классификации рассматривается как объект, обладающий одним из нескольких возможных ненаблюдаемых множеств характеристик. Каждое такое множество характеристик служит причиной появления одного или нескольких следствий, и любой алгоритм классификации принимает решение о классе по наблюдениям следствий из анализируемой причины.

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, grusho@yandex.ru

<sup>2</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, info@itake.ru

<sup>3</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, m.zabehailo@yandex.ru

<sup>4</sup>ПАО Сбербанк России, dvismirnov@sberbank.ru

<sup>5</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, eltimon@yandex.ru

Использование причинно-следственных связей в классификации основано на двух свойствах причин и следствий. Первое свойство состоит в том, что появление причины детерминированно влечет появление следствия в случае конструктивного взаимодействия причины с пространством, в котором может возникнуть следствие. Второе свойство означает, что нет свойств, появление которых не имеет причин. Если причина искажена, то следствие либо не возникает, либо обусловлено другими причинами.

## 2 Модель причинно-следственных связей для объектов

Будем считать, что существует конечное множество  $U$  характеристик. Любой объект  $O$  определяется как некоторое подмножество множества  $U$ . Пусть задано некоторое множество объектов  $O$  и только объекты этого множества могут появляться для анализа. Свойство  $A$  объекта  $O$  представляет собой некоторое подмножество характеристик объекта  $O$ , при этом свойство  $A$  может не быть объектом. Состав характеристик свойства  $A$  может быть неизвестен. Пусть необходимо неоднократно классифицировать поступающие объекты  $O$  из  $O$  на предмет наличия в них одного из свойств  $A_1, A_2, \dots, A_k$  в предположении, что в каждом классифицируемом объекте содержится ровно одно из этих свойств.

По определению свойство  $A$  в объекте  $O$  служит причиной появления следствия  $B$ , которое доступно наблюдению в информационном пространстве  $I$ , если характеристики  $A$  могут породить объект в пространстве  $I$ , содержащий следствие  $B$ , и в этом случае при появлении  $A$  детерминированно появляется  $B$ .

В рассматриваемом случае каждое из свойств  $A_1, A_2, \dots, A_k$  служит причиной детерминированного появления непустого множества своих следствий  $\{B(A_i, j), j = 1, \dots, m_i\}$ ,  $i = 1, \dots, k$ , в информационных пространствах  $\{I(A_i, j), j = 1, \dots, m_i\}$ ,  $i = 1, \dots, k$ . Допускается, что у некоторых следствий может быть несколько причин, т.е. некоторые следствия, порождаемые разными причинами, могут совпадать.

## 3 Определение множеств следствий с помощью обучения

Рассмотрим задачу построения множеств следствий в условиях, когда для каждого  $A_i$  определена

обучающая выборка объектов  $S_i \subseteq O$ , содержащих это свойство. Для каждого объекта  $S$  из обучающей выборки  $S_i$  известны взаимодействия причины  $A_i$  с информационными пространствами  $I(A_i, j)$ ,  $j = 1, \dots, m_i$ , в которых можно наблюдать следствия этой причины. Из детерминированности появления следствий причины  $A_i$  в этих пространствах  $I(A_i, j)$ ,  $j = 1, \dots, m_i$ , получаем одинаковый набор следствий  $V_i = \{B(A_i, j), j = 1, \dots, m_i\}$ . В доступных для  $A_i$  информационных пространствах, где этот набор не появился, причина  $A_i$  не имеет следствий.

Если объект  $O$  не принадлежит ни одной обучающей выборке, но обладает свойством  $A_i$ , то он также порождает множество следствий  $V_i = \{B(A_i, j), j = 1, \dots, m_i\}$ . Отсюда следует возможность классифицировать объекты, не принадлежащие обучающим выборкам. Если число классов велико, так же как множества следствий каждого  $A_i$ , то задача классификации может оказаться трудно вычислимой из-за того, что возможны повторения следствий в множествах следствий. Поэтому целесообразно искать упрощенные схемы классификации объектов по находящимся в них причинах следствий. Для этого может быть использован аппарат СРП.

## 4 Системы различных представителей

Рассмотрим следующую конечную теоретико-множественную схему. Пусть даны  $k$  подмножеств  $X_1, X_2, \dots, X_k$  конечного множества  $X$ .

**Определение.** Система различных представителей для подмножеств  $X_1, X_2, \dots, X_k$  конечного множества  $X$  — это множество элементов  $\{x_1, x_2, \dots, x_k\}$  этих множеств, таких что  $x_i \in X_i$  и  $x_i \neq x_j$ , если  $i \neq j$ .

Положим  $X_i = V_i$ ,  $i = 1, \dots, k$ , где  $V_i$  — множество следствий причины  $A_i$ .

**Теорема 1.** Если все элементы множества

$$\bigcup_{1 \leq i \leq k} V_i = \bigcup_{1 \leq i \leq k} (B(A_i, j), j = 1, \dots, m_i)$$

различны, то существует СРП и ее элементы однозначно идентифицируют причины  $A_1, A_2, \dots, A_k$ .

**Доказательство.** Для каждой причины  $A_i$  существует по крайней мере одно следствие (это следует из постановки задачи в разд. 2), и все они различные. Тогда представители различных множеств следствий образуют СРП, которая однозначно определяет причины  $A_1, A_2, \dots, A_k$ .



В общем случае можно воспользоваться теоремой Филиппа Холла [5, 6]. Вернемся к обозначениям определения СРП.

**Теорема 2.** Пусть даны  $k$  подмножеств  $X_1, X_2, \dots, X_k$  множества  $X$ . Необходимым и достаточным условием существования СРП, т. е. множества элементов  $\{x_1, x_2, \dots, x_k\}$ , таких что  $x_i \in X_i$  и  $x_i \neq x_j$ , если  $i \neq j$ , выступает условие  $C$ : для каждого  $r = 1, \dots, k$  и каждой последовательности  $r$  различных индексов  $i_1, i_2, \dots, i_r$  в объединении подмножеств  $X_{i_1}, \dots, X_{i_r}$  содержится не менее  $r$  разных элементов.

К сожалению, в условиях задачи классификации причин по следствиям непосредственно воспользоваться этой теоремой нельзя. Если причина  $A$ , состоящая из характеристик множества  $U$ , порождает следствие  $B$  в объекте информационного пространства характеристик  $U^*$ , а  $B$  как причина порождает в объекте информационного пространстве характеристик  $U^{**}$  следствие  $E$ , то из условия детерминированности причинно-следственных связей следует транзитивность причины  $A$  на следствие  $E$ . Это означает, что  $A$  также служит причиной появления следствия  $E$ .

Возможность появления транзитивных следствий у причины  $A$  означает, что разрывать элементы цепочек таких причинно-следственных связей нельзя. Приведем пример.

**Пример.** Пусть необходимо классифицировать объект  $O$  на принадлежность одному из двух классов, определяемых причинами  $A_1$  и  $A_2$ . Предположим, что возможны повторы следствий, при этом  $A_1$  и  $A_2$  служат причинами следствий  $B_1$  и  $B_2$ , а  $B_1$  служит причиной  $B_2$ .

Для множеств  $X_1 = \{B_1, B_2\}$  и  $X_2 = \{B_1, B_2\}$  существует система различных представителей ( $B_1$ ) для идентификации  $X_1$  и ( $B_2$ ) для идентификации  $X_2$ . Однако для идентификации причин  $A_1$  и  $A_2$  построенная СРП не дает однозначной классификации причин  $A_1$  и  $A_2$ , так как в обоих случаях при появлении  $B_1$  автоматически появляется  $B_2$  и их нельзя разделять.

Возможность появления неразделяемых цепочек следствий означает, что такие цепочки следует рассматривать как один элемент. Тогда изменяется число элементов в множествах, участвующих в формулировке теоремы Филиппа Холла, и необходимые условия могут не выполняться.

Рассмотрим эту проблему детальнее. Для выполнения условий теоремы Филиппа Холла необходима возможность формировать любые подмножества исходных множеств и считать число разных элементов. Тогда преобразование каждой из одинаковых цепочек причинно-следственных связей

в один общий новый элемент в множествах следствий формирует возможности проверять условия теоремы Филиппа Холла. Однако по результатам обучения эти цепочки неизвестны. Отсюда следуют три возможности.

1. Если набор следствий причины  $A_i$  встретился в обучении один раз, то любой его элемент может быть взят в систему различных представителей. Это следует из доказательства теоремы 1 и того факта, что все элементы класса, порожденного причиной  $A_i$ , будут обладать одним набором следствий.
2. Если наборы следствий причин  $A_i$  и  $A_j$ ,  $i \neq j$ , обладают одинаковыми подмножествами следствий, а в других наборах следствий в случаях появления ранее встречавшихся следствий  $A_i$  они появляются в составе тех же подмножеств, то все такие подмножества следует заменить на один новый символ. Также надо переобозначить другие множества одинаковых следствий в других множествах следствий.
3. Если в наборах следствий причин  $A_i$  и  $A_j$ ,  $i \neq j$ , встретились разные подмножества одинаковых элементов, то каждое из этих подмножеств следует переобозначить разными буквами, которые можно использовать в построении СРП. Это следует из того, что эти подмножества соответствуют разным транзитивным следствиям одной причины.

После указанных редуций множеств следствий можно исследовать возможности построения СРП на основе условий теоремы Филиппа Холла. Если СРП существует, то она однозначно классифицирует объект в соответствии с обладанием ровно одной из причин  $A_1, A_2, \dots, A_k$ .

Проверка выполнения условий теоремы Филиппа Холла и построение СРП можно осуществить перебором  $2^k - 1$  подмножеств редуцированных множеств следствий. В книге [6] приведены другие алгоритмы поиска СРП, обладающие меньшими оценками сложности.

## 5 Классификация в условиях шума

В работе [6] приведено следствие теоремы Филиппа Холла о числе СРП.

**Следствие.** Пусть даны  $k$  подмножеств  $X_1, X_2, \dots, X_k$  множества  $U$ , которые имеют СРП. Если наименьшее из этих множеств содержит  $t$  элементов, то при  $k \leq t$  существует не меньше чем  $t(t-1) \cdot \dots \cdot (t-k+1)$  различных СРП, а при  $t < k$  существует не меньше чем  $t!$  различных СРП.

Применение этого следствия в задаче классификации причин по следствиям возможно только после редукции множеств следствий, о которой было сказано выше.

В работе [7] рассматривалась бинарная классификация на основе следствий в условиях случайного шума, затрудняющего идентификацию следствий. Используя приведенное следствие теоремы Филиппа Холла, можно создать избыточное множество различных СРП, если это возможно, и использовать различные СРП как взаимно дополняющую информацию для уточнения классификации.

## 6 Заключение

Задачу конечной классификации можно представлять как задачу идентификации набора из нескольких причин по наблюдаемым следствиям. При этом саму классификацию можно построить на системах различных представителей множеств следствий. Это позволяет быстро и с высокой надежностью проводить классификацию объектов на наличие причины из конечного множества.

Определение множеств следствий по обучающей выборке не требует большого числа элементов такой выборки, так как основана на детерминизме появления следствий при появлении причины.

Глубокая и трудоемкая работа требуется для представления задачи классификации в форме идентификации причин по наблюдаемым следствиям. Другая сложная проблема состоит в поиске

возможных следствий выделенных причин в различных информационных пространствах.

Непосредственное применение теорем и алгоритмов поиска СРП осложнено необходимостью учитывать существование транзитивных причин.

## Литература

1. Pearl J. Causal inference // Causality: Objectives and assessment / Eds. I. Guyon, D. Janzing, B. Scholkopf. — Proceedings of machine learning research ser. — Whistler, Canada, 2010. Vol. 6. P. 39–58.
2. Richens J. G., Lee C. M., Johri S. Improving the accuracy of medical diagnosis with causal machine learning // Nat. Commun., 2020. Vol. 11. Art. 3923. 9 p. doi: 10.1038/s41467-020-17419-7.
3. Забейжайло М. И., Грушо А. А., Грушо Н. А., Тимонина Е. Е. Поддержка решения задач диагностического типа // Системы и средства информатики, 2021. Т. 1. № 1. С. 69–81. doi: 10.14357/08696527210106.
4. Zhang C., Zhang K., Li Y. A causal view on robustness of neural networks. — Cornell University, 2021. arXiv: 2005.01095v3 [cs.LG]. 21 p.
5. Hall Ph. On representation of subsets // J. Lond. Math. Soc., 1935. Vol. 10. P. 26–30. doi: 10.1112/JLMS/S1-10.37.26.
6. Халл М. Комбинаторика / Пер. с англ. — М.: Мир, 1970. 424 с. (Hall M. Combinatorial theory. — London: Blaisdell Pub. Co., 1967. 310 p.)
7. Грушо А. А., Забейжайло М. И., Кульченков В. В., Смирнов Д. В., Тимонина Е. Е., Шоргин С. Я. Причинно-следственные связи в задачах анализа ненаблюдаемых процессов // Системы и средства информатики, 2023. Т. 33. № 2. С. 71–78. doi: 10.14357/08696527230207.

Поступила в редакцию 05.06.23

---

---

## CLASSIFICATION BY CAUSE-AND-EFFECT RELATIONSHIPS

A. A. Grusho<sup>1</sup>, N. A. Grusho<sup>1</sup>, M. I. Zabezhailo<sup>1</sup>, D. V. Smirnov<sup>2</sup>, and E. E. Timonina<sup>1</sup>

<sup>1</sup>Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

<sup>2</sup>Sberbank of Russia, 19 Vavilov Str., Moscow 117999, Russian Federation

**Abstract:** By definition, property  $A$  in object  $O$  is the cause for the occurrence of consequence  $B$  which is available for observation in information space  $I$  if characteristics of  $A$  can generate an object in space  $I$  containing consequence  $B$ . In this case,  $B$  determinedly appears with the appearance of  $A$ . Therefore, one can consider the classification problem as calculating the consequences of the characteristics of the object where the consequences act as characteristics of the class. In this case, the characteristics of the classification object can be considered as the cause that deterministically (classification as mapping) generates consequences (characteristics of the class). Each of the properties  $A_i$ ,  $i = 1, \dots, k$ , is the cause of the deterministic appearance of a nonempty set of its consequences. If the number of classes is large as well as the sets of consequences of each, then the classification problem can be complex to compute due to the fact that repetitions of consequences in the sets of consequences are possible. Therefore, it is advisable to look for simplified schemes for classifying objects according to the causes for the consequences in them. For this, an apparatus of systems of various representatives can be used. In the context of the problem of classifying causes due to consequences, it is impossible to directly use F. Hall’s theorem on systems

of various representatives, since elements of cause-and-effect chains cannot be broken. The paper shows that the transformation of each of the same chains of cause-and-effect relationships into one common new element in the sets of consequences forms the possibility of applying the conditions of F. Hall's theorem.

**Keywords:** cause-and-effect relationships; finite classification; searching for the properties in unobservable data

**DOI:** 10.14357/19922264230310

**EDN:** AKWBZD

## References

- Pearl, J. 2010. Causal inference. *Causality: Objectives and assessment*. Eds. I. Guyon, D. Janzing, and B. Scholkopf. Proceedings of machine learning research ser. Whistler, Canada. 6:39–58.
- Richens, J. G., C. M. Lee, and S. Johri. 2020. Improving the accuracy of medical diagnosis with causal machine learning. *Nat. Commun.* 11:3923. 9 p. doi: 10.1038/s41467-020-17419-7.
- Zabekhailo, M. I., A. A. Grusho, N. A. Grusho, and E. E. Timonina. 2021. Podderzhka resheniya zadach diagnosticheskogo tipa [Support for solving diagnostic type problems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(1):69–81. doi: 10.14357/08696527210106.
- Zhang, C., K. Zhang, and Y. Li. 2021. A causal view on robustness of neural networks. *arXiv.org*. 21 p. Available at: <https://arxiv.org/abs/2005.01095v3> (accessed June 28, 2023).
- Hall, P. 1935. On representation of subsets. *J. Lond. Math. Soc.* 10:26–30. doi: 10.1112/JLMS/S1-10.37.26.
- Hall, M. 1967. *Combinatorial theory*. London: Blaisdell Pub. Co. 310 p.
- Grusho, A. A., M. I. Zabekhailo, V. V. Kulchenkov, D. V. Smirnov, E. E. Timonina, and S. Ya. Shorgin. 2023. Prichinno-sledstvennyye svyazi v zadachakh analiza nenablyudaemykh protsessov [Cause-and-effect relationships in analysis of unobservable process properties]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(2):71–78. doi: 10.14357/08696527230207.

*Received June 5, 2023*

## Contributors

**Grusho Alexander A.** (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; [grusho@yandex.ru](mailto:grusho@yandex.ru)

**Grusho Nikolai A.** (b. 1982) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; [info@itake.ru](mailto:info@itake.ru)

**Zabekhailo Michael I.** (b. 1956) — Doctor of Science in physics and mathematics, principal scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; [m.zabekhailo@yandex.ru](mailto:m.zabekhailo@yandex.ru)

**Smirnov Dmitry V.** (b. 1984) — business partner for IT security department, Sberbank of Russia, 19 Vavilov Str., Moscow 117999, Russian Federation; [dvsmirnov@sberbank.ru](mailto:dvsmirnov@sberbank.ru)

**Timonina Elena E.** (b. 1952) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; [eltimon@yandex.ru](mailto:eltimon@yandex.ru)

# О КЛАСТЕРИЗАЦИИ ОБЪЕКТОВ СЕТЕВОЙ ВЫЧИСЛИТЕЛЬНОЙ ИНФРАСТРУКТУРЫ НА ОСНОВЕ АНАЛИЗА СТАТИСТИЧЕСКИХ АНОМАЛИЙ В ТРАФИКЕ\*

А. К. Горшенин<sup>1</sup>, С. А. Горбунов<sup>2</sup>, Д. Ю. Волканов<sup>3</sup>

**Аннотация:** Рассматривается задача выявления статистических аномалий (т. е. существенных превышений от типичных значений полученного и исходящего трафика) нагрузки на узлы сетевой вычислительной инфраструктуры. Рост нагрузки в реальных системах ведет к необходимости регулярного масштабирования вычислительных ресурсов и хранилищ, а также перенаправления потоков данных. Предложена процедура выявления статистических аномалий в сетевом трафике с использованием аппроксимации наблюдений обобщенным гамма-распределением для дальнейшей кластеризации объектов сетевой вычислительной инфраструктуры с целью оценки потребности в ресурсах. Все вычислительные статистические процедуры, описанные в статье, реализованы с использованием языка программирования R и применены к сетевому трафику, полученному в рамках моделирования на специализированном архитектурно-программном стенде. Предложенные подходы могут быть использованы и для более широкого класса телекоммуникационных задач.

**Ключевые слова:** сетевая инфраструктура; сетевой трафик; обобщенное гамма-распределение; вычислительная статистика; проверка статистических гипотез; выявление аномалий; кластеризация

**DOI:** 10.14357/19922264230311

**EDN:** XHTMVI

## 1 Введение

В современной сетевой вычислительной инфраструктуре по мере развития информационных ресурсов растет нагрузка на вычислительные ресурсы инфраструктуры [1, 2]. Этот рост периодически вызывает необходимость масштабирования вычислительных ресурсов и ресурсов хранилищ данных на узлах сетевой вычислительной инфраструктуры или же перенаправления потоков данных. На сегодняшний день эта проблема чаще всего решается в «ручном» режиме на основе опыта сетевых и системных администраторов. Для автоматизации процессов управления масштабированием ресурсов на вычислительных узлах сетевой вычислительной инфраструктуры необходимы методы объективного выявления аномалий в нагрузке на такие узлы и оценки размера аномалий для определения объемов необходимых ресурсов [3–6]. Под аномалиями в данной работе понимаются существенные превышения типичных значений полученного и исходящего трафика. Они могут быть

вызваны, например, работой несанкционированного программного обеспечения на узле или нехваткой ресурсов для обработки пользовательского трафика. Один из возможных путей решения этой проблемы заключается в мониторинге и анализе сетевого трафика в сетевой вычислительной инфраструктуре и выявления в нем статистических аномалий, возникающих в том числе в результате влияния набора случайных факторов [7, 8].

Хорошо известна возможность статистического описания процессов в телекоммуникационных сетях с использованием различных семейств гамма-распределений: классических гамма — для распределений времени обслуживания [9]; гамма–гамма — для аппроксимации некоторых характеристик в сетях сотовой связи [10], подводных коммуникационных системах [11], глобальных вычислительных сетях [12]; конечных гамма-смесей — для описания тонкой структуры информационных потоков [13–17]; обобщенных гамма (Generalized Gamma, GG) — для распределений времени пребывания пользователя в ячейке сотовой сети [18].

\* Работа выполнена при поддержке Программы развития МГУ, проект № 23-Ш03-03. При обработке и анализе трафика использовалась инфраструктура Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

<sup>1</sup> Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук; Московский государственный университет имени М. В. Ломоносова, agorshenin@frcsc.ru

<sup>2</sup> Московский государственный университет имени М. В. Ломоносова; Московский центр фундаментальной и прикладной математики, s.gorbunov.cmc@gmail.com

<sup>3</sup> Московский государственный университет имени М. В. Ломоносова, volkanov@asvk.cs.msu.ru

В работе предложен статистический подход к решению задачи кластеризации объектов сетевой вычислительной инфраструктуры по уровню нагрузки с точки зрения информационного обмена (передаваемого и получаемого трафика) на основе процедуры выявления аномальных наблюдений в сетевом трафике с использованием специального GG-теста. Продемонстрированы примеры применения разработанной методики для данных сетевого трафика, полученных в рамках моделирования на специализированном архитектурно-программном стенде кафедры автоматизации систем вычислительных комплексов факультета вычислительной математики и кибернетики МГУ имени М. В. Ломоносова.

## 2 Статистическая модель трафика на основе обобщенного гамма-распределения

В качестве анализируемых данных в статье используются наборы, полученные на специализированном стенде, на котором моделируются различные сценарии реальных сетевых взаимодействий. Рассматриваются суточные данные (с пятиминутной агрегацией). В каждый момент времени из-

вестно суммарное число отправленных и полученных бит (для удобства дальнейшего анализа данные нормированы значением  $2^{20}$  — около 131 Кбайт). За указанный промежуток времени моделировалось взаимодействие для 1920 объектов в сети. На рис. 1 приведен пример данных входящего трафика для одного из объектов за все время наблюдений, а на рис. 2 — данные сразу для всех объектов, но в некоторый фиксированный момент времени.

Ранее авторами было установлено высокое статистическое согласие данных мобильного трафика сотового оператора с семейством обобщенных гамма-распределений [19] — это сосредоточенное на положительной полупрямой трехпараметрическое семейство вероятностных распределений, определяемое плотностью вида

$$f(x; r, \gamma, \mu) = \frac{|\gamma| \mu^\gamma}{\Gamma(r)} x^{\gamma r - 1} e^{-\mu x^\gamma},$$

где  $x > 0$ ,  $r > 0$ ,  $\mu > 0$ ,  $\gamma \in \mathbb{R} \setminus \{0\}$ . Его выбор может объясняться и тем фактом, что указанное семейство содержит практически все самые популярные абсолютно непрерывные распределения, сосредоточенные на положительной полупрямой, в том числе распределения с тяжелыми хвостами (некоторые примеры приведены в таблице). Естественным образом возникает идея использования

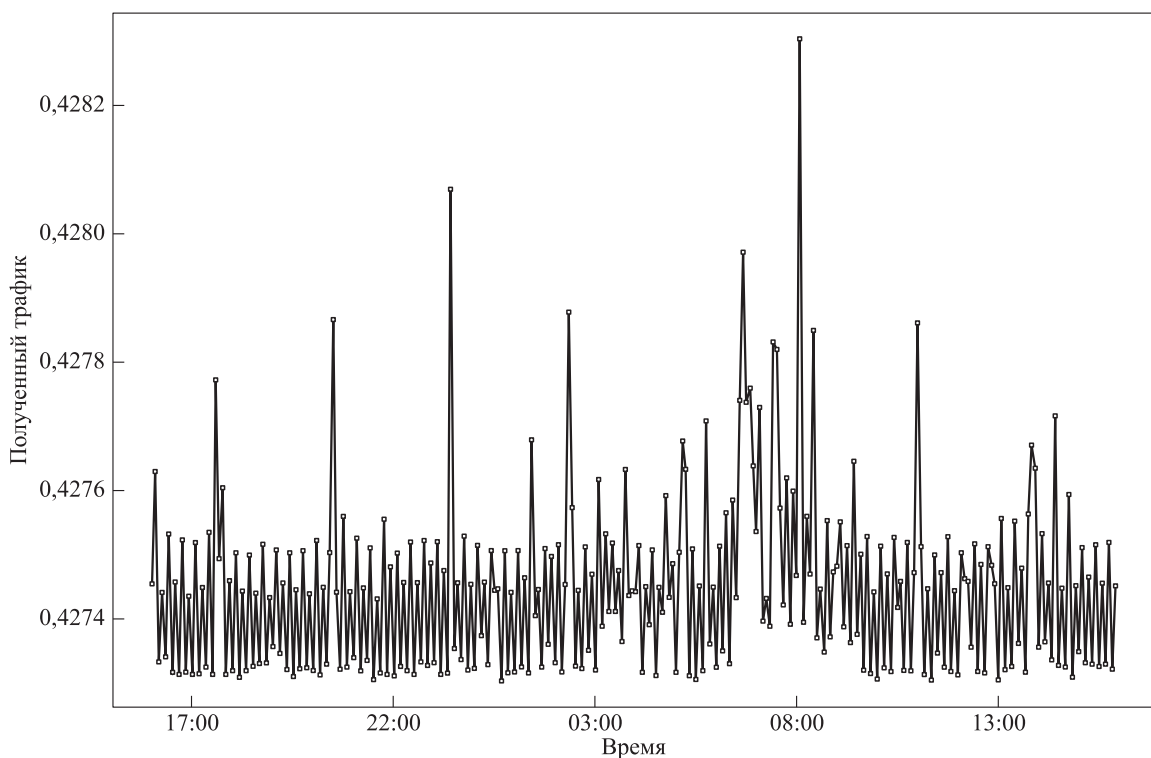


Рис. 1 Временной ряд полученного трафика для одного объекта

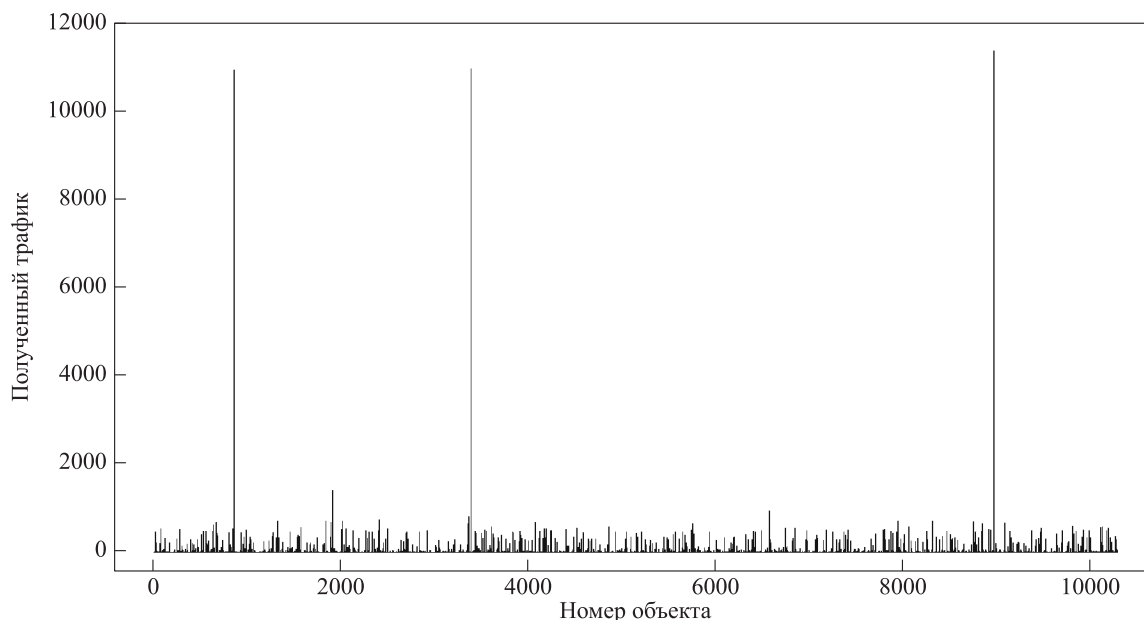


Рис. 2 Выборка полученного графика за одну отметку времени для всех объектов

подобной статистической модели для иных информационных систем.

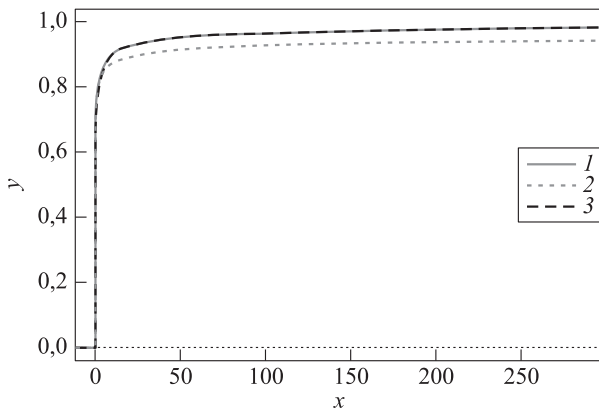
Для оценивания параметров GG-распределения можно применять различные методы, начиная от классического метода максимального правдоподобия. Однако в этом случае для анализируемых данных наблюдались существенные ошибки при аппроксимации хвостов распределения. Поэтому в качестве вычислительной процедуры был реализован алгоритм на основе минимизации  $l^2$ -нормы между значениями эмпирической и теоретической функций распределения, вычисленными в узлах некоторой сетки  $y = (y_1, \dots, y_m)$ . Для оценки параметров по выборке  $\mathbb{X} = (X_1, \dots, X_n)$  решается оптимизационная задача следующего вида:

$$\begin{aligned}
 & (\hat{r}, \hat{\gamma}, \hat{\mu}) = \\
 & = \arg \min_{r>0, \mu>0, \gamma \in \mathbb{R} \setminus \{0\}} \sum_{i=1}^m \left[ \int_0^{y_i} \frac{|\gamma| \mu^r}{\Gamma(r)} x^{\gamma r - 1} e^{-\mu x^\gamma} dx - \right. \\
 & \quad \left. - \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i < y_i) \right]^2.
 \end{aligned}$$

Типичный пример функций, полученных в результате применения различных методов, продемонстрирован на рис. 3. Оптимизационный подход позволил не только точнее, но и примерно в 30 раз быстрее оценивать параметры для всех анализируемых данных по сравнению с методом максимального правдоподобия.

Некоторые частные случаи обобщенного гамма-распределения

Семейство	Значения параметров
Гамма-распределение	$\gamma = 1$
Обратное гамма-распределение	$\gamma = -1$
Распределение Леви	$\gamma = -1, r = 0,5$
Показательное распределение	$\gamma = 1, r = 1$
Распределение Эрланга	$\gamma = 1, r \in \mathbb{N}$
Распределение хи-квадрат	$\gamma = 1, \mu = 0,5$
Распределение Накагами	$\gamma = 2$
Полунормальное распределение	$\gamma = 2, r = 0,5$
Распределение Рэлея	$\gamma = 2, r = 1$
Хи-распределение	$\gamma = 2, \mu = 1/\sqrt{2}$
Распределение Максвелла	$\gamma = 2, r = 1,5$
Распределение Фреше (распределение экстремальных значений II типа)	$r = 1, \gamma < 0$
Распределение Вейбулла–Гнеденко (распределение экстремальных значений III типа)	$r = 1, \gamma > 0$



**Рис. 3** Сравнение эмпирической функции распределения (1) и функций, полученных двумя методами оценки параметров — максимального правдоподобия (2) и функциональной оптимизации (3)

### 3 Метод кластеризации объектов сетевой вычислительной инфраструктуры на основе анализа аномалий трафика

В реальных задачах аномальность трафика определяется не только его абсолютным значением, но и соотношением с другими наблюдениями. В данном разделе рассмотрим процедуру, которая позволяет статистически корректно учитывать это.

Сначала рассмотрим статистическую процедуру выявления аномальных наблюдений для данных, описываемых обобщенным гамма-распределением. Впервые такого рода статистический тест был предложен в статье [20] для метеорологических рядов (осадков). Учитывая особенности анализируемых данных (параметр  $\gamma$  для трафика может иметь как положительные, так и отрицательные значения), обобщим описанную в упомянутой статье процедуру.

Пусть  $V_1, \dots, V_m$  — независимая выборка из обобщенного гамма-распределения с некоторыми параметрами  $r > 0, \gamma \neq 0, \mu > 0, V_1 \geq V_j, \forall j \geq 2$ . Рассмотрим статистику вида

$$\hat{\mathcal{R}} = \left( \frac{(m-1)V_1^\gamma}{V_2^\gamma + \dots + V_m^\gamma} \right)^{\text{sgn}(\gamma)}.$$

Тогда при условии, что верна гипотеза  $H_0$ : «значение  $V_1$  не является аномально большим»,

$$\hat{\mathcal{R}} \sim \begin{cases} F(r, (m-1)r) & \text{для } \gamma > 0; \\ F((m-1)r, r) & \text{для } \gamma < 0, \end{cases}$$

где через  $F$  обозначено распределение Снедекора–Фишера (F-распределение) с соответствующими параметрами.

Для того чтобы выявить аномальные наблюдения в выборке из обобщенного гамма-распределения  $\mathbb{V} = \{V_1, \dots, V_m\}$  с параметрами  $(r, \gamma, \mu)$ , необходимо зафиксировать уровень значимости  $\alpha$ , для каждого  $V_i, i \in \{1, \dots, m\}$ , рассчитать значение статистики  $\hat{\mathcal{R}}$  и сравнить это значение с  $\alpha$ -квантилями F-распределения с соответствующими параметрами. Наблюдение признается аномальным тогда и только тогда, когда наблюдаемое значение статистики больше квантиля F-распределения.

На практике для выборок большого объема подсчет знаменателя статистики  $\hat{\mathcal{R}}$  для каждого наблюдения может заметно увеличить время работы алгоритма, поэтому рациональнее будет один раз посчитать  $S = V_1^\gamma + \dots + V_m^\gamma$ , а затем для  $i$ -го наблюдения вычислять значение статистики следующим образом:

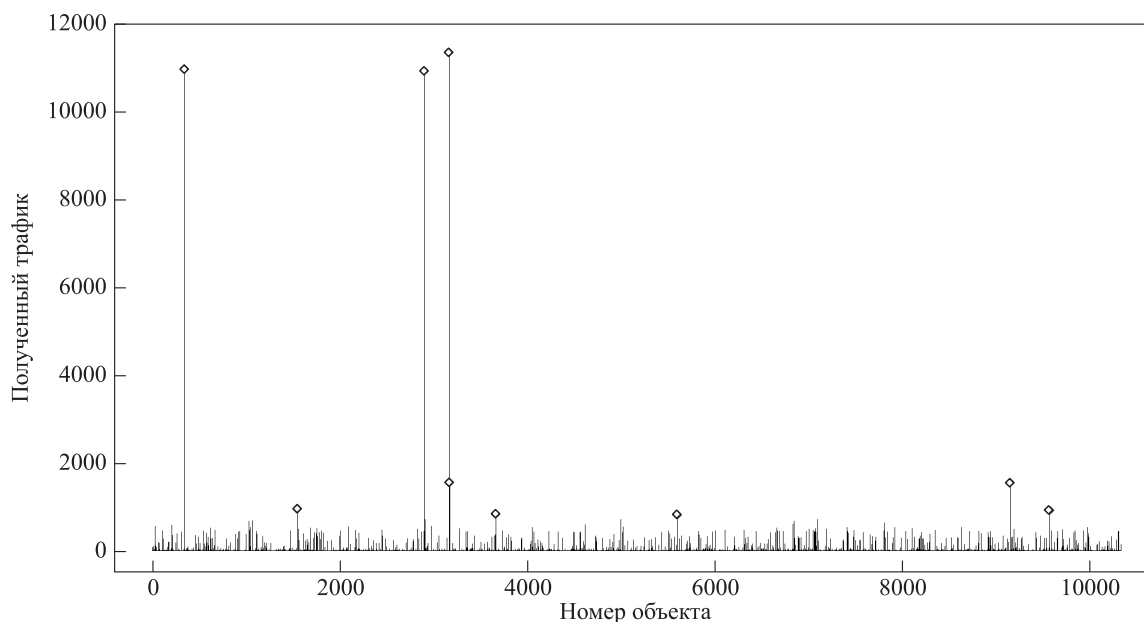
$$\hat{\mathcal{R}} = \left( \frac{(m-1)V_i^\gamma}{S - V_i^\gamma} \right)^{\text{sgn}(\gamma)}.$$

На рис. 4 приведены примеры выявления аномальных наблюдений при уровне значимости  $\alpha = 0,05$  для фиксированного момента времени (выделены маркерами) с помощью описанного метода.

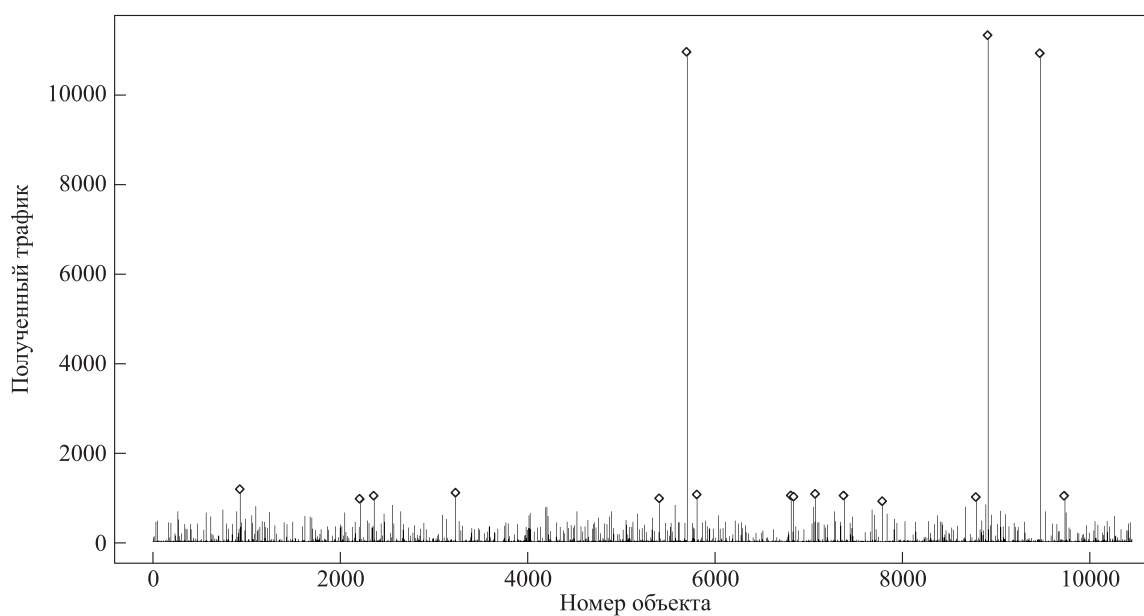
Для кластеризации можно использовать такую разметку наблюдений для выборок за каждый момент времени. Причем формировать выборки можно как на непересекающихся последовательных интервалах времени, так и воспользовавшись методом скользящего окна.

Сначала разметим аномальные наблюдения описанным выше GG-тестом, формируя выборки непересекающимися часовыми окнами, и нанесем получившуюся разметку на временные ряды по конкретным объектам. Если рассматривать разметку GG-теста на выборках, по которым он обучался (рис. 5), то можно заметить, что существует некоторое пороговое значение, разделяющее аномальные и обычные значения.

Однако при рассмотрении той же разметки для конкретного объекта (рис. 6) такого порога может и не оказаться. Так происходит из-за того, что выборки, на которых размечаются аномалии, формируются окнами по времени, поэтому разметка содержит в себе информацию о среднем тренде по всем объектам (рис. 7). На примере объекта с рис. 6 можно выделить две зоны с большим числом аномалий: в районе 22:00, где трафик рос быстрее среднего по всем объектам, и в районе 14:00, где трафик конкретного объекта достиг максимума, а перцентили имели нисходящий тренд. Имея разметку по каждому объекту, можно провести бинарную кластеризацию, задавая некоторый порог доли данных, которые были размечены аномальными.



(а)



(б)

**Рис. 4** Аномалии в полученном (а) и отправленном (б) трафике

Для визуализации полученных результатов составим признаковое описание объектов: каждый объект опишем медианой, квантилями и интерквантильным размахом по полученному и отправленному трафику по всем имеющимся наблюдениям.

Применим метод главных компонент, чтобы перевести признаковое описание на двумерную плоскость. Установлено, что первые две главные компоненты суммарно описывают 95,8% вариации, причем одна приписывает больший вес перценти-

лям (см. ось абсцисс на рис. 8), а другая — интерквантильному размаху (см. ось ординат на рис. 8). На рис. 8 отмечены объекты, признанные аномальными по полученному (1), отправленному (2) и обоим видам трафика (3).

По оси абсцисс на рис. 8 отложены значения первой главной компоненты, которая имеет выраженную положительную корреляцию с перцентилями, по оси ординат — значения второй главной компоненты с выраженной отрицательной корреляцией с размахами. При фиксированном уров-



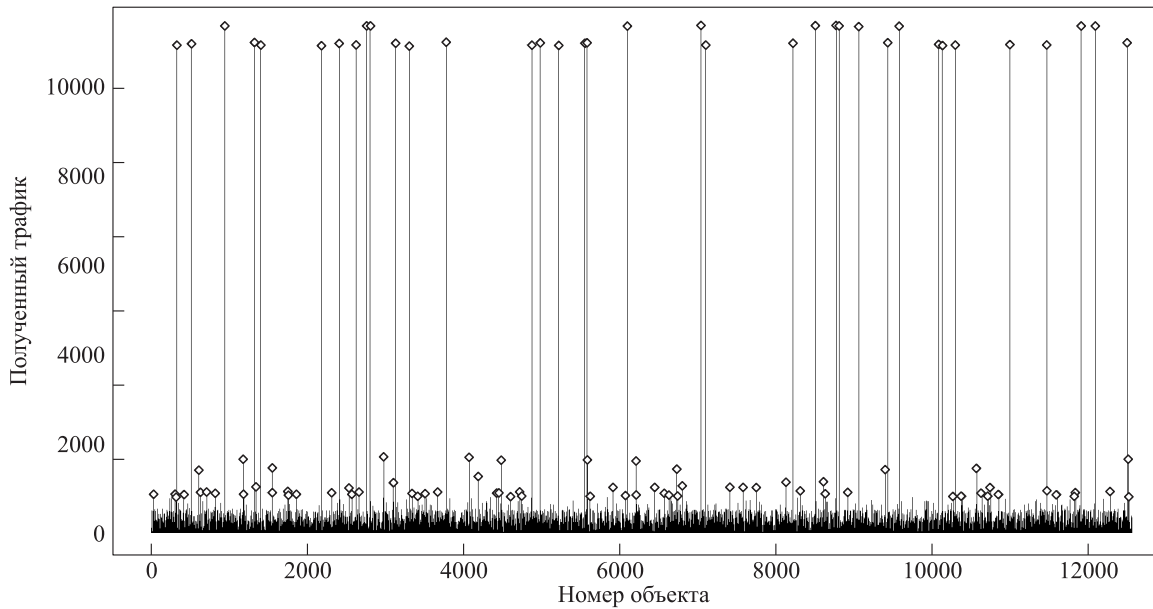


Рис. 5 Аномалии на часовом горизонте наблюдения

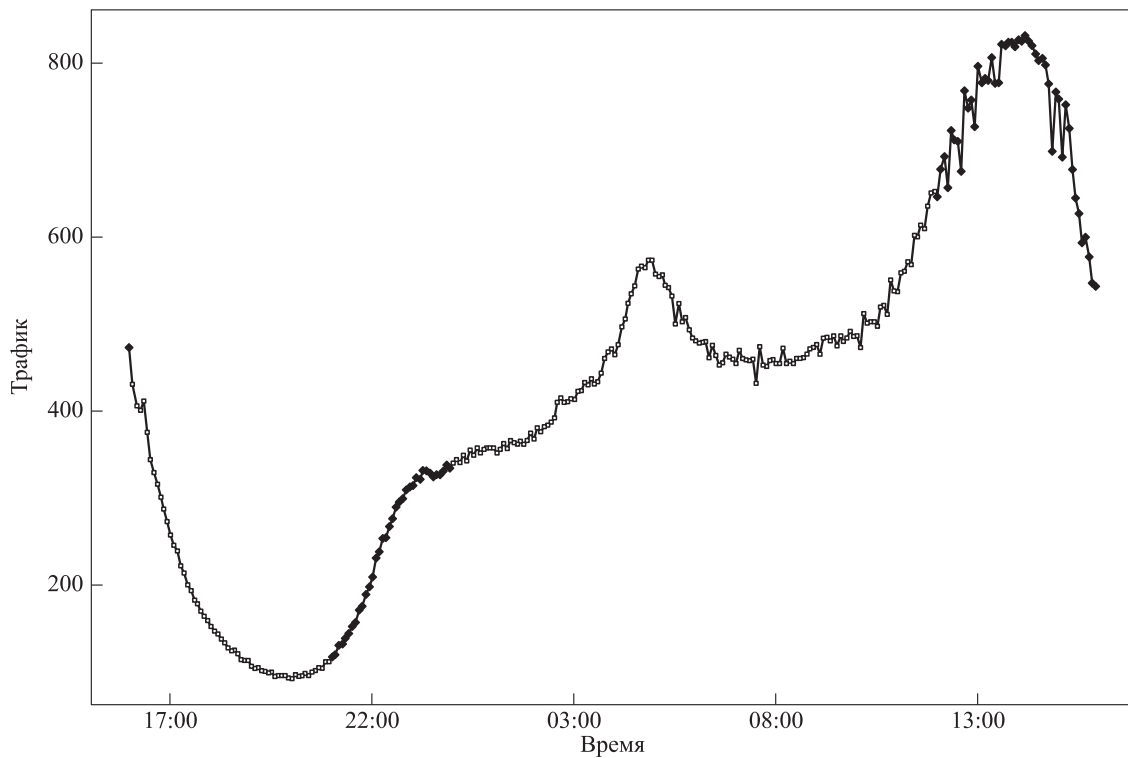


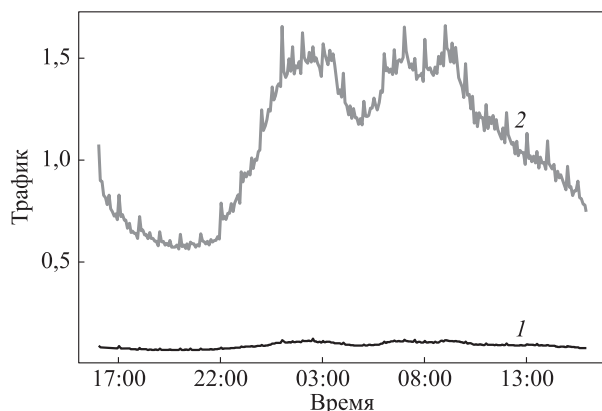
Рис. 6 Аномалии на временном ряде конкретного объекта

не перцентилей (первой компоненты) аномальны объекты, у которых меньше размах, и, наоборот, при фиксированном уровне второй компоненты — у которых больше значения перцентилей.

Недостаток метода непересекающихся окон заключается в том, что каждое наблюдение попадает только в одно окно, поэтому наблюдение для те-

кущего окна может быть признано аномальным, хотя при увеличении интервала времени оно так-вым уже не является (в частности, см. обсуждение подхода в статье [20]) (см. рис. 5).

Если же использовать скользящее окно некоторой ширины  $w$ , то каждое наблюдение (за исключением расположенных на краях) попадет в  $w$  окон.



**Рис. 7** Медиана (1) и верхний квартиль полученного трафика для всех объектов: 2 — 75% перцентиль

Таким образом, наблюдение может быть признано аномальным от 0 до  $w$  раз и можно говорить о некотором сопоставлении уровней аномальности конкретного наблюдения относительно других. Будем считать аномальное наблюдение:

- абсолютно аномальным (признано аномальным ровно  $w$  раз);
- относительно аномальным (признано аномальным от  $\lceil w/2 \rceil$  до  $w - 1$  раз);
- промежуточно аномальным (признано аномальным от 1 до  $\lceil w/2 \rceil - 1$  раз).

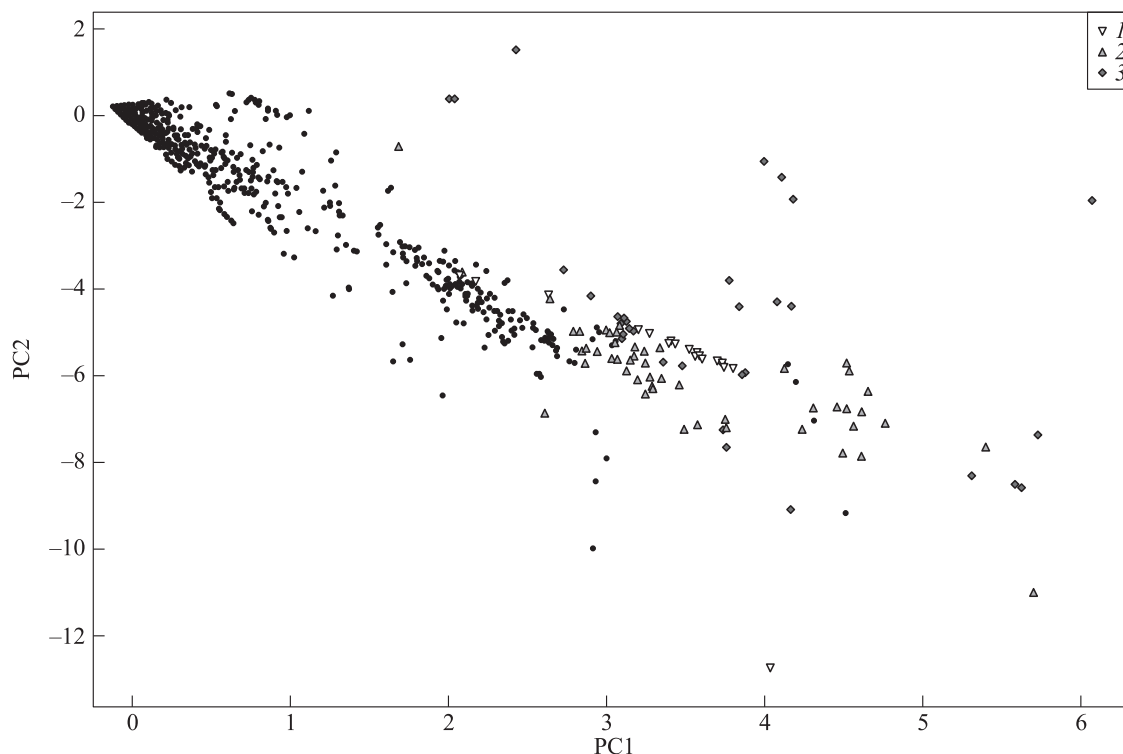
На рис. 9 приведены примеры разметки некоторых объектов в режиме скользящего окна шириной 1 ч.

В этом случае наибольшие по абсолютной величине значения не обязательно признаются аномальными, так как они могут соответствовать периодам общей высокой нагрузки на сеть. Для каждого наблюдения  $V_j$  по каждому объекту имеется некоторый уровень аномальности  $z_j \in \{0, \dots, w\}$ , и в качестве правила кластеризации можно использовать некоторую функцию от  $\{z_1, \dots, z_n\}$ . Например, на рис. 10 представлена кластеризация на основе решающего правила  $\mathbb{I}(\{(1/n) \sum_i z_i > 0,5\})$ .

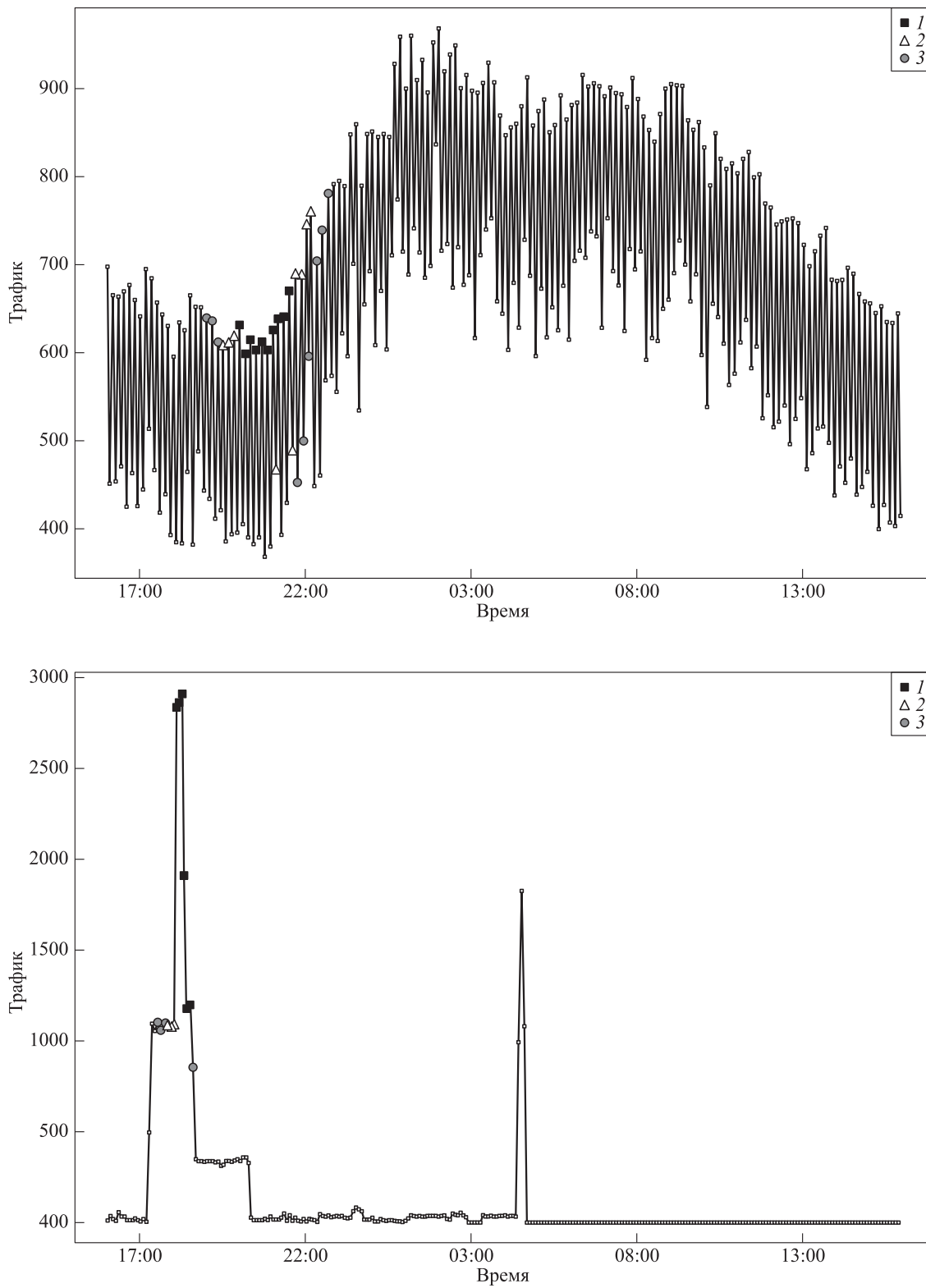
На рис. 8 не были отображены 3 объекта с наибольшими значениями главных компонент, располагающиеся в правом верхнем углу на рис. 10, так как подход на основе непересекающихся окон отнес к аномальным по общему трафику 37 объектов, большая часть из которых продемонстрирована. Метод скользящего окна в качестве аномальных сразу по обоим типам трафика отметил только эти 3 наблюдения, что действительно позволяет говорить об их существенных отличиях от других объектов.

## 4 Заключение

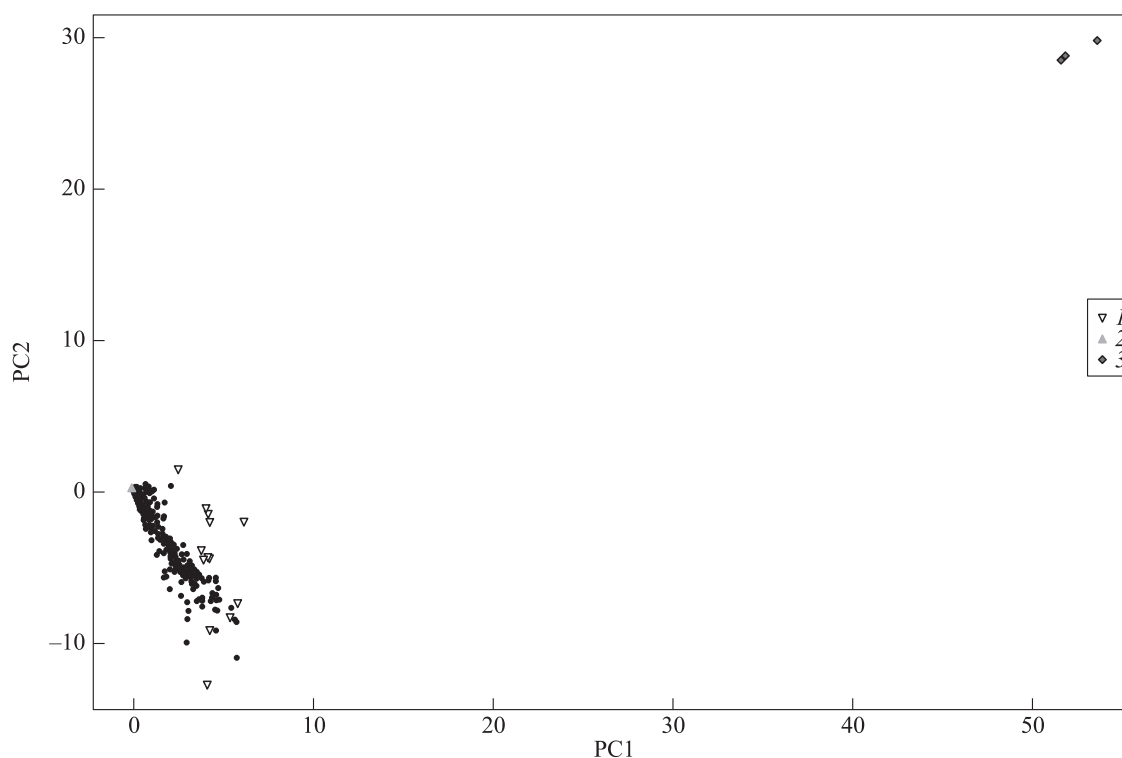
В работе рассмотрен статистический подход к выявлению аномальных нагрузок на узлах сетевой



**Рис. 8** Кластеризация объектов на группы с высоким входящим, исходящим и общим трафиком: 1 — аномален по полученному трафику; 2 — аномален по отправленному трафику; 3 — аномален по обоим типам трафика



**Рис. 9** Примеры разметки временных рядов трафика, скользящее окно: 1 — абсолютно аномальные; 2 — относительно аномальные; 3 — промежуточно аномальные



**Рис. 10** Кластеризация объектов на группы с высоким входящим, исходящим и общим трафиком с использованием скользящего окна: 1 — аномален по полученному трафику; 2 — аномален по отправленному трафику; 3 — аномален по обоим типам трафика

вычислительной архитектуры. С использованием языка программирования R реализована процедура определения аномальных наблюдений в трафике в рамках хорошо соответствующего реальным данным предположения о возможности описания наблюдений обобщенным гамма-распределением. На основе анализа появления таких наблюдений в узлах может быть реализована процедура кластеризации объектов сети. Дальнейшие направления исследований в данной области могут быть связаны с построением статистических моделей для иных характеристик вычислительных узлов (например, загрузки процессора или используемой памяти) и решение описанной задачи кластеризации в расширенном признаковом пространстве.

Предложенная методика имеет определенный потенциал и для других задач в области телекоммуникаций, например связанных с распределением ресурсов для виртуальных машин [21, 22]. Алгоритмы такого рода могут быть реализованы в виде сервисов отдельных аналитических систем [23, 24] или в рамках цифровых платформ [25].

Авторы выражают признательность члену-корреспонденту РАН Р.Л. Смелянскому за ценные советы, касающиеся телекоммуникационной составляющей статьи, а также профессору В. Ю. Королеву за плодотворные обсуждения вопросов мо-

делирования реальных данных с использованием различных семейств вероятностных распределений.

## Литература

1. Смелянский Р.Л. Иерархические периферийные вычисления // Моделирование и анализ информационных систем, 2019. Т. 26. Вып. 1. С. 146–169. doi: 10.18255/1818-1015-2019-1-146-169.
2. Smeliansky R. Network powered by computing // Edge computing — technology, management and integration. — IntechOpen, 2023. 21 p. doi: 10.5772/intechopen.110178.
3. Rossem S.V., Tavernier W., Colle D., Pickavet M., De-meester P. Automated monitoring and detection of resource-limited NFV-based services // Conference on Network Softwarization: Softwarization Sustaining a Hyper-Connected World: en Route to 5G, NetSoft 2017. — Piscataway, NJ, USA: IEEE, 2017. Art. 8004220. doi: 10.1109/NETSOFT.2017.8004220.
4. Malak D., Medard M., Andrews J. G. Spatial concentration of caching in wireless heterogeneous networks // IEEE T. Wirel. Commun., 2021. Vol. 20. Iss. 6. P. 3397–3414. doi: 10.1109/TWC.2021.3049812.
5. Zhang Z., Lu J., Chen H. Controller robust placement with dynamic traffic in software-defined networking // Comput. Commun. 2022. Vol. 194. P. 458–467. doi: 10.1016/j.comcom.2022.07.018.

6. Mesodiakaki A., Zola E., Kassler A. Robust and energy-efficient user association and traffic routing in B5G Het-Nets // *Comput. Netw.*, 2022. Vol. 217. Art. 109305. doi: 10.1016/j.comnet.2022.109305.
7. Hosseinzadeh S., Amirmazlaghani M., Shajari M. An aggregated statistical approach for network flood detection using gamma-normal mixture modeling // *Comput. Commun.*, 2020. Vol. 152. P. 137–148. doi: 10.1016/j.comcom.2020.01.028.
8. Abood M. S., Mustafa A. S., Mahdi H. F., Mohammed A. F. A., Hamdi M. M., Hussein N. A. The analysis of teletraffic and handover performance in cellular system // 3rd Congress (International) on Human–Computer Interaction, Optimization and Robotic Applications Proceedings. — Piscataway, NJ, USA: IEEE, 2021. P. 1–5. doi: 10.1109/HORA52670.2021.9461300.
9. Parulekar M., Makowski A. M.  $M|G|\infty$  input processes: A versatile class of models for network traffic // *IEEE INFOCOM Proceedings*. — Piscataway, NJ, USA: IEEE, 1997. Vol. 2. P. 419–426. doi: 10.1109/INFCOM.1997.644490.
10. Tabassum H., Dawy Z., Hossain E., Alouini M.-S. Interference statistics and capacity analysis for uplink transmission in two-tier small cell networks: A geometric probability approach // *IEEE T. Wirel. Commun.*, 2014. Vol. 13. Iss. 7. P. 3837–3852. doi: 10.1109/TWC.2014.2314101.
11. Noor K., Shahid H., Obaid H. M., Rauf A., Yousaf A., Shahid A. Hybrid underwater intelligent communication system // *Wireless Pers. Commun.*, 2022. Vol. 125. Iss. 3. P. 2219–2238. doi: 10.1007/s11277-022-09653-7.
12. Padhan A. K., Kumar S. H., Sahu P. R., Samantaray S. R. Performance analysis of smart grid wide area network with RIS assisted three hop system // *IEEE Transactions Signal Information Processing Networks*, 2023. Vol. 9. P. 48–59. doi: 10.1109/TSIPN.2023.3239652.
13. Gorshenin A., Korolev V., Kuzmin V., Zeifman A. Coordinate-wise versions of the grid method for the analysis of intensities of non-stationary information flows by moving separation of mixtures of gamma-distribution // 27th European Conference on Modelling and Simulation Proceedings / Eds. W. Rekdalsbakken, R. T. Bye, and H. Zhang. — Dudweiler, Germany: Digitaldruck Pirrot GmbH, 2013. P. 565–568. doi: 10.7148/2013-0565.
14. Gorshenin A., Korolev V. Modelling of statistical fluctuations of information flows by mixtures of gamma distributions // 27th European Conference on Modelling and Simulation Proceedings / Eds. W. Rekdalsbakken, R. T. Bye, and H. Zhang. — Dudweiler, Germany: Digitaldruck Pirrot GmbH, 2013. P. 569–572. doi: 10.7148/2013-0569.
15. Gorshenin A., Kuzmin V. Online system for the construction of structural models of information flows // 7th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings. — Piscataway, NJ, USA: IEEE, 2015. P. 216–219. doi: 10.1109/ICUMT.2015.7382430.
16. Gorshenin A., Kuzmin V. On an interface of the online system for a stochastic analysis of the varied information flows // *AIP Conf. Proc.*, 2016. Vol. 1738. Art. 220009. 4 p. doi: 10.1063/1.4952008.
17. Горшенин А. К. О некоторых математических и программных методах построения структурных моделей информационных потоков // *Информатика и её применения*, 2017. Т. 11. Вып. 1. С. 58–68. doi: 10.14357/19922264170105.
18. Zonozi M. M., Dassanayake P., Faulkner M. Teletraffic modelling of cellular mobile networks // *IEEE VTC P.*, 1996. Vol. 2. P. 1274–1277. doi: 10.1109/VETEC.1996.501517.
19. Stacy E. W. A generalization of the gamma distribution // *Ann. Math. Stat.*, 1962. Vol. 33. Iss. 3. P. 1187–1192. doi: 10.1214/aoms/1177704481.
20. Korolev V. Yu., Gorshenin A. K. Probability models and statistical tests for extreme precipitation based on generalized negative binomial distributions // *Mathematics*, 2020. Vol. 8. Iss. 4. Art. 604. doi: 10.3390/math8040604.
21. Huang B., Chen J., He Q., Wang B., Liu Z., Cheng Y. HASSO: A hot-page aware scheduling optimization method in virtualized NUMA systems // 7th Conference (International) on Information and Communication Systems Proceedings. — Piscataway, NJ, USA: IEEE, 2016. P. 68–73. doi: 10.1109/IACS.2016.7476088.
22. Tian H., Li S., Wang A., Wang W., Wu T., Yang H. Owl: Performance-aware scheduling for resource-efficient function-as-a-service cloud // 13th Symposium on Cloud Computing Proceedings. — New York, NY, USA: Association for Computing Machinery, 2022. P. 78–93. doi: 10.1145/3542929.3563470.
23. Горшенин А. К. Концепция онлайн-комплекса для стохастического моделирования реальных процессов // *Информатика и её применения*, 2016. Т. 10. Вып. 1. С. 72–81. doi: 10.14357/19922264160107.
24. Gorshenin A. K., Kuzmin V. Yu. Research support system for stochastic data processing // *Pattern Recognition Image Analysis*, 2017. Vol. 27. No. 3. P. 518–524. doi: 10.1134/S1054661817030117.
25. Gorshenin A. Toward modern educational IT-ecosystems: From learning management systems to digital platforms // 10th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings. — Piscataway, NJ, USA: IEEE, 2018. P. 329–333. doi: 10.1109/ICUMT.2018.8631229.

Поступила в редакцию 15.07.23

# TOWARD CLUSTERING OF NETWORK COMPUTING INFRASTRUCTURE OBJECTS BASED ON ANALYSIS OF STATISTICAL ANOMALIES IN NETWORK TRAFFIC

A. K. Gorshenin<sup>1,2</sup>, S. A. Gorbunov<sup>2,3</sup>, and D. Yu. Volkanov<sup>2</sup>

<sup>1</sup>Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

<sup>2</sup>M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

<sup>3</sup>Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

**Abstract:** The problem of detecting statistical anomalies (that is, outliers in relation to the typical values of upload and download traffic) of the load on the nodes of the network computing infrastructure is considered. The regular scaling in computing resources and storage as well as redirection of data flows is needed due to the increase of load in real systems. The procedure for detecting statistical anomalies in network traffic is proposed using the approximation of observations by the generalized gamma distribution for further clustering of network computing infrastructure objects in order to evaluate resource need. All computational statistical procedures described in the paper are implemented using the R programming language and they are applied for network traffic, simulated using a specialized architectural and software stand. The proposed approaches can also be used for a wider class of telecommunication problems.

**Keywords:** network infrastructure; network traffic; generalized gamma distribution; computational statistics; statistical hypothesis testing; anomaly detection; clustering

**DOI:** 10.14357/19922264230311

**EDN:** XHTMVI

## Acknowledgments

This work was done with the support of MSU Program of Development, Project No. 23-SCH03-03. The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (CKP “Informatics”) of FRC CSC RAS (Moscow).

## References

1. Smelyansky, R. L. 2019. Ierarkhicheskie periferiynye vy-chisleniya [Hierarchical edge computing]. *Modelirovaniye i analiz informatsionnykh sistem* [Modeling and Analysis of Information Systems] 26(1):146–169. doi: 10.18255/1818-1015-2019-1-146-169.
2. Smeliansky, R. 2023. Network powered by computing. *Edge computing — technology, management and integration*. IntechOpen. 21 p. doi: 10.5772/intechopen.110178.
3. Rossem, S. V., W. Tavernier, D. Colle, M. Pickavet, and P. Demeester. 2017. Automated monitoring and detection of resource-limited NFV-based services. *Conference on Network Softwarization: Softwarization Sustaining a Hyper-Connected World: en Route to 5G*. Piscataway, NJ: IEEE. 8004220. doi: 10.1109/NETSOFT.2017.8004220.
4. Malak, D., M. Medard, and J. G. Andrews. 2021. Spatial concentration of caching in wireless heterogeneous networks. *IEEE T. Wirel. Commun.* 20(6):3397–3414. doi: 10.1109/TWC.2021.3049812.
5. Zhang, Z., J. Lu, and H. Chen. 2022. Controller robust placement with dynamic traffic in software-defined networking. *Comput. Commun.* 194:458–467. doi: 10.1016/j.comcom.2022.07.018.
6. Mesodiakaki, A., E. Zola, and A. Kasser. 2022. Robust and energy-efficient user association and traffic routing in B5G HetNets. *Comput. Netw.* 217:109305. doi: 10.1016/j.comnet.2022.109305.
7. Hosseinzadeh, S., M. Amirmazlaghani, and M. Shajari. 2020. An aggregated statistical approach for network flood detection using gamma-normal mixture modeling. *Comput. Commun.* 152:137–148. doi: 10.1016/j.comcom.2020.01.028.
8. Abood, M. S., A. S. Mustafa, H. F. Mahdi, A.-F. A. Mohammed, M. M. Hamdi, and N. A. Hussein. 2021. The analysis of teletraffic and handover performance in cellular system. *3rd Congress (International) on Human–Computer Interaction, Optimization and Robotic Applications Proceedings*. Piscataway, NJ: IEEE. 1–5. doi: 10.1109/HORA52670.2021.9461300.
9. Parulekar, M., and A. M. Makowski. 1997.  $M|G|\infty$  input processes: A versatile class of models for network traffic. *IEEE INFOCOM Proceedings*. Piscataway, NJ: IEEE. 2:419–426. doi: 10.1109/INFCOM.1997.644490.
10. Tabassum, H., Z. Dawy, E. Hossain, and M.-S. Alouini. 2014. Interference statistics and capacity analysis for

- uplink transmission in two-tier small cell networks: A geometric probability approach. *IEEE T. Wirel. Commun.* 13(7):3837–3852. doi: 10.1109/TWC.2014.2314101.
11. Noor, K., H. Shahid, H. M. Obaid, A. Rauf, A. Yousaf, and A. Shahid. 2022. Hybrid underwater intelligent communication system. *Wireless Pers. Commun.* 125(3):2219–2238. doi: 10.1007/s11277-022-09653-7.
  12. Padhan, A. K., S. H. Kumar, P. R. Sahu and S. R. Samantaray. 2023. Performance analysis of smart grid wide area network with RIS assisted three hop system. *IEEE Transactions Signal Information Processing Networks* 9:48–59. doi: 10.1109/TSIPN.2023.3239652.
  13. Gorshenin, A., V. Korolev, V. Kuzmin, and A. Zeifman. 2013. Coordinate-wise versions of the grid method for the analysis of intensities of non-stationary information flows by moving separation of mixtures of gamma-distribution. *27th European Conference on Modelling and Simulation Proceedings*. Eds. W. Rekdalsbakken, R. T. Bye, and H. Zhang. Dudweiler, Germany: Digitaldruck Pirrot GmbH. 565–568. doi: 10.7148/2013-0565.
  14. Gorshenin, A., and V. Korolev. 2013. Modelling of statistical fluctuations of information flows by mixtures of gamma distributions. *27th European Conference on Modelling and Simulation Proceedings*. Eds. W. Rekdalsbakken, R. T. Bye, and H. Zhang. Dudweiler, Germany: Digitaldruck Pirrot GmbH. 569–572. doi: 10.7148/2013-0569.
  15. Gorshenin, A. K., and V. Kuzmin. 2015. Online system for the construction of structural models of information flows. *7th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings*. Piscataway, NJ: IEEE. 216–219. doi: 10.1109/ICUMT.2015.7382430.
  16. Gorshenin, A., and V. Kuzmin. 2016. On an interface of the online system for a stochastic analysis of the varied information flows. *AIP Conference Proceedings* 1738(1):220009. 4 p. doi: 10.1063/1.4952008.
  17. Gorshenin, A. K. 2017. O nekotorykh matematicheskikh i programnykh metodakh postroeniya strukturnykh modeley informatsionnykh potokov [On some mathematical and programming methods for construction of structural models of information flows]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):58–68. doi: 10.14357/19922264170105.
  18. Zonoozi, M. M., P. Dassanayake, and M. Faulkner. 1996. Teletraffic modelling of cellular mobile networks. *IEEE VTC P. 2*:1274–1277. doi: 10.1109/VETEC.1996.501517.
  19. Stacy, E. W. 1962. A generalization of the gamma distribution. *Ann. Math. Stat.* 33(3):1187–1192. doi: 10.1214/aoms/1177704481.
  20. Korolev, V. Yu., and A. K. Gorshenin. 2020. Probability models and statistical tests for extreme precipitation based on generalized negative binomial distributions. *Mathematics* 8(4):604. doi: 10.3390/math8040604.
  21. Huang, B., J. Chen, Q. He, B. Wang, Z. Liu, and Y. Cheng. 2016. HASO: A hot-page aware scheduling optimization method in virtualized NUMA systems. *7th Conference (International) on Information and Communication Systems Proceedings*. Piscataway, NJ: IEEE. 68–73. doi: 10.1109/IACS.2016.7476088.
  22. Tian, H., S. Li, A. Wang, W. Wang, T. Wu, and H. Yang. 2022. Owl: Performance-aware scheduling for resource-efficient function-as-a-service cloud. *13th Symposium on Cloud Computing Proceedings*. New York, NY: Association for Computing Machinery. 78–93. doi: 10.1145/3542929.3563470.
  23. Gorshenin, A. K. 2016. Kontseptsiya onlayn-kompleksa dlya stokhasticheskogo modelirovaniya real'nykh protsessov [Concept of online service for stochastic modeling of real processes]. *Informatika i ee Primeneniya — Inform. Appl.* 10(1):72–81. doi: 10.14357/19922264160107.
  24. Gorshenin, A. K., and V. Yu. Kuzmin. 2017. Research support system for stochastic data processing. *Pattern Recognition Image Analysis* 27(3):518–524. doi: 10.1134/S10546661817030117.
  25. Gorshenin, A. 2018. Toward modern educational IT-ecosystems: From learning management systems to digital platforms. *10th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings*. Piscataway, NJ: IEEE. 329–333. doi: 10.1109/ICUMT.2018.8631229.

Received July 15, 2023

## Contributors

**Gorshenin Andrey K.** (b. 1986) — Doctor of Science in physics and mathematics, associate professor, principal scientist, head of department, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; agorshenin@frcsc.ru

**Gorbunov Sergei A.** (b. 2000) — master student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; mathematician, Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; shadesilent@yandex.ru

**Volkanov Dmitrii Yu.** (b. 1979) — Candidate of Science (PhD) in physics and mathematics, associate professor, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; volkanov@asvk.cs.msu.ru

# ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ БИНАРНЫХ НЕЙРОННЫХ СЕТЕЙ ПРИ ДЕТЕКТИРОВАНИИ ОБЪЕКТА НА ИЗОБРАЖЕНИИ

Д. О. Королев<sup>1</sup>, О. Г. Малеев<sup>2</sup>

**Аннотация:** Глубокие сверточные нейронные сети широко применяются для задач детектирования объектов. Однако современные модели глубоких сверточных нейронных сетей требуют больших вычислительных ресурсов, что затрудняет их развертывание на мобильных и встроенных устройствах с ограниченными ресурсами. Бинарные нейронные сети позволяют снизить требования к устройствам. В бинарных нейронных сетях активации и/или веса принимают только двоичные значения  $(-1, 1)$ . Представленный в работе метод применяет балансировку и нормализацию целочисленных значений весов при прямом распространении и двухэтапную аппроксимацию функции знака при обратном. Приведены результаты сравнения точности обнаружения на наборе данных PASCAL Face и скорости работы и размера модели на мобильном устройстве для представленного в работе метода, модели без применения бинаризации, сети TinyML и методов Bi-Real Net и ABC-Net.

**Ключевые слова:** бинарные нейронные сети; сверточные нейронные сети; обнаружение объектов; ускорение модели

**DOI:** 10.14357/19922264230312

**EDN:** TOCVAL

## 1 Введение

Вопросам разработки методов, алгоритмов и программно-аппаратных комплексов обработки и классификации изображений с детектированием объектов посвящено достаточно много научных работ [1–4]. Практические решения данной задачи находят широкое применение в системах управления и безопасности, видеоаналитике, системах контроля производства, сельском хозяйстве и т. д. Распространенным подходом к классификации изображений и обнаружению объектов стало применение сверточных нейронных сетей [5]. Благодаря глубокой структуре с множеством уровней и миллионами параметров, глубокие сверточные нейронные сети демонстрируют высокую точность в широком классе задач. Например, сеть VGG-16, содержащая около 140 млн 32-битных параметров с плавающей запятой, достигает точности 92,7% в задаче классификации изображений на наборе данных ImageNet [6]. Вместе с тем такая сеть требует более 500 МБ дискового пространства и для выполнения классификации требует около  $1,6 \cdot 10^{10}$  арифметических операций с плавающей запятой. Понятно, что использование такой архитектуры приводит к необходимости применения достаточно высокопроизводительного оборудования, включа-

ющего, как правило, графические ускорители для проведения матричных вычислений [6].

Предлагаемая в статье реализация алгоритма детектирования объектов с применением бинарных нейронных сетей требует значительно меньшего объема вычислений и памяти, что позволяет проводить детектирование объектов в реальном масштабе времени на встроенных устройствах.

## 2 Бинарные нейронные сети

В бинарных нейронных сетях значения выходов нейронов и/или весов принимают только двоичные значения  $(-1, 1)$ . При этом 32-битные операции умножения и накопления могут быть заменены поразрядными логическими операциями XNOR и операциями подсчета накопления bitcount, заключающимися в подсчете количества единиц [6]. Такой подход позволяет получить значительное увеличение быстродействия и уменьшение размера модели.

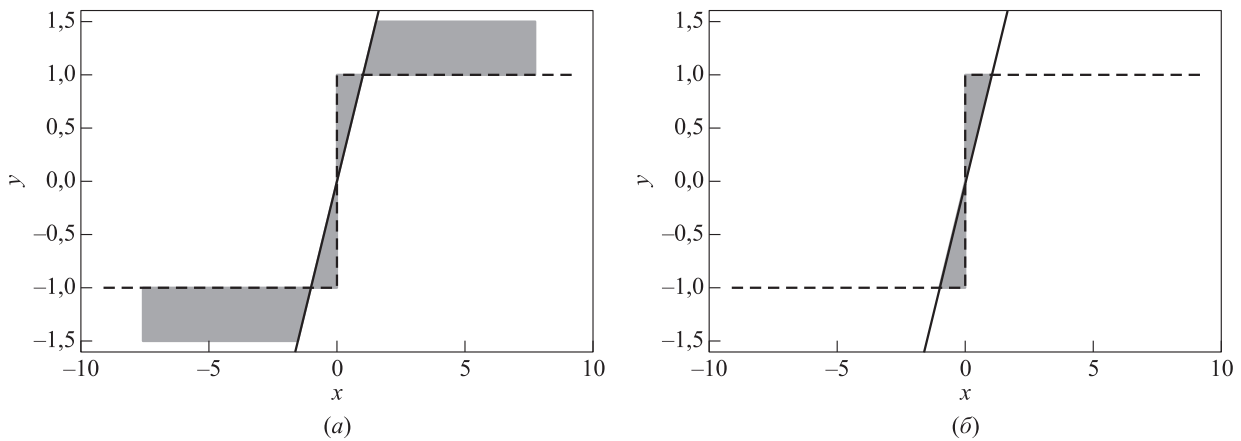
Для прямого распространения в качестве функции активации может быть применена функция знака:

$$\text{sgn}(x) = \begin{cases} 1, & \text{если } x \geq 0; \\ 0, & \text{если } x < 0. \end{cases}$$

<sup>1</sup>Институт компьютерных наук и технологий Санкт-Петербургского политехнического университета Петра Великого, korolev.do512@gmail.com

<sup>2</sup>Институт компьютерных наук и технологий Санкт-Петербургского политехнического университета Петра Великого, maleev.og@spbstu.ru





Форма функций Identity (a) и Hardtanh (б)

Функция знака проста в реализации и неплохо работает на практике, однако ее невозможно использовать при обучении в алгоритме обратного распространения ошибки из-за ее недифференцируемости при  $x = 0$ .

Для решения этой проблемы при обучении применяют функции Identity и Hardtanh [6]. Функция Identity, в отличие от Hardtanh, передает информацию о градиенте без учета эффекта бинаризации. Функция Hardtanh учитывает бинаризацию, однако из-за наличия интервала отсечения отбрасывает информацию за его пределами. Обе эти функции теряют информацию о градиенте и на практике могут приводить к уменьшению точности:

$$\text{Identity} : y = x ;$$

$$\text{Hardtanh} : y = \text{Hardtanh}(x).$$

На рисунке представлена форма функций Identity и Hardtanh [5].

Из вышесказанного понятно, что реализация бинарной нейронной сети — задача нетривиальная. Далее рассмотрим основные этапы предлагаемого авторами метода построения бинарной сети.

### 3 Алгоритм бинаризации

В основе использованного в работе метода лежит способ IR-Net (Information Retention Network) [5]. Для дополнительной минимизации ошибки бинаризации применяется целочисленный битовый сдвиг со значением  $s$  над двоичными весами. Это позволяет расширить возможности представления двоичных весов без необходимости выполнять операции с целочисленными значениями в процессе обучения. При этом бинаризация весов и активаций сети имеет следующий вид:

$$Q_w(\hat{w}_{\text{std}}) = B_w \ll s = \text{sign}(\hat{w}_{\text{std}}) \ll s ;$$

$$Q_a(a) = B_a = \text{sign}(a) .$$

Значение битового сдвига  $s$  определяется как

$$s = \text{round} \left( \log_2 \left( \frac{\|\hat{w}_{\text{std}}\|_1}{n} \right) \right) ,$$

где  $n$  — размерность вектора весов;  $\|\hat{w}_{\text{std}}\|_1$  — L1-норма вектора весов (сумма абсолютных значений вектора).

Основная операция сети для прямого распространения может быть представлена как

$$z = (B_w \odot B_a) \ll s .$$

#### 3.1 Бинаризация весов и активаций при прямом распространении

Для уменьшения влияния ошибок бинаризации в рассматриваемом методе предлагается уравнивать веса, вычитая среднее значение из весов полной точности:

$$\hat{w} = w - \bar{w} .$$

Кроме того, для увеличения стабильности процесса обучения и уменьшения влияния эффекта снижения точности, связанного с чрезмерным ростом величины весов и, следовательно, ростом величины градиента, дополнительно предлагается проводить нормализацию уравновешенных весов:

$$\hat{w}_{\text{std}} = \frac{\hat{w}}{\sigma(\hat{w})} ,$$

где  $\sigma(\cdot)$  — стандартное отклонение.

Эти операции делают веса полной точности равномерно распределенными, что повышает стабильность обучения бинарных нейронных сетей и, как следствие, точность.

Для прямого распространения в качестве функции активации используется функция знака. При этом двоичные значения могут быть получены следующим образом:

$$Q_w(\hat{w}_{std}) = B_w = \text{sign}(\hat{w}_{std});$$

$$Q_a(a) = B_a = \text{sign}(a).$$

### 3.2 Двухэтапная аппроксимация при обратном распространении

Бинарная функция активации может приводить к потере точности, затрудняя передачу градиента при обратном распространении. Для эффективного обучения сети должны быть применены методы, обеспечивающие большую способность к обновлению в начале процесса обучения и небольшую ошибку градиента в конце.

Для обратного распространения предлагается применять двухэтапный способ аппроксимации функции знака [5]:

- (1) на первом этапе для аппроксимации функции знака применяется функция Identity. При этом на первом этапе не проводится отсечения градиентов (значение отсечения очень велико). При переходе ко второму этапу форма аппроксимации изменяется от функции Identity к форме Hardtanh, а значение отсечения постепенно снижается до единицы;
- (2) на втором этапе значение интервала отсечения становится равным единице. Форма аппроксимации постепенно принимает вид функции знака, обеспечивающей согласованность прямого и обратного распространения.

Функция аппроксимации  $g(x)$  для обратного распространения определяется как

$$g(x) = k \tanh tx,$$

где  $g(x)$  — аппроксимация функции знака;  $t$  и  $k$  — управляющие переменные, изменяющие свои значения в процессе обучения. Эти переменные могут быть определены как

$$t = T_{\min} \cdot 10^{(i/N)\log(T_{\max}/T_{\min})};$$

$$k = \max\left(\frac{1}{t}, 1\right),$$

где  $i$  — текущая эпоха;  $N$  — число эпох;  $T_{\min}$  и  $T_{\max}$  — параметры, равные  $10^{-1}$  и  $10^1$ .

## 4 Результаты

Для экспериментальной нейронной сети в работе выбрана широко известная архитектура SSD (single shot detector) [7] с добавлением слоя пакетной нормализации [8]. Модель бинарной сети, реализованная на основе представленного метода, была обучена для детектирования объектов с определением координат ограничивающей рамки [9]. В качестве обучающей выборки использовался набор данных WIDER FACE [10], содержащий 12 881 изображение.

Тестирование производительности бинарной нейронной сети проводилось на мобильном устройстве с процессором MediaTek Helio P22 MT6762 ( $4 \times$  Cortex-A53 2000 МГц,  $4 \times$  Cortex-A53 1500 МГц). Тестирование точности производилось на наборе данных PASCAL Face [11], содержащем 851 изображение и 1341 лицо, ограниченные прямоугольными рамками.

Для развертывания предварительно обученной модели использовался фреймворк daBNN [12], поддерживающий мобильные устройства с операционной системой Android.

Результаты экспериментов в сравнении с ранее известными методами Bi-Real Net [13] и ABC-Net [14] и применении автоматического преобразования модели с использованием фреймворка TensorFlow Lite приведены в таблице.

Результаты работы моделей

Модель	Точность, %	Время работы, мс	Размер, МБ
Полной точности	98,61	1811	89,69
Предложенный метод	98,04	223	11,91
Bi-Real Net	97,98	254	12,56
ABC-Net	97,02	240	11,64
TensorFlow Lite	98,56	645	23,78

## 5 Заключение

Предлагаемый в данной статье метод построения бинарной нейронной сети позволяет уменьшить время обработки изображения в 8,1 раза, а объем памяти, занимаемый моделью, в 7,5 раза. При этом происходит незначительное падение точности обнаружения на тестовом наборе данных PASCAL Face (0,57%).

Реализованный метод превосходит ранее известные методы Bi-Real Net [13] и ABC-Net [14] по точности и имеет лучшее быстродействие.

Также предлагаемый авторами метод продемонстрировал в 2,8 раза лучшее быстродействие и более чем в 2 раза меньшие требования по памяти, чем

распространенный в настоящее время инструмент TensorFlow Lite [15], поддерживающий выполнение квантования весов до 8 бит точности.

Реализованный метод показал высокую точность обнаружения, практически не уступающую полноразмерным сверточным сетям, и может эффективно применяться для построения моделей нейронных сетей, решающих задачи обнаружения объектов, на мобильных и встроенных устройствах с ограниченными ресурсами.

## Литература

1. Pathak A. R., Pandey M., Rautaray S. S. Application of deep learning for object detection // *Procedia Comput. Sci.*, 2018. Vol. 132. P. 1706–1717. doi: 10.1016/j.procs.2018.05.144.
2. Jiao L., Zhang F., Liu F., Yang S., Li L., Feng Z., Qu R. A survey of deep learning-based object detection // *IEEE Access*, 2019. Vol. 7. P. 128837–128868. doi: 10.1109/ACCESS.2019.2939201.
3. Lu S., Wang B., Wang H., Chen L., Linjian M., Zhang X. A real-time object detection algorithm for video // *Comput. Electr. Eng.*, 2019. Vol. 77. P. 398–408. doi: 10.1016/j.compeleceng.2019.05.009.
4. Wu X., Sahoo D., Hoi S. C. Recent advances in deep learning for object detection // *Neurocomputing*, 2020. Vol. 396. P. 39–64. doi: 10.1016/j.neucom.2020.01.085.
5. Qin H., Gong R., Liu X., Shen M., Wei Z., Yu F., Song J. Forward and backward information retention for accurate binary neural networks // *Conference on Computer Vision and Pattern Recognition Proceedings*. — Seattle, WA, USA: IEEE, 2020. P. 2247–2256. doi: 10.1109/CVPR42600.2020.00232.
6. Qin H., Gong R., Liu X., Bai X., Song J., Sebe N. Binary neural networks: A survey // *Pattern Recogn.*, 2020. Vol. 105. Art. 107281. 14 p. doi: 10.1016/j.patcog.2020.107281.
7. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S. E., Fu C., Berg A. C. SSD: Single shot MultiBox detector // *Computer vision / Eds. B. Leibe, J. Matas, N. Sebe, M. Welling*. — Lecture notes in computer science ser. — Cham: Springer, 2016. Vol. 9905. P. 21–37. doi: 10.1007/978-3-319-46448-0\_2.
8. Cai Z., He X., Sun J., Vasconcelos N. Deep learning with low precision by half-wave Gaussian quantization // *Conference on Computer Vision and Pattern Recognition Proceedings*. — IEEE, 2017. P. 5406–5414. doi: 10.1109/CVPR.2017.574.
9. Ren S., He K., Girshick R. B., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks // *IEEE T. Pattern Anal.*, 2015. Vol. 39. P. 1137–1149. doi: 10.1109/TPAMI.2016.2577031.
10. Yang S., Luo P., Loy C. C., Tang X. WIDER FACE: A face detection benchmark // *Conference on Computer Vision and Pattern Recognition Proceedings*. — IEEE, 2016. P. 5525–5533. doi: 10.1109/CVPR.2016.596.
11. Yan J., Zhang X., Lei Z., Li S. Face detection by structural models // *Image Vision Comput.*, 2014. Vol. 32. P. 790–799. doi: 10.1016/j.imavis.2013.12.004.
12. Zhang J., Pan Y., Yao T., Zhao H., Mei T. daBNN: A super fast inference framework for binary neural networks on ARM devices // *27th ACM Conference (International) on Multimedia Proceedings*. — New York, NY, USA: ACM, 2019. P. 2272–2275. doi: 10.1145/3343031.3350534.
13. Liu Z., Wu B., Luo W., Yang X., Liu W., Cheng K. Bi-Real Net: Enhancing the performance of 1-bit CNNs with improved representational capability and advanced training algorithm // *Computer vision / Eds. V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss*. — Lecture notes in computer science ser. — Cham: Springer, 2018. Vol. 11219. P. 747–763. doi: 10.1007/978-3-030-01267-0\_44.
14. Lin X., Zhao C., Pan W. Towards accurate binary convolutional neural network // *31st Conference (International) on Neural Information Processing Systems Proceedings*. — Red Hook, NY, USA: Curran Associates Inc., 2017. P. 345–353.
15. TensorFlow. TensorFlow Lite guide. <https://www.tensorflow.org/lite>.

Поступила в редакцию 22.07.22

## EFFICIENCY OF BINARY NEURAL NETWORKS FOR OBJECT DETECTION ON AN IMAGE

D. O. Korolev and O. G. Maleev

Peter the Great St. Petersburg Polytechnic University, 29 Polytechnicheskaya Str., St. Petersburg 195251, Russian Federation

**Abstract:** Deep convolutional neural networks are widely used for object detection. However, modern deep convolutional neural network models are computationally expensive hindering their deployment in resource-constrained mobile and embedded devices. Binary neural networks help to alleviate the resource requirements of devices. Activations and weights in binary neural networks are limited by binary values  $(-1, 1)$ . The proposed method implements balancing and standardization of floating-point weights in forward propagation and two-stage

sign function approximation in back propagation. The paper presents the results of detection accuracy on the PASCAL Face dataset as well as the results of speed and model size on the mobile device for the proposed method, the model without binarization, the TinyML network, and Bi-Real Net and ABC-Net methods.

**Keywords:** binary neural networks; convolutional neural networks; objects detection; model acceleration

**DOI:** 10.14357/19922264230312

**EDN:** TOCVAL

## References

1. Pathak, A. R., M. Pandey, and S. S. Rautaray. 2018. Application of deep learning for object detection. *Procedia Comput. Sci.* 132:1706–1717. doi: 10.1016/j.procs.2018.05.144.
2. Jiao, L., F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, and R. Qu. 2019. A survey of deep learning-based object detection. *IEEE Access* 7:128837–128868. doi: 10.1109/ACCESS.2019.2939201.
3. Lu, S., B. Wang, H. Wang, L. Chen, M. Linjian, and X. Zhang. 2019. A real-time object detection algorithm for video. *Comput. Electr. Eng.* 77:398–408. doi: 10.1016/j.compeleceng.2019.05.009.
4. Wu, X., D. Sahoo, and S. C. Hoi. 2020. Recent advances in deep learning for object detection. *Neurocomputing* 396:39–64. doi: 10.1016/j.neucom.2020.01.085.
5. Qin, H., R. Gong, X. Liu, M. Shen, Z. Wei, F. Yu, and J. Song. 2020. Forward and backward information retention for accurate binary neural networks. *Conference on Computer Vision and Pattern Recognition Proceedings*. Seattle, WA: IEEE. 2247–2256. doi: 10.1109/CVPR42600.2020.00232.
6. Qin, H., R. Gong, X. Liu, X. Bai, J. Song, and N. Sebe. 2020. Binary neural networks: A survey. *Pattern Recogn.* 105:107281. 14 p. doi: 10.1016/j.patcog.2020.107281.
7. Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg. 2016. SSD: Single shot MultiBox detector. *Computer vision*. Eds. B. Leibe, J. Matas, N. Sebe, and M. Welling. Lecture notes in computer science ser. Cham: Springer. 9905:21–37. doi: 10.1007/978-3-319-46448-0\_2.
8. Cai, Z., X. He, J. Sun, and N. Vasconcelos. 2017. Deep learning with low precision by half-wave Gaussian quantization. *Conference on Computer Vision and Pattern Recognition Proceedings*. IEEE. 5406–5414. doi: 10.1109/CVPR.2017.574.
9. Ren, S., K. He, R. B. Girshick, and J. Sun. 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE T. Pattern Anal.* 39(6):1137–1149. doi: 10.1109/TPAMI.2016.2577031.
10. Yang, S., P. Luo, C. C. Loy, and X. Tang. 2016. WIDER FACE: A face detection benchmark. *Conference on Computer Vision and Pattern Recognition Proceedings*. IEEE. 5525–5533. doi: 10.1109/CVPR.2016.596.
11. Yan, J., X. Zhang, Z. Lei, and S. Li. 2014. Face detection by structural models. *Image Vision Comput.* 32(10):790–799. doi: 10.1016/j.imavis.2013.12.004.
12. Zhang, J., Y. Pan, T. Yao, H. Zhao, and T. Mei. 2019. daBNN: A super fast inference framework for binary neural networks on ARM devices. *27th ACM Conference (International) on Multimedia Proceedings*. New York, NY: ACM. 2272–2275. doi: 10.1145/3343031.3350534.
13. Liu, Z., B. Wu, W. Luo, X. Yang, W. Liu, and K. Cheng. 2018. Bi-Real Net: Enhancing the performance of 1-bit CNNs with improved representational capability and advanced training algorithm. *Computer vision*. Eds. V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss. Lecture notes in computer science ser. Cham: Springer. 11219:747–763. doi: 10.1007/978-3-030-01267-0\_44.
14. Lin, X., C. Zhao, and W. Pan. 2017. Towards accurate binary convolutional neural network. *31st Conference (International) on Neural Information Processing Systems Proceedings*. Red Hook, NY: Curran Associates Inc. 345–353.
15. TensorFlow. TensorFlow Lite guide. Available at: <https://www.tensorflow.org/lite> (accessed July 13, 2023).

Received July 22, 2022

## Contributors

**Korolev Denis O.** (b. 1999) — student, Institute of Computer Science and Technology, Peter the Great St. Petersburg Polytechnic University, 29 Polytechnicheskaya Str., St. Petersburg 195251, Russian Federation; korolev.do512@gmail.com

**Maleev Oleg G.** (b. 1971) — Candidate of Science (PhD) in technology, associate professor, Institute of Computer Science and Technology, Peter the Great St. Petersburg Polytechnic University, 29 Polytechnicheskaya Str., St. Petersburg 195251, Russian Federation; maleev\_og@spbstu.ru

# ФОРМАЛИЗОВАННОЕ ОПИСАНИЕ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ИНФОРМАЦИИ В БАЗАХ ДАННЫХ\*

В. В. Вакуленко<sup>1</sup>, И. М. Зацман<sup>2</sup>

**Аннотация:** Рассматривается последовательность этапов статистической обработки текстовой информации, начиная с конкретных информационных объектов (КИО) баз данных (БД) и заканчивая значениями числовых характеристик множеств этих объектов. Например, если в БД хранятся описания полнотекстовых научных статей, то они считаются КИО. При соответствующем наполнении такой БД многоэтапный процесс их обработки позволяет определить значения числовых характеристик публикационной активности исследователя, научного подразделения или научной организации в целом. Такие процессы начинаются с обработки КИО и завершаются вычислением значений характеристик множеств этих объектов. На промежуточных этапах обработки могут формироваться таблицы и другие вербально-числовые объекты. Если этапы статистической обработки спроектированы как обратимые и в БД реализована функция верификации значений числовых характеристик, то процесс их проверки начинается со значений характеристик и завершается доступом к КИО, которые были использованы для вычисления этих значений. Предлагается формализованное описание этапов статистической обработки текстовой информации в БД. Такую ее трансформацию в числовые значения предлагается назвать информационно-математической (ИМ-трансформация). Она сочетает обработку КИО, формирование вербально-числовых объектов и математические вычисления значений числовых характеристик. Такая трансформация текстовой информации может на отдельных этапах включать математические преобразования, но в целом она к ним не сводится. Цель статьи — предложить принципы формализованного описания ИМ-трансформации текстов в БД. В качестве ее иллюстрации рассмотрен пример формализации процесса определения числа вариантов перевода коннекторов, выражающих внутритекстовые отношения между текстовыми фрагментами в надкорпусной БД (НБД) коннекторов, созданной в ФИЦ ИУ РАН.

**Ключевые слова:** информационно-математическая трансформация; текстовая информация; статистическая обработка текстовой информации; надкорпусная база данных

**DOI:** 10.14357/19922264230313

**EDN:** TYCAEX

## 1 Введение

Процесс статистической обработки текстовой информации в НБД параллельных текстов начинается с *этапа поиска* вхождений исследуемой языковой единицы (ЯЕ) в параллельных текстах. Одна из целей статистической обработки состоит в определении числа вариантов перевода исследуемой ЯЕ, которая может быть словом, словосочетанием, грамматической категорией (например, время глагола) или знаком препинания<sup>3</sup>.

Например, если исследуются варианты перевода на французский язык коннектора<sup>4</sup> *a то* в русско-французских параллельных текстах художественных произведений, то на первом этапе из текстов НБД извлекаются фрагменты (большой частью предложения), содержащие этот коннектор, и их переводы на французский язык (см. три примера

найденных пар фрагментов оригинала и их переводов в табл. 1). Всего в НБД коннекторов, созданной в ФИЦ ИУ РАН, на 08.05.2023 было 144 088 таких пар, каждая из которых — это конкретный информационный объект НБД. Из них 311 КИО содержат коннектор *a то* в русском оригинале, включая три примера, приведенных в табл. 1. Методы поиска исследуемых ЯЕ в НБД описаны в работах [2–4].

На *втором этапе* для найденных пар выполняется формирование аннотируемых переводных соответствий (АПС), каждое из которых включает структурированное описание коннектора *a то*, его контекста, перевода этого коннектора на французский язык и его контекста согласно общим принципам аннотирования [5] и разработанной на их основе методологии [3].

Всего на 08.05.2023 для 220 из 311 найденных пар были сформированы 220 АПС с описанием

\* Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

<sup>1</sup> Федеральный исследовательский центр «Информатика и управление» Российской академии наук; vvak@pm.me

<sup>2</sup> Федеральный исследовательский центр «Информатика и управление» Российской академии наук, izatsman@yandex.ru

<sup>3</sup> Знаку препинания в переводе может соответствовать слово или словосочетание.

<sup>4</sup> Коннекторами называются лексические единицы, основная функция которых состоит в выражении одного или нескольких внутритекстовых отношений между фрагментами текста, например отношения причины [1].

Таблица 1 Примеры результатов поиска фрагментов текста с коннектором *a to* и их переводов на французский язык

Фрагмент с коннектором в оригинале	Перевод
Я вот теперь выехала, чтоб узнать, <b>а то</b> засяду уже безвыездно, если, избави Бог, скарлатина. [Л. Н. Толстой. Анна Каренина (1873–1877)]	Je suis sortie aujourd’hui pour savoir où vous en étiez, <b>car</b> j’ai peur de ne plus pouvoir bouger de longtemps. [Перевод Анри Монго (1952)]
Я нарочно прибрала, чтобы кто не поднял, <b>а то</b> потом поминай как звали! [М. А. Булгаков. Мастер и Маргарита (1929–1940)]	Je l’ai trouvé, dans une serviette, et justement je l’avais mis de côté pour que personne ne la ramasse, <b>sinon</b> , hein, adieu la valise! [Перевод Клода Линьи (1968)]
— Он женится! Хочешь об заклад, что не женится? — возразил он. — Да ему Захар и спать-то помогает, <b>а то</b> жениться! [И. А. Гончаров. Обломов (1848–1859)]	— Lui, se marier! je parie qu’il ne se mariera pas! répliquait-il. Il a besoin de Zakhar même pour dormir, <b>et</b> tu veux qu’il se marie! [Перевод Артюра Адамова (1959)]

Таблица 2 Примеры АПС с коннектором *a to*

Фрагмент с коннектором в оригинале	Коннектор и внутритекстовое отношение	Перевод	Коннектор и внутритекстовое отношение
Я вот теперь выехала, чтоб узнать, <b>а то</b> засяду уже безвыездно, <i>если</i> , избави Бог, скарлатина.	<b>а то</b> (отношение причины)	Je suis sortie aujourd’hui pour savoir où vous en étiez, <b>car</b> j’ai peur de ne plus pouvoir bouger de longtemps.	<b>car</b> (отношение причины)
Я нарочно прибрала, чтобы кто не поднял, <b>а то</b> потом поминай как звали!	<b>а то</b> (отношение отрицательной альтернативы)	Je l’ai trouvé, dans une serviette, et justement je l’avais mis de côté pour que personne ne la ramasse, <b>sinon</b> , hein, adieu la valise!	<b>sinon</b> (отношение отрицательной альтернативы)
— Он женится! Хочешь об заклад, что не женится? — возразил он. — Да ему Захар и спать-то помогает, <b>а то</b> жениться!	<b>а то</b> (отношение несоответствия)	— Lui, se marier! Qu’est-ce que tu paries qu’il ne se mariera pas? Il ne peut même pas s’endormir sans Zakhar, <b>et</b> tu veux qu’il se marie!	<b>et</b> (соединительное отношение)

коннектора *a to* и его переводов (см. в табл. 2 три примера АПС, соответствующих трем парам фрагментов оригинала и их переводов из табл. 1). Отметим, что в этих трех АПС коннектор *a to* выражает три разных внутритекстовых отношения: причина, отрицательная альтернатива, несоответствие. При этом внутритекстовое отношение в оригинале и переводе может быть разным (см. нижнюю строку табл. 2).

Сформированные АПС дают возможность определить на *третьем этапе* число вариантов перевода коннектора *a to* на французский язык. Рассмотренные три этапа процесса статистической обработки текстовой информации описаны вербально (словами естественного языка) на примере перевода коннектора *a to* в НБД.

В статье предлагается подход к формализованному описанию статистической обработки текстовой информации в БД. Ее трансформацию в числовые значения предлагается назвать ИМ-трансформацией. Она сочетает обработку конкрет-

ных текстов, формирование в результате обработки вербально-числовых объектов и вычисление значений числовых характеристик. Такая трансформация текстовой информации в числовые характеристики может на отдельных этапах включать математические преобразования, но в целом она к ним не сводится.

Цель статьи — предложить принципы формализованного описания ИМ-трансформации текстов в БД и ввести понятия: (1) полностью обратимая, (2) частично обратимая и (3) необратимая статистическая обработка текстовой информации.

## 2 Обратимость статистической обработки

Три понятия (полностью обратимая, частично обратимая и необратимая статистическая обработка) опишем на примере определения числа вариантов перевода коннектора *a to*. В предыдущем разделе были рассмотрены три этапа статистиче-

ской обработки текстовой информации, которые сформулируем в общем виде:

- (1) поиск вхождений исследуемой ЯЕ в текстовой информации БД;
- (2) формирование АПС для исследуемых ЯЕ;
- (3) определение числовых статистических характеристик информации, хранимой в БД.

В рассмотренном примере на первом этапе были найдены 311 КИО. Из них для 220 КИО на втором этапе были сформированы 220 АПС. Сейчас более подробно рассмотрим третий этап. В НБД структура АПС включает два отдельных поля: одно для исследуемого коннектора *a to* (см. второй столбец табл. 2) и одно для его перевода, например *car* (см. четвертый столбец табл. 2; если коннектор не был переведен, то ставится специальная метка). В результате сопоставления значений этих двух полей в 220 АПС была сформирована табл. 3 с числом вариантов перевода коннектора *a to* и числом АПС, соответствующих каждому варианту. Всего эти АПС содержат 40 вариантов перевода. Девять вариантов с наибольшим числом АПС, включая вариант отсутствия перевода, включены в первые девять строк табл. 3. Сумма АПС для этих девяти вариантов равна 174. Для остальных вариантов (их в НБД 31) отведена одна десятая строка. Сумма АПС для 31 варианта равна 46.

В НБД этапы статистической обработки информации спроектированы как полностью обратимые, что дает возможность проверки результатов статисти-

ческой обработки, начиная с числовых значений (см. табл. 3) и заканчивая получением списка тех АПС, которые были использованы для вычисления этих значений<sup>1</sup>.

Например, число 11 из четвертой строки табл. 3 в НБД служит гиперссылкой на список из 11 АПС, включая первое АПС из табл. 2, каждое из которых имеет гиперссылку на соответствующий КИО, т. е. пару текстового оригинала с коннектором *a to* и перевода с французским коннектором *car*, включая первую пару из табл. 1. Если бы в АПС эта гиперссылка отсутствовала, то статистическая обработка была бы частично обратимой. А если и числовые значения в табл. 3 не служили бы гиперссылками на списки АПС, то статистическая обработка стала бы полностью необратимой.

### 3 Описание этапов статистической обработки

Статистический анализ текстовой информации — распространенный метод исследований в компьютерной лингвистике, использующий БД и корпуса текстов, включая параллельные [6–9]. Вербальное описание этапов статистического анализа текстовой информации рассмотрим на примере контрастивного исследования пары предлогов *from/fra* в английском и норвежском языке (первый предлог — английский, второй — норвежский) [10] с использованием текстов англо-норвежского параллельного корпуса университета Осло [11].

Статистические характеристики этой пары предлогов определялись для трех видов переводных соответствий<sup>2</sup>: предлог *from* переведен предлогом *fra*, предлог *from* переведен любым другим предлогом, кроме *fra*, и предлог в переводе не использовался [12]. Целью работы [10] ставилось вычисление частотности каждого вида переводного соответствия для каждого смыслового значения предлога *from*. Приведем краткое описание этапов их вычисления из этой работы.

Для определения частотности сначала был сформирован массив текстов, отобранных из англо-норвежского параллельного корпуса. В массив вошли только художественные тексты. В нем был выполнен поиск пар предложений, содержащих предлог *from* в английском оригинале.

Затем был выполнен семантический анализ найденных пар предложений, в результате которого

**Таблица 3** Варианты перевода коннектора *a to* на французский язык (согласно НБД коннекторов ФИЦ ИУ РАН по состоянию на 08.05.2023)

№	Вариант перевода коннектора <i>a to</i>	Число АПС варианта
1	<i>sinon</i>	83
2	<i>коннектор в переводе отсутствует</i>	43
3	<i>et</i>	11
4	<i>car</i>	11
5	<i>mais</i>	9
6	<i>si</i>	5
7	<i>sans quoi</i>	4
8	<i>sans cela</i>	4
9	<i>ou</i>	4
10	31 вариант перевода с числом АПС 3 и меньше	46
Итого	40 моделей перевода	220

<sup>1</sup> Аннотируемые переводные соответствия могут редактироваться. Поэтому после завершения статистической обработки отобранных АПС возможность их редактирования должна быть отключена.

<sup>2</sup> Эти три вида переводного соответствия введены в работе [11] для исследования переводов значений английского предлога *at* на арабский язык.

были экспертно определены контекстные значения предлога *from* и проставлены виды переводных соответствий для *from* (этот предлог переведен предлогом *fra*, другим предлогом или никакой предлог в переводе не использовался). После определения контекстных значений предлога *from* и простановки одного из трех видов были определены их частотности для каждого контекстного значения в сформированном массиве текстов. Например, проведенное исследование показало: если предлог *from* в оригинальных английских предложениях имел значение «причина» (*suffered from high blood pressure*), то в их переводах норвежский предлог *fra* не использовался (нулевая частотность). Если предлог *from* имел значение «источник» (*lights from the cars in the road*), то в 50% переводов использовался предлог *fra*, в 38% использовались другие предлоги и в 12% переводов при переводе *from* предлоги не использовались. Вычисленные частотности каждого вида переводного соответствия для каждого смыслового значения предлога *from* приведены в табл. 4.8 на с. 63 работы [10].

Таким образом, в этой работе дано вербальное описание всех этапов статистической обработки параллельных текстов, включающее таблицу частотностей. В отличие от статистической обработки информации НБД коннекторов, в работе [10] не ставилась задача обеспечить обратимость этапов обработки, т. е. «пройти» от частотности до исходной текстовой информации.

#### 4 Принципы информационно-математической трансформации

Сформулируем предлагаемые принципы формализованного описания многоэтапного процесса статистической обработки текстовой информации в БД. Сначала определяется список этапов статистической обработки и схема связей между ними. Затем для каждого этапа определяются:

- множества объектов (например, текстовых, вербально-числовых и/или числовых) на входе и выходе этапа;
- идентификаторы для доступа в БД к объектам входных и выходных множеств этапа или способ(ы) их определения;
- процессы преобразования входных множеств объектов в выходные;
- классификационные системы, используемые в процессе преобразования;

- возможность доступа к входным множествам объектов этапа либо в процессе обратного преобразования выходных множеств объектов во входные (обратимое преобразование), либо сохранение в БД входных множеств этапа (обратимый этап).

Используя перечисленные принципы, сформулируем описание процесса определения числа вариантов (моделей) перевода *a to* в НБД коннекторов, рассмотренного в начале статьи. Этот процесс включает три этапа, которые выполняются последовательно.

На первом этапе есть только одно входное множество (обозначим его как Input Set 1 — IS1), которое содержит параллельные тексты НБД коннекторов. Эти тексты состоят из пар, включающих русское предложение и его перевод на французский язык. Каждая пара имеет уникальный идентификатор в НБД. На 08.05.2023 было 144 088 идентификаторов таких пар входного множества IS1 (список идентификаторов обозначим как Input List 1 — IL1). Выходное множество на этом этапе формируется также одно (обозначим его как Output Set 1 — OS1), и оно содержит предложения русского языка с коннектором *a to* и их переводы на французский язык (см. табл. 1). На 08.05.2023 было 311 идентификаторов тех пар выходного множества OS1, русские предложения которого содержат коннектор *a to* (список идентификаторов элементов множества OS1 обозначим как OL1).

Преобразование входного множества IS1(IL1) в выходное OS1(OL1) выполняется с помощью поиска тех пар в IS1, которые содержат коннектор *a to*. Его обозначим как  $R1(a\ to)$ :  $IS1(IL1) \rightarrow OS1(OL1)$ . При выполнении этого преобразования классификационные системы не используются. Первый этап обратимый, что обозначим как  $RR1(a\ to)$ , так как в НБД коннекторов сохраняется входное множество этого этапа.

Входным множеством IS2 на втором этапе служит выходное множество OS1 первого этапа, идентификаторы элементов (т. е. 311 пар, русские предложения которых содержат коннектор *a to*) которого образуют список OL1. На этом этапе формируются АПС с использованием методологии аннотирования [3]. Выходное множество (обозначим его как OS2) содержит сформированные АПС (см. табл. 2), идентификаторы которых образуют список OL2. Преобразование входного множества IS2(OL1) в выходное OS2(OL2) выполняется с помощью аннотирования тех пар, которые содержат коннектор *a to*. Его обозначим как  $A2(a\ to)$ :  $IS2(OL1) \rightarrow OS2(OL2)/CS(LSR)$ , где  $CS(LSR)$  — это классификационная система логико-семанти-



ческих отношений, четыре рубрики которой были использованы в трех АПС в табл. 2 (отношения причины, отрицательной альтернативы, несоответствия и соединительное отношение). Второй этап тоже обратимый, что обозначим как  $AR2(a\ to)$ , так как в НБД коннекторов сохраняется входное множество этого этапа.

Входным множеством  $IS3$  третьего этапа служит множество  $OS2$ , которое содержит  $OL2$  сформированных АПС. На этом этапе вычисляется количество вариантов перевода коннектора  $a\ to$  на французский язык (обозначим как  $OL3$ ) и определяется множество  $OS3$ , каждый элемент которого — это число АПС для каждого варианта в НБД коннекторов. Преобразование входного множества  $IS3(OL2)$  в выходное  $OS3(OL3)$  выполняется с помощью сравнения содержимого двух полей АПС (для исследуемого коннектора и его перевода) и определения числа пар этих полей, содержимое которых совпадает. Его обозначим как  $M3(a\ to): IS3(OL2) \rightarrow OS3(OL3)$ . Преобразование  $M2(a\ to)$  третьего этапа обратимо, что обозначим как  $MR3(a\ to)$ , так как число АПС для каждого варианта перевода в НБД коннекторов служит гиперссылкой на список этих АПС. Классификационные системы на этом этапе не используются.

Таким образом, на основе сформулированных принципов ИМ-трансформации было получено следующее формализованное описание трех этапов статистической обработки текстовой информации в НБД коннекторов:

- $RR1(a\ to): IS1(IL1) \rightarrow OS1(OL1)$ ;
- $AR2(a\ to): IS2(OL1) \rightarrow OS2(OL2)/CS(LSR)$ ;
- $MR3(a\ to): IS3(OL2) \rightarrow OS3(OL3)$ .

## 5 Заключение

Предлагаемые принципы описания ИМ-трансформации частично уже использовались при проектировании БД наукометрической информации [13–16], которые применялись для информационного мониторинга при решении задачи программно-целевого управления [17]. Эти принципы дают возможность получать формализованное описание статистической обработки информации в БД до начала ее проектирования. Это описание может быть использовано как условия, которые необходимо выполнить при проектировании БД, если статистическая обработка информации в ней должна быть обратимой. Их выполнение позволит лицам, принимающим решения, верифицировать с помощью БД используемые ими результаты статистической

обработки. Для проектировщиков БД формализованное описание, включающее спецификацию преобразований каждого этапа, которая в статье не рассматривается, задает требование к построению необходимых классификационных систем, а также разграничивает обратимые этапы и обратимые преобразования в БД.

В заключение отметим, что необходимость в предлагаемых принципах описания ИМ-трансформации, возможно, обусловлена теми же причинами, которые Д. А. Поспелов сформулировал при построении логико-лингвистических моделей [18]. Сопоставительный анализ предлагаемых принципов и причин создания этих моделей в будущем позволил бы проверить эту гипотезу.

## Литература

1. Гончаров А. А., Инькова О. Ю. Методика поиска имплицитных логико-семантических отношений в тексте // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 97–104. doi: 10.14357/19922264190314.
2. Зализняк Анна А., Кружков М. Г. База данных безличных глагольных конструкций русского языка // Информатика и её применения, 2016. Т. 10. Вып. 4. С. 132–141. doi: 10.14357/19922264160414.
3. Гончаров А. А., Инькова О. Ю., Кружков М. Г. Методология аннотирования в надкорпусных базах данных // Системы и средства информатики, 2019. Т. 29. № 2. С. 148–160. doi: 10.14357/08696527190213.
4. Нуриев В. А., Кружков М. Г. Корпусные данные при контрастивном изучении пунктуации // Системы и средства информатики, 2023. Т. 33. № 1. С. 14–23. doi: 10.14357/08696527230102.
5. Handbook of linguistic annotation / Eds. N. Ide, J. Pustejovsky. — Dordrecht, The Netherlands: Springer Science + Business Media, 2017. 1468 p.
6. Parallel corpora for contrastive and translation studies: New resources and applications / Eds. I. Doval, M. T. Sánchez Nieto. — Amsterdam, Philadelphia: John Benjamins, 2019. 301 p.
7. Translating and comparing languages: Corpus-based insights / Eds. S. Granger, M.-A. Lefer. — Louvain-la-Neuve, Belgique: Presses universitaires de Louvain, 2020. 298 p.
8. Corpora in translation and contrastive research in the digital age: Recent advances and explorations / Eds. J. Lavid-López, C. Maíz-Arévalo, J. R. Zamorano-Mansilla. — Amsterdam, Philadelphia: John Benjamins, 2021. 351 p.
9. Extending the scope of corpus-based translation studies / Eds. S. Granger, M.-A. Lefer. — London: Bloomsbury Academic, 2022. 288 p.
10. Savchenko E. A contrastive study of the English and Norwegian cognates from and fra: Master's Thesis. — Oslo, Norway: University of Oslo, 2013. 120 p. <https://www.duo.uio.no/handle/10852/37026>.

11. The English–Norwegian Parallel Corpus (ENPC). <https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc>.
12. Hasan A. A., Abdullah I. H. A cross mapping of temporal at-ba “Forward and backward translation” // *English Language Teaching*, 2009. Vol. 2. Iss. 1. P. 80–84. doi: 10.5539/elt.v2n1p80.
13. Зацман И. М. Полидоменные модели в системах оценки инновационного потенциала и результативности научных исследований // *Компьютерная лингвистика и интеллектуальные технологии: Тр. Междунар. конф. Диалог-2006*. — М.: РГГУ, 2006. С. 178–183.
14. Зацман И. М. Полидоменные модели электронных библиотек систем мониторинга сферы науки // *Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Тр. 8-й Всеросс. научн. конф.* — Ярославль: ЯрГУ им. П. Г. Демидова, 2006. С. 75–81.
15. Зацман И. М., Веревкин Г. Ф., Дрынова И. В., Курчавова О. А., Ларин Н. В., Норемян Т. П. Моделирование систем информационного мониторинга как проблема информатики // *Системы и средства информатики*, 2006. Т. 16. № 3. С. 257–278.
16. Зацман И. М., Веревкин Г. Ф., Шубников С. К. Моделирование систем мониторинга. — М.: ИПИ РАН, 2008. 115 с.
17. Зацман И. М., Веревкин Г. Ф. Информационный мониторинг сферы науки в задачах программно-целевого управления // *Системы и средства информатики*, 2006. Т. 16. № 1. С. 185–210.
18. Поспелов Д. А. Логико-лингвистические модели в системах управления. — М.: Энергоиздат, 1981. 231 с.

Поступила в редакцию 20.06.23

## FORMALIZED DESCRIPTION OF STATISTICAL INFORMATION PROCESSING IN DATABASES

V. V. Vakulenko and I. M. Zatsman

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The paper presents an overview of the stages of statistical processing of text data, from specific informational objects in databases to the values of the numerical characteristics of these objects. For example, if the database contains the descriptions of full-text research articles, then they represent specific informational objects. With the appropriate population of such a database, the multistage procedure of their processing makes it possible to determine the values of the numerical characteristics of the publication activity of a researcher, a scientific division, and a scientific organization as a whole. Such procedures begin with the processing of specific informational objects and end with computing of the values of the numerical characteristics of these objects. At intermediate stages, tables and other both verbal and numerical objects may form. If the stages of the statistical processing are designed to be reversible and the database implements the function of verifying the values of the numerical characteristics, then the procedure of their verification begins with the values of the characteristics and ends with access to specific informational objects that were used to compute these values. The paper proposes a formalized description of the stages of statistical processing of text data in databases. Informational-mathematic transformation (IM-transformation) is the proposed name for such transformation of text data into numerical values. It combines the processing of specific informational objects, the formation of verbal and numerical objects, and the mathematical computation of the values of numerical characteristics. Such transformation of text data may include mathematic processes at certain stages; however, it does not completely reverse back to them. The goal of the paper is to propose the principles of formalized description of IM-transformation of texts in databases. To illustrate this, the paper provides the example of formalizing the process of determining the frequency of translation variants of connectives expressing intertextual relations between text fragments in the supracorpora database of connectives developed in the FRC CSC RAS.

**Keywords:** informational-mathematic transformation; text information; statistical processing of text information; supracorpora database

**DOI:** 10.14357/19922264230313

**EDN:** TYCAEX

### Acknowledgments

The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (СКР “Informatics”) of FRC CSC RAS (Moscow).

## References

- Goncharov, A. A., and O. Yu. Inkova. 2019. Metodika poiska implitsitnykh logiko-semanticheskikh otnosheniy v tekste [Methods for identification of implicit logical-semantic relations in texts]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):97–104. doi: 10.14357/19922264190314.
- Zalizniak, Anna A., and M. G. Kruzhev. 2016. Baza danykh bezlichnykh glagol'nykh konstruksiy russkogo yazyka [Database of Russian impersonal verbal constructions]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):132–141. doi: 10.14357/19922264160414.
- Goncharov, A. A., O. Yu. Inkova, and M. G. Kruzhev. 2019. Metodologiya annotirovaniya v nadkorpornykh bazakh danykh [Annotation methodology of supracorpora databases]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(2):148–160. doi: 10.14357/08696527190213.
- Nuriev, V. A., and M. G. Kruzhev. 2023. Korpusnye danye pri kontrastivnom izuchenii punktuatsii [The parallel corpora perspective on studying contrastive punctuation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(1):14–23. doi: 10.14357/08696527230102.
- Ide, N., and J. Pustejovsky, eds. 2017. *Handbook of linguistic annotation*. Dordrecht, The Netherlands: Springer Science + Business Media. 1468 p.
- Doval, I., and M. T. Sanchez Nieto, eds. 2019. *Parallel corpora for contrastive and translation studies: New resources and applications*. Amsterdam, Philadelphia: John Benjamins. 310 p.
- Granger, S., and M.-A. Lefer, eds. 2020. *Translating and comparing languages: Corpus-based insights*. Louvain-la-Neuve, Belgique: Presses universitaires de Louvain. 298 p.
- Lavid-López, J., C. Maíz-Arévalo, and J. R. Zamorano-Mansilla, eds. 2021. *Corpora in translation and contrastive research in the digital age: Recent advances and explorations*. Amsterdam, Philadelphia: John Benjamins. 351 p.
- Granger, S., and M.-A. Lefer, eds. 2022. *Extending the scope of corpus-based translation studies*. London: Bloomsbury Academic. 288 p.
- Savchenko, E. 2013. *A contrastive study of the English and Norwegian cognates from and fra*. Oslo, Norway: University of Oslo. Master's Thesis. 120 p. Available at: <https://www.duo.uio.no/handle/10852/37026> (accessed July 10, 2023).
- The English–Norwegian parallel corpus (ENPC). Available at: <https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc> (accessed July 10, 2023).
- Hasan, A. A., and I. H. Abdullah. 2009. A cross mapping of temporal at-ba “Forward and backward translation.” *English Language Teaching* 2(1):80–84. doi: 10.5539/elt.v2n1p80.
- Zatsman, I. M. 2006. Polidomennyye modeli v sistemakh otsenki innovatsionnogo potentsiala i rezul'tativnosti nauchnykh issledovaniy [Polydomain models for evaluation systems of innovative potential and performance of researches]. *Computational Linguistics and Intellectual Technologies: Conference (International) “Dialog 2006” Proceedings*. Moscow: RGGU. 178–183.
- Zatsman, I. M. 2006. Polidomennyye modeli elektronnykh bibliotek sistem monitoringa sfery nauki [Polydomain models for digital libraries of monitoring systems in scientific sphere]. *Elektronnyye biblioteki: perspektivnyye metody i tekhnologii, elektronnyye kolleksii: Tr. 8-y Vseross. nauchn. konf.* [Digital Libraries: Advanced Methods and Technologies, Digital Collections. 8th All-Russian Research Conference Proceedings]. Yaroslavl': P. G. Demidov Yaroslavl' State University. 75–81.
- Zatsman, I. M., G. F. Verevkin, I. V. Drynova, O. A. Kurchavova, N. V. Larin, and T. P. Norekhan. 2006. Modelirovaniye sistem informatsionnogo monitoringa kak problema informatiki [Model for systems of information monitoring as a problem of informatics]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 16(3):257–278.
- Zatsman, I. M., G. F. Verevkin, and S. K. Shubnikov. 2008. *Modelirovaniye sistem monitoringa* [Modeling of monitoring systems]. Moscow: IPI RAN. 115 p.
- Zatsman I. M., and G. F. Verevkin. 2006. Informatsionnyy monitoring sfery nauki v zadachakh programmno-tselevogo upravleniya [Information monitoring for scientific sphere in programme budgeting problems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 16(1):185–210.
- Pospelov, D. A. 1981. *Logiko-lingvisticheskie modeli v sistemakh upravleniya* [Logical-linguistic models in control systems]. Moscow: Energoizdat. 231 p.

Received June 20, 2023

## Contributors

**Vakulenko Vasily V.** (b. 1995) — engineer-researcher, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; [vvak@pm.me](mailto:vvak@pm.me)

**Zatsman Igor M.** (b. 1952) — Doctor of Science in technology, head of department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; [izatsman@yandex.ru](mailto:izatsman@yandex.ru)

# КРИТЕРИИ ОПРЕДЕЛЕНИЯ СЕМАНТИЧЕСКОЙ БЛИЗОСТИ ДИСКУРСИВНЫХ ОТНОШЕНИЙ\*

О. Ю. Инькова<sup>1</sup>, М. Г. Кружков<sup>2</sup>

**Аннотация:** Работа посвящена результатам разработки структурированных определений дискурсивных отношений на основе их классификации, а также критериям, позволяющим определить их семантическую близость. Авторы показывают недостатки существующих подходов, которые приводят к противоречивым или часто необоснованным результатам, а также раскрывают преимущества альтернативного решения: классификации дискурсивных отношений на основе их структурированных определений. Приводятся примеры таких определений, сформированных в Надкорпусной базе данных коннекторов (НБДК), а также критерии, позволяющие определить семантическую близость дискурсивных отношений. Поскольку структурированные определения представляют собой набор различительных признаков, авторы обсуждают проблему присвоения коэффициента близости каждому из признаков. Полученные данные, в том числе количественные, позволяют выдвинуть гипотезу, согласно которой из трех групп признаков: «Уровень», «Базовая операция» и «Семейство признаков» — наибольший вес имеет последняя. Предлагаются пути дальнейшего исследования этой проблемы, в частности с учетом таких факторов, как данные по сочетаемости дискурсивных отношений, по соответствиям дискурсивных отношений и их показателей в тексте оригинала и в тексте перевода, а также тех случаев, когда один показатель может выражать несколько дискурсивных отношений.

**Ключевые слова:** надкорпусная база данных; логико-семантические отношения; коннекторы; аннотирование; фасетная классификация

**DOI:** 10.14357/19922264230314

**EDN:** UJZJZI

## 1 Определения дискурсивных отношений и вопросы их семантической близости

Вопросы, связанные с изучением дискурсивных отношений и их показателей, коннекторов, не теряют своей актуальности в силу того, что эти отношения играют значительную роль в обеспечении связности текста [1–3], а список их показателей постоянно расширяется (ср. для русского языка, например, [4, с. 663–686; 5] и др.), что затрудняет автоматическую обработку текста. Кроме того, в существующих подходах дискурсивные отношения определены, как правило, довольно противоречиво, поскольку нет четких критериев для их выделения (см., например, наиболее известную в данной области теорию риторических отношений [6], а также [7–9]).

В некоторых классификациях [10, 11] отношения оформлены в виде иерархической структуры, однако критерии объединения отношения в группы и, в частности, критерий выделения того или иного числа высших иерархических уровней, от-

сутствуют. Так, в классификации, используемой в Пенсильванском аннотированном корпусе (Penn Discourse Treebank, RDTB) [10], высших классифицирующих уровней четыре: Temporal, Contingency, Comparison, Expansion. Если основания для выделения уровней Temporal и Contingency интуитивно ясны, то основания для объединения отношений в классы Comparison и Expansion неочевидны.

Вызывает, например, возражение отнесение к группе Comparison уступительных отношений (*Хотя на улице дождь, Петя не захотел взять зонтик*), в основе интерпретации которых лежит не сравнение, а импликация (Если на улице идет дождь, то мы, как правило, берем зонтик), а значит, им место в классе Contingency вместе с причинными и условными отношениями (подробнее см. [12]).

В результате в одном классе оказываются семантически далекие отношения: например, в классе Expansion оказывается отношение альтернативы и спецификации; ср. *Она учится в университете или работает?* с отношением альтернативы и *Подари ей четвероного друга, например хомьяка* с отношением спецификации.

\* Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Женевский университет, oiyainkova@yandex.ru

<sup>2</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, magnit75@yandex.ru

И наоборот: семантически близкие отношения оказываются в разных группах, как в случае уступительных и условных отношений. Отсутствие критериев для выделения иерархических уровней и отнесения к ним того или иного набора логико-семантических отношений (ЛСО) приводит также к тому, что классификация, используемая в PDTB, претерпевает постоянные изменения; ср. две ее версии, представленные в [10] и [13].

Основные проблемы классификации дискурсивных отношений описаны в предыдущей работе авторов [14], где представлены также первые результаты ее альтернативного решения, в частности классификация, используемая в НБДК, разработанной в ИПИ ФИЦ ИУ РАН<sup>1</sup>, и созданные на ее основе структурированные определения дискурсивных отношений, или, в терминологии авторов, ЛСО. Коротко перечислим основные положения данного подхода.

В основе классификации лежат четыре базовые семантические операции, на которые опирается то или иное ЛСО: импликация; расположение на шкале времени; сравнение; соотнесение частного и общего или элемента и множества. Классификация различает также уровни, на которых может быть установлено ЛСО: пропозициональный уровень, уровень высказывания (иллокутивный), метаязыковой (подробнее см. [12]). Каждое ЛСО определяется, следовательно, на основе этих двух критериев, к которым добавляется критерий, характеризующий ЛСО, основанные на импликации и сравнении: полярность, т. е. устанавливается ли ЛСО непосредственно между положениями вещей  $p$  и  $q$ , описанными в связываемых им фрагментах текста, или же при его интерпретации должны быть учтены также их отрицательные корреляты  $\neg p$  и  $\neg q$ . Учитываются и семантические, и прагматические характеристики контекста.

## 2 Структурированные определения логико-семантических отношений

Разработанная концепция классификации дает возможность описывать ЛСО при помощи структурированных определений, представляющих собой набор различительных признаков. На момент написания статьи в НБДК описаны 26 ЛСО. Определения ЛСО пропозициональной альтернативы и спецификации приведены в табл. 1 (другие определения см. в [14, 18] и ниже).

Структурированные определения ЛСО пропозициональной альтернативы и спецификации показывают, что они имеют лишь один общий различительный признак: оба они установлены на пропозициональном уровне. Этого недостаточно для того, чтобы отнести их к единому иерархическому уровню, как предлагается в PDTB (см. обоснование в разд. 3). Напротив, у уступительных и условных ЛСО, оказывающихся в PDTB на разных уровнях, общих признаков больше (табл. 2). В частности, общим для них является помимо пропозиционального уровня базовая операция.

В этой связи следует заметить, что если некоторые различительные признаки, например «пропозициональный уровень» или «операция импликации», характеризуют несколько ЛСО, то другие выступают уникальными свойствами того или иного ЛСО, позволяющими его идентифицировать и отличать от других. К ним относится, например, признак « $Y$  содержит более частное понятие  $q$ , сужающее экстенционал  $p$ »<sup>2</sup>, характеризующий ЛСО спецификации. В табл. 3 приведены данные по использованию различительных признаков для ЛСО, получивших определения в НБД.

Таблица 1 Примеры структурированных определений ЛСО

ЛСО	Базовая семантическая операция	Уровень	Дополнительные характеристики
Пропозициональная альтернатива	<ul style="list-style-type: none"> <li>Операция сравнения, устанавливающая сходство <math>p</math> и <math>q</math></li> </ul>	<ul style="list-style-type: none"> <li>Пропозициональный</li> </ul>	<ul style="list-style-type: none"> <li><math>p</math> и <math>q</math> — положения вещей, имеющие статус гипотезы;</li> <li>говорящий предлагает сделать выбор между <math>p</math> и <math>q</math></li> </ul>
Спецификация	<ul style="list-style-type: none"> <li>Операция соотнесения общего и частного</li> </ul>	<ul style="list-style-type: none"> <li>Пропозициональный</li> </ul>	<ul style="list-style-type: none"> <li><math>X</math> содержит обобщенное понятие или положение вещей <math>p</math>;</li> <li><math>Y</math> содержит более частное понятие <math>q</math>, сужающее экстенционал <math>p</math></li> </ul>

<sup>1</sup>Подробнее о структуре НБД, ее возможностях и результатах, полученных с ее использованием, см., например, [15–17]. Представительный фрагмент НБД доступен по адресу: <http://a179.frccsc.ru/RFH41002/main.aspx>.

<sup>2</sup>Строчные буквы  $p$  и  $q$  обозначают положение дел, прописные  $P$  и  $Q$  — акты высказывания, прописные  $X$  и  $Y$  — фрагменты текста.

**Таблица 2** Структурированные определения уступительных и условных ЛСО

ЛСО	Базовая семантическая операция	Уровень	Дополнительные характеристики
Уступительные	• Операция импликации	• Пропозициональный	• $p$ и $q$ — положения вещей; • как правило, если имеет место $q$ , то не имеет места $p$
Условные	• Операция импликации	• Пропозициональный	• $p$ и $q$ — положения вещей, имеющие статус гипотезы или неосуществившегося положения вещей; • если имеет место $p$ , то имеет место $q$

**Таблица 3** Использование признаков в описаниях ЛСО

Признак ЛСО	Количество ЛСО
Пропозициональный уровень	14
Операция сравнения, устанавливающая несходство $p$ и $q$	12
Метаязыковой уровень	6
Уровень высказывания	6
$p$ отвергается, принимается $q$	5
Операция сравнения, устанавливающая сходство между $p$ и $q$	5
Говорящий предлагает сделать выбор между $p$ и $q$	5
...	...
Как правило, если имеет место $q$ , то не имеет места $p$	1
$p$ и $q$ — положения вещей, не связанные никаким другим ЛСО	1
$p$ верно, только если исключить осуществление $q$	1
Положение вещей $q$ служит аргументом в пользу ожидаемого вывода не- $r$	1
Положение вещей $p$ служит аргументом в пользу ожидаемого вывода $r$	1
$p$ и $q$ имеют одинаковый экстенционал	1
$q$ — возможное описание того же положения вещей $r$	1
$q$ — обобщенное («без частных») представление положения вещей, сделанное на основании свойств $p$	1
...	...

Из 52 использованных в НБДК различительных признаков 19 характеризуют более одного ЛСО, а 33 уникальны. Тем не менее эти уникальные признаки могут быть объединены в семейства, фиксирующие концептуальную близость признаков, несмотря на их формальные различия. Например, семейство признаков «используется отрицательный коррелят  $p/P$ » включает в себя следующий набор признаков:

- (1)  $P$  отвергается, принимается  $q$ ;
- (2)  $q$  имеет место, если не имеет места  $p$ ;
- (3)  $q$  имеет место, если не имеет места положение вещей, описанное в  $P$ ;
- (4)  $p$  отвергается, принимается  $q$ ;
- (5) как правило, если имеет место  $q$ , то не имеет места  $p$ .

### 3 Критерии определения степени семантической близости логико-семантических отношений

Структурированные определения ЛСО позволяют создать их непротиворечивую классификацию, обосновав число высших иерархических уровней и объединение ЛСО в семантические классы общностью лежащей в их основе семантической операции. Общность этого признака не всегда, однако, служит необходимым и достаточным условием для того, чтобы считать ЛСО семантически близкими. Было замечено, что, с одной стороны, показателю ЛСО в одном языке может соответствовать в переводе показатель ЛСО, относящийся к другому иерархическому уровню, причем как при наличии системного переводного эквивалента у данного показателя (а), так и при его отсутствии (б), и такие

Таблица 4 Структурированные определения ЛСО одновременности, сопутствования и сопоставления

ЛСО	Базовая семантическая операция	Уровень	Дополнительные характеристики
Одновременность	• Расположение на оси времени	• Пропозиционный	• $p$ и $q$ — положения вещей; • $Tr$ включает в себя $Tq$
Сопутствование	• Операция сравнения, устанавливающая сходство между $p$ и $q$	• Пропозиционный	• $q$ — положение вещей, зависимое от $p$ ; • $Tr$ включает в себя $Tq$
Сопоставление	• Операция сравнения, устанавливающая несходство $p$ и $q$	• Пропозиционный	• $p$ и $q$ актуальны для говорящего в момент речи $Td$ ; • сходство $p$ и $q$ относительно некоторого «общего знаменателя»

случаи не единичны. Например, в высказывании (а) итальянский коннектор *intanto* ‘между тем’ переведен русским сопоставительным союзом *а*:

(а) Si volse, e prese ad arrampicarsi di traverso lungo la proda assoluta dell'argine. Si aiutava con la mano destra, afferrandosi ai ciuffi dell'erba; *intanto*, la sinistra levata all'altezza del capo, veniva togliendosi e rimettendosi il cerchietto ferma-capelli. — Она повернулась и стала карабкаться по залитому солнцем спуску, хватаясь правой рукой за траву, а левой, поднятой над головой, поправляла обруч на волосах. [Giorgio Bassani. *Il giardino dei Finzi-Contini* (1962), перевод И. Соболева (2008)].

В (б) тот же итальянский коннектор переводит русский показатель ЛСО сопутствования:

(б) Он говорил громко *и при этом* делал такие удивленные глаза, что можно было подумать, что он лгал. — Egli parlava ad alta voce *e intanto* faceva degli occhi così meravigliati che si pensava ch'egli mentisse. [А. П. Чехов. Палата № 6 (1892), перевод Ф. Malcovati].

С другой стороны, известно, что показатели ЛСО многозначны: на это указывают, в частности, словари и грамматики. Так, для русского коннектора *между тем*, выступающего системным эквивалентом итальянского *intanto*, словарь [19] дает три значения: одновременность, сопоставление и противительность (ЛСО «вопреки ожидаемому»).

Если, однако, сравнить определения ЛСО сопутствования, сопоставления и одновременности, то увидим, что они, несмотря на то что в их основе лежат разные семантические операции, имеют общие различительные признаки (табл. 4).

Общим для трех отношений является признак «пропозиционный уровень», для ЛСО одновременности и сопутствования — признак « $Tr$  включает в себя  $Tq$ », который входит в то же семейство признаков («единство временного интервала»), что и признак « $p$  и  $q$  актуальны для говорящего в момент речи  $Td$ », характеризующий ЛСО сопоставления. Таким образом, с одной стороны, в качестве переводного эквивалента показателя некоторого ЛСО выбираются показатели, разделяющие с ним различительные признаки, а с другой — набор значений, выражаемых коннектором, также не является произвольным, а определяется семантической близостью ЛСО.

Однако при определении семантической близости ЛСО не все различительные признаки имеют одинаковый вес, или «коэффициент близости». Так, признак «пропозиционный уровень» не может иметь высокий коэффициент, поскольку не является дискриминирующим. Он входит в группу признаков («уровень, на котором установлено ЛСО»), включающую всего три элемента, а значит, он становится общим для большого числа ЛСО, которые в подавляющем большинстве устанавливаются именно между пропозициями. Кроме того, ЛСО, если оно может устанавливаться на всех трех уровнях, в большинстве случаев устанавливается именно на уровне пропозиций. В табл. 5 приводятся данные для ЛСО альтернативы и коррекции, для которых в НБДК сделана сплошная выборка.

Признаки из группы «Базовая семантическая операция» также, по имеющимся данным, не имеют решающего значения для определения семантической близости ЛСО, хотя их коэффициенты, по-видимому, должны быть выше, чем у признаков

Таблица 5 Распределение аннотаций с ЛСО альтернативы и коррекции по уровням

ЛСО	Уровень		
	Пропозиционный	Иллокутивный	Метаязыковой
Альтернатива	2138	15	144
Коррекция	231	36	36

группы «Уровень», поскольку они являются классифицирующими и служат для обоснования того, почему ЛСО объединяются в семантические классы.

Проанализированные данные позволяют предположить, что наивысший коэффициент должен быть присвоен тем случаям, когда различительные признаки ЛСО относятся к одному семейству признаков. Это хорошо видно на примере ЛСО одновременности, сопоставления и сопутствования (см. табл. 4), а также на примере других ЛСО, получивших определения в НБДК. Дальнейшее исследование позволит присвоить числовые параметры коэффициентам близости для различительных признаков.

## 4 Заключение

Разработанная классификация ЛСО и сформированные на ее основе в НБДК структурированные определения позволяют не только избежать противоречий или необоснованных решений в классификации ЛСО, но и определить степень близости ЛСО с учетом общности их различительных признаков. В дальнейшем при определении коэффициентов близости, присваиваемых признакам, можно учитывать данные по сочетаемости ЛСО [20], по соответствиям ЛСО в оригинальном и переводном тексте, а также те случаи, когда разные ЛСО выражаются одним и тем же показателем. Полученные результаты позволяют улучшить автоматическую обработку текста, а также качество машинного перевода.

## Литература

1. *Hobbs J. R.* A computational approach to discourse analysis. — New York, NY, USA: Department of Computer Science, City College, City University of New York, 1976. Research Report 76-2. P. 28–38.
2. *Hobbs J. R.* Why is discourse coherent? — Menlo Park, CA, USA: SRI International, 1978. SRI Technical Note 176. 44 p.
3. *Hobbs J. R.* Coherence and coreference // *Cognitive Sci.*, 1979. Vol. 3. No. 1. P. 67–90.
4. Русская грамматика. / Под ред. Н. Ю. Шведовой. — М.: Наука, 1980. Т. 2. 714 с.
5. Лексикографические портреты служебных слов / Под ред. Е. А. Стародумовой, Е. С. Шереметьевой, В. Н. Завьялова. — Владивосток: ДВФУ, 2022. 322 с.
6. *Mann W. C., Thompson S. A.* Rhetorical structure theory: Towards a functional theory of text organization // *Text*, 1988. Vol. 8. No. 3. P. 243–281. doi: 10.1515/text.1.1988.8.3.243.
7. *Knott A., Dale R.* Using linguistic phenomena to motivate a set of coherence relations // *Discourse Process.*, 1994. Vol. 18. No. 1. P. 35–62. doi: 10.1080/01638539409544883.
8. *Rudolph E.* Contrast: Adversative and concessive expressions on sentence and text level. — Berlin/Boston: Walter de Gruyter, 1996. 564 p.
9. *Fraser B.* An account of discourse markers // *International Review Pragmatics*, 2009. Vol. 1. No. 2. P. 293–320. doi: 10.1163/187730909X12538045489818.
10. PDTB Research Group. The Penn Discourse Treebank 2.0 annotation manual. — Philadelphia, PA, USA: Institute for Research in Cognitive Science, University of Pennsylvania, 2008. Technical Report IRCS-08-01. 99 p. <https://www.cis.upenn.edu/~elenimi/pdtb-manual.pdf>.
11. *Breindl E., Volodina A., Waßner U. H.* Handbuch der deutschen Konnektoren 2. Semantik der deutschen Satzverknüpfers. — Berlin: Walter de Gruyter, 2014. 1327 p.
12. *Инькова О. Ю.* Логико-семантические отношения: проблемы классификации // *Связность текста: ме-реологические логико-семантические отношения.* — М.: ЯСК, 2019. С. 11–98.
13. *Webber B., Prasad R., Lee A., Joshi A.* The Penn Discourse Treebank 3.0 annotation manual, 2019. 81 p. [https:// catalog ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf](https://catalog ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf).
14. *Инькова О. Ю., Кружков М. Г.* Структурированные определения дискурсивных отношений в Надкорпусной базе данных коннекторов // *Информатика и её применения*, 2021. Т. 15. Вып. 4. С. 27–32. doi: 10.14357/19922264210404.
15. Семантика коннекторов: контрастивное исследование / Под. ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2018. 368 с.
16. Структура коннекторов и методы ее описания / Под. ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2019. 316 с.
17. *Кружков М. Г.* Концепция построения надкорпусных баз данных // *Системы и средства информатики*, 2021. Т. 31. № 3. С. 101–112. doi: 10.14357/08696527210309.
18. *Инькова О. Ю.* Определения дискурсивных отношений: опыт Надкорпусной базы данных коннекторов // *Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной Междунар. конф. «Диалог».* — М.: РГГУ, 2021. Вып. 20(27). С. 328–338.
19. Словарь структурных слов русского языка / Под ред. В. В. Морковкина. — М.: Лазурь, 1997. 422 с.
20. *Инькова О. Ю., Кружков М. Г.* Сочетаемость логико-семантических отношений: количественные методы анализа // *Информатика и её применения*, 2019. Т. 13. Вып. 2. С. 83–91. doi: 10.14357/19922264190212.

Поступила в редакцию 10.07.23



# EVALUATION CRITERIA FOR DISCOURSE RELATIONS SEMANTIC AFFINITY

O. Yu. Inkova<sup>1,2</sup> and M. G. Kruzhkov<sup>1</sup>

<sup>1</sup>Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

<sup>2</sup>University of Geneva, 22 Bd des Philosophes, CH-1205 Geneva 4, Switzerland

**Abstract:** The paper presents an overview of structured definitions of discourse relations created based on classification principles and criteria for evaluating their semantic affinity. The authors point out the shortcomings of existing classification approaches that are sometimes inconsistent or contradictory and outline the benefits of an alternative approach to classification of discourse relations which is based on their structured definition. The paper provides examples of such definitions created within the Supracorpora Database of Connectives and discusses the criteria for evaluating their semantic closeness. As the structured definitions are represented by sets of distinguishing features, the authors discuss the problem of identifying proximity factors for each of these features. The gathered data suggest a hypothesis that among the three groups of features: “Level,” “Basic operation,” and “Feature family”, it is the last one that should have the most impact. Finally, directions for further research of this problem are considered, namely, the option of taking into account such factors as compatibility of discourse relations, correspondence of relations between the source text and its translation, and such cases where certain relation markers may express different discourse relations in various contexts.

**Keywords:** supracorpora database; logical-semantic relations; connectives; annotation; faceted classification

**DOI:** 10.14357/19922264230314

**EDN:** UJZJZI

## Acknowledgments

The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (CKP “Informatics”) of FRC CSC RAS (Moscow).

## References

- Hobbs, J. R. 1976. A computational approach to discourse analyses. New York, NY: Department of Computer Science, City College, City University of New York. Research Report 76-2. 28–38.
- Hobbs, J. R. 1978. Why is discourse coherent? Menlo Park, CA: SRI International. SRI Technical Note 176. 44 p.
- Hobbs, J. R. 1979. Coherence and coreference. *Cognitive Sci.* 3(1):67–90.
- Shvedova, N. Yu., ed. 1980. *Russkaya grammatika*. [Russian grammar]. Moscow: Nauka. Vol. 2. 714 p.
- Starodumova, E. A., E. S. Sheremet’eva, and V. N. Zav’yalov, eds. 2022. *Leksikograficheskie portrety sluzhebnykh slov* [Lexicographic portraits of auxiliary words]. Vladivostok: FEFU. 322 p.
- Mann, W. C., and S. A. Thompson. 1988. Rhetorical structure theory: Towards a functional theory of text organization. *Text* 8(3):243–281. doi: 10.1515/text.1.1988.8.3.243.
- Knott, A., and R. Dale. 1994. Using linguistic phenomena to motivate a set of coherence relations. *Discourse Process.* 18(1):35–62. doi: 10.1080/01638539409544883.
- Rudolph, E. 1996. *Contrast: Adversative and concessive expressions on sentence and text level*. Berlin/Boston: Walter de Gruyter. 564 p.
- Fraser, B. 2009. An account of discourse markers. *International Review Pragmatics* 1(2):293–320. doi: 10.1163/187730909X12538045489818.
- PDTB Research Group. 2008. The Penn Discourse Treebank 2.0 annotation manual. Philadelphia, PA: Institute for Research in Cognitive Science, University of Pennsylvania. Technical Report IRCS-08-01. 99 p. Available at: <https://www.cis.upenn.edu/~elenimi/pdtb-manual.pdf> (accessed August 1, 2023).
- Breindl, E., A. Volodina, and U. H. Waßner. 2014. *Handbuch der deutschen Konnektoren 2. Semantik der deutschen Satzverknüpfers*. Berlin: Walter de Gruyter. 1327 p.
- Inkova, O. Yu. 2019. Logiko-semanticheskie otnosheniya: problemy klassifikatsii [Logical-semantic relations: Classification problems]. *Svyaznost’ teksta: mereologicheskie logiko-semanticheskie otnosheniya* [Text Coherence: Mereological Logical Semantic Relations]. Moscow: LRC Publ. 11–98.
- Webber, B., R. Prasad, A. Lee, and A. Joshi. 2019. The Penn Discourse Treebank 3.0 annotation manual. 81 p. Available at: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf> (accessed August 1, 2023).
- Inkova, O. Yu., and M. G. Kruzhkov. 2021. Strukturirovannye opredeleniya diskursivnykh otnosheniy v Nadkor-

- pusnoy baze dannykh konnektorov [Structured definitions of discourse relations in the Supracorpora Database of Connectives]. *Informatika i ee Primeneniya — Inform. Appl.* 15(4):27–32. doi: 10.14357/19922264210404.
15. Inkova, O. Yu., ed. 2018. *Semantika konnektorov: Kontrastivnoe issledovanie* [Semantics of connectives: Contrastive study]. Moscow: TORUS PRESS. 368 p.
  16. Inkova, O. Yu., ed. 2019. *Struktura konnektorov i metody ee opisaniya* [Structure of connectors and methods of its description]. Moscow: TORUS PRESS. 316 p.
  17. Kruzhkov, M. G. 2021. Kontsepsiya postroeniya nadkorporusnykh baz dannykh [Conceptual framework for supracorpora databases]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(3):101–112. doi: 10.14357/08696527210309.
  18. Inkova, O. Yu. 2021. Opredeleniya diskursivnykh otnosheniy: opyt Nadkorporusnoy bazy dannykh konnektorov [Definition of discursive relations: The experience of the supracorpora database of connectors]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Po mat-m ezhegodnoy Mezhdunar. konf. "Dialog"* [Computational Linguistics and Intellectual Technologies: Papers from the Annual Conference (International) "Dialogue"]. Moscow: RGGU. 20(27):328–338.
  19. Morkovkin, V. V., ed. 1997. *Slovar' strukturnykh slov russkogo yazyka* [Dictionary of structural words of the Russian language]. Moscow: Lazur'. 422 p.
  20. Inkova, O. Yu., and M. G. Kruzhkov. 2019. Sochetaemost' logiko-semanticheskikh otnosheniy: kolichestvennyye metody analiza [Compatibility of logical semantic relations: Methods of quantitative analysis]. *Informatika i ee Primeneniya — Inform. Appl.* 13(2):83–91. doi: 10.14357/19922264190212.

Received July 10, 2023

## Contributors

**Inkova Olga Yu.** (b. 1965) — Doctor of Science in philology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; faculty member, University of Geneva, 22 Bd des Philosophes, CH-1205 Geneva 4, Switzerland; olyainkova@yandex.ru

**Kruzhkov Mikhail G.** (b. 1975) — senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; magnit75@yandex.ru

# ТРАНСФОРМАЦИЯ ИЕРАРХИИ АКОФФА В НАУЧНОЙ ПАРАДИГМЕ ИНФОРМАТИКИ\*

И. М. Зацман<sup>1</sup>

**Аннотация:** Рассматривается иерархия DIKW (data, information, knowledge, wisdom — данные, информация, знания, мудрость), которая была опубликована в 1989 г. Расселом Акоффом. В ней мудрость находится на вершине иерархии, затем следуют знание, информация и, в самом низу, данные. Первоначально предполагалось, что иерархию DIKW можно будет использовать для описания отношений между ее четырьмя компонентами. Однако проблему описания взаимных преобразований двух соседних компонентов, особенно для знания и информации, оказалось весьма сложно решить в рамках иерархии DIKW. Сложность ее решения заключается в том, что иерархия DIKW подразумевает генерацию знания в результате процесса фильтрации соседней с ней информации, но средства реализации этого процесса не были определены Акоффом. Не удастся также описать смысловую интерпретацию данных, так как они непосредственно не примыкают к знанию в иерархии DIKW, которая подразумевает наличие отношений только между соседними компонентами. Цель статьи состоит в трансформации иерархии DIKW в рамках научной парадигмы информатики, основанной на средовом делении ее предметной области на ментальную, информационную, цифровую и ряд других сред. В то время как Акофф использовал принцип вертикального размещения компонентов иерархии, в данной статье вместо этого предлагается соотнести используемые в информатике интерфейсы и знаковые системы с отношениями между тремя компонентами иерархии: данными, информацией и знанием. Если использовать принцип вертикального размещения не компонентов, а сред предметной области информатики, то тогда можно предложить подход к решению проблемы описания взаимных преобразований трех компонентов иерархии, сопоставив их с интерфейсами информатики и знаковыми системами. Такое сопоставление даст возможность увидеть те пары компонентов, для которых интерфейсы в настоящее время не формализованы, не имеют компьютерной реализации и выполняются экспертами. В статье приводится пример технологии извлечения знания, сочетающей автоматические и экспертные (неформализованные) технологические этапы.

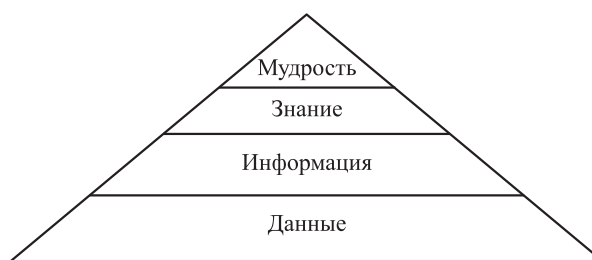
**Ключевые слова:** научная парадигма информатики; данные; информация; знание; иерархия DIKW; технологии извлечения знания

**DOI:** 10.14357/19922264230315

**EDN:** UMVRRV

## 1 Введение

Графически иерархию DIKW, которая состоит из четырех компонентов (данные, информация, знания, мудрость), часто представляют в форме пирамиды (рис. 1 [1]). Такое визуальное представление вполне соответствует первоначальному ее описанию: «Мудрость находится на вершине иерархии. . . и, в самом низу, данные. Каждое из перечисленных понятий [кроме данных] содержит в себе нижестоящие. . .» [2]. Утверждение о том, что *перечисленные понятия содержат в себе нижестоящие*, Акофф комментирует следующим образом. Данные он трактует как наборы символов, которые *характеризуют свойства объектов и событий, а также их окружение*. Эти наборы формируются в процессе наблюдения или зондирования. Информация определяется как *результат анализа данных*. Знание рассматривается им в контексте управления организационными системами в сфере экономики:



**Рис. 1** Иерархия DIKW (данные, информация, знание, мудрость)

это то, что делает возможным *преобразование информации в инструкции*, и это делает возможным управление такими системами [2].

Процесс анализа данных и получение информации на их основе, как правило, предполагают их ментальное понимание человеком, формирование концептов (составляющих знания) и их представление в виде текста, диаграмм, графиков и т. д. Таким

\* Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

<sup>1</sup> Федеральний исследовательский центр «Информатика и управление» Российской академии наук; izatsman@yandex.ru

образом, до получения информации в результате анализа данных сначала формируется знание, но оно в иерархии не имеет общей границы с данными.

Согласно Дэвиду Вайнбергеру, переход от информации к знаниям в иерархии DIKW оказался еще более проблематичен, чем переход от данных к информации. Ее недостатки он описал так: «знание — это не просто результат фильтрации [компонентов]... Оно — результат гораздо более сложного процесса, который является *социальным, целенаправленным, контекстно и культурно обусловленным*... Наиболее важным в этом отношении является то, что там, где решения сложные и знание [необходимое для решения] получить трудно, оно не определяется информацией, поскольку именно процесс познания в первую очередь определяет, какая информация необходима и как ее следует использовать... Представление о том, что знание (а тем более мудрость) является результатом применения фильтров на каждом уровне, рисует неверную картину (курсив мой. — И. З.)» [3].

В данной статье предлагается изменить принцип построения иерархии и соотнести используемые в информатике интерфейсы и знаковые системы с отношениями между тремя компонентами иерархии: данными, информацией и знанием. Акофф использовал принцип вертикального размещения компонентов иерархии, которому свойственны недостатки, проанализированные, в частности, в работах [1, 3].

Цель статьи состоит в трансформации иерархии DIKW в рамках научной парадигмы информатики, основанной на средовом делении ее предметной области на ментальную, информационную, цифровую и ряд других сред. Если использовать принцип вертикального размещения не компонентов, а сред предметной области информатики, то тогда можно предложить подход к решению проблемы описания взаимных преобразований трех компонентов иерархии (кроме мудрости), сопоставив их с интерфейсами информатики и знаковыми системами. Такое сопоставление даст возможность увидеть те пары компонентов, для которых интерфейсы в настоящее время не формализованы и не имеют компьютерной реализации (т. е. преобразования таких компонентов выполняются экспертами).

## 2 Среды и объекты предметной области информатики

В работе [4] было начато описание научной парадигмы информатики, основанной на средовом делении ее предметной области. Было описано

основание для построения верхнего уровня классификации сущностей ее предметной области (позиционируемых как феномены разной природы) и даны определения следующих пяти ее сред, каждая из которых включает сущности одной и той же природы:

- (1) *ментальная среда* — это совокупность когнитивных феноменов, формируемых в процессах познания, происходящих в сознании людей (далее — концепты как составляющие знания человека);
- (2) *информационная среда* — это совокупность сенсорно воспринимаемых феноменов, находящихся вне сознания;
- (3) *цифровая среда* — это совокупность компьютерных кодов;
- (4) *нейросреда* — это электрические потенциалы и магнитные поля, генерируемые мозгом, которые используются в информационных технологиях (ИТ) управления роботизированной рукой [5] и в других ИТ, применяющих интерфейсы «мозг–компьютер»;
- (5) *ДНК-среда* — это совокупность цепочек РНК и ДНК.

В соответствии с перечисленными средами верхний уровень классификации сущностей предметной области информатики включает как минимум пять классов, каждый из которых содержит объекты одной среды: ментальной, информационной, цифровой, нейросреды или ДНК-среды. При этом с ростом разнообразия природы сущностей верхний уровень классификации может пополняться новыми классами, если при проектировании ИТ обнаружатся сущности, которые по своей природе не относятся ни к одной из ранее уже выделенных сред [6, 7].

## 3 Трансформация иерархии DIKW

Согласно используемой парадигме информатики и определению верхнего уровня классификации сущностей ее предметной области, *ментальная среда* охватывает концепты как составляющие знания. *Информационная среда* содержит как минимум сенсорно воспринимаемые данные и знаковую информацию. При их компьютерном кодировании получаем соответственно цифровые данные и цифровую информацию, которые в парадигме информатики позиционируются как две принципиально разные сущности *цифровой среды*. Нейросреда и ДНК-среда в трансформации иерархии DIKW в этой статье не рассматриваются.



Рис. 2 Средовая версия иерархия DIKW как результат трансформации

Если использовать принцип вертикального размещения не компонентов, а сред предметной области информатики, то получим новый вариант иерархии DIKW (рис. 2), который назовем *средовой версией* иерархии DIKW. Она включает три среды и как минимум восемь сущностей:

- (1) мудрость, знание, ментальные образы данных и концепты ментальной среды;
- (2) знаковую информацию и сенсорно воспринимаемые данные информационной среды;

- (3) цифровые данные и цифровую информацию цифровой среды.

Сопоставим известные интерфейсы информатики, реализуемые с помощью кодовых таблиц, и знаковые системы с двумя следующими границами средовой версии иерархии DIKW:

- (1) между информационной и цифровой средами;
- (2) между информационной и ментальной средами.

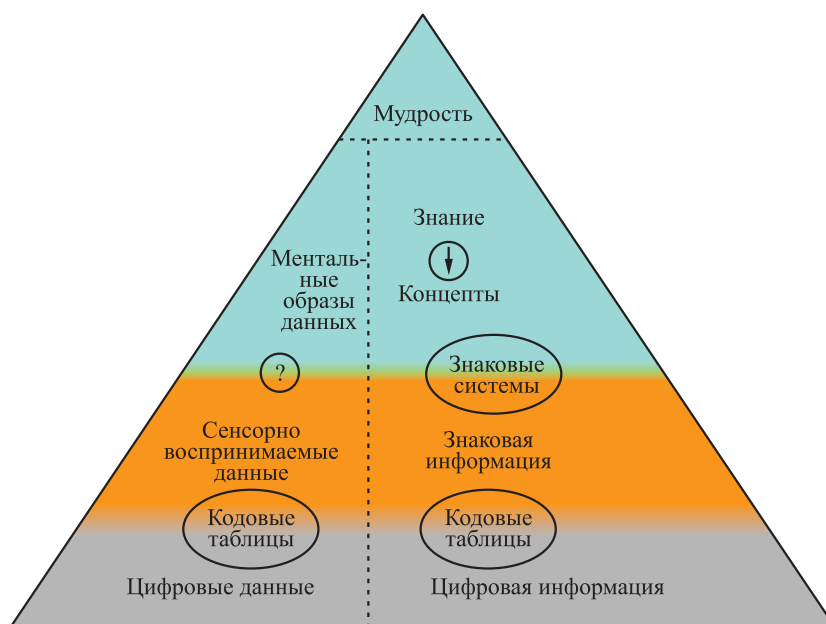


Рис. 3 Средовая версия иерархия DIKW: знаковые системы и кодовые таблицы

В результате сопоставления можно увидеть (рис. 3), что первой границе в информатике соответствуют кодовые таблицы, если данные могут быть представлены в символической форме. Что касается второй границы, то для представления концептов знания широко используются знаковые системы.

Наименее исследованными остаются отношения между сенсорно воспринимаемыми данными и их ментальными образами. Однако для проектирования в информатике одного из ключевых этапов ИТ извлечения нового знания из данных необходимо тем или иным способом реализовать эти отношения. Отметим, что их исследование и описание относятся к когнитивным наукам, а не к предметной области информатики.

В современных ИТ визуальной аналитики [8, 9] и извлечения нового знания из данных [10, 11] для их смысловой *интерпретации* привлекаются, как правило, эксперты, обладающие опытом в соответствующей предметной области.

#### 4 Модель технологии извлечения знания

Для применения средовой версии иерархии DIKW в процессе проектирования ИТ извлечения нового знания из данных и/или текстовой информации необходимо разделить индивидуальные знания экспертов и их коллективные знания [12–14]. Вариант средовой версии (см. рис. 3), дополненный разделением на индивидуальное и коллективное знание экспертов, подробно описан в работе [15]. В этой же работе определены основные процессы и стадии извлечения нового знания, а также предложена модель технологии извлечения знания на основе спиральной модели генерации знания, которую создал Икуджиро Нонака [12].

Основные отличия предлагаемой модели от ранее созданной спиральной модели состоят в следующем. Во-первых, предлагаемая модель включает в себя потенциальные источники нового знания, которые должны соответствовать цели его генера-

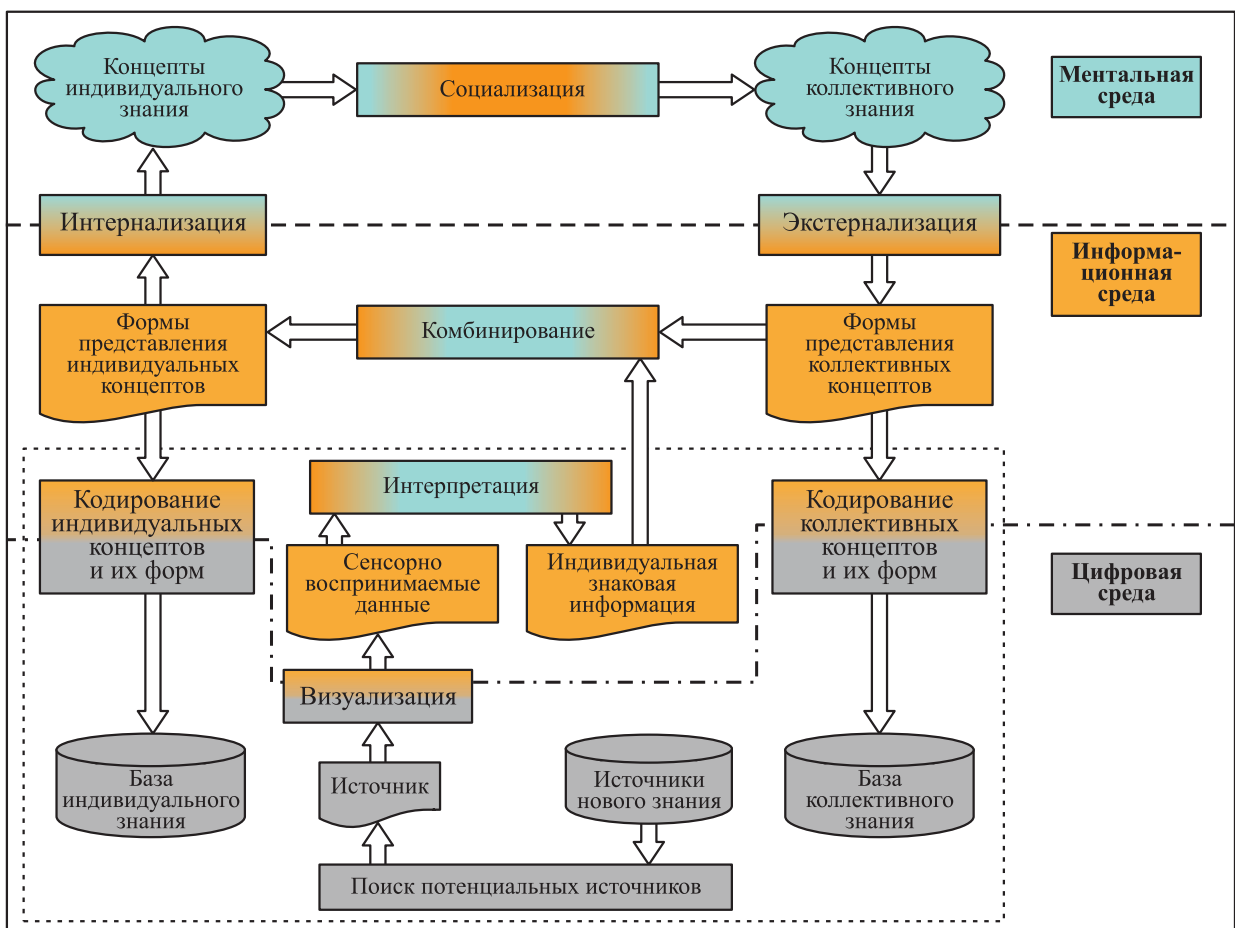


Рис. 4 Модель ИТ извлечения знания

ции. Во-вторых, она охватывает сущности трех сред разной природы: ментальной, информационной и цифровой, а спиральная модель подразумевает использование только ментальной и информационной сред. В-третьих, она включает восемь процессов извлечения знания (рис. 4), а спиральная модель — четыре процесса (интернализация, социализация, экстернализация и комбинирование, которые подробно описаны в работах [12–14]). Предлагаемая модель кроме этих четырех процессов спиральной модели включает еще четыре: *поиск* потенциальных источников нового знания, их *визуализация*, *интерпретация* экспертами и *кодирование* индивидуальных и коллективных концентров знания экспертов.

В предлагаемой модели процессы могут состоять из одной, четырех, пяти или шести стадий, которые описаны в работе [15]. Верхняя часть этой модели на рис. 4 практически полностью совпадает со спиральной моделью. Нижняя ее часть начинается с поиска потенциальных источников нового знания в базе данных, соответствующих цели его извлечения. Например, в задачах медицинской информатики это могут собранные клинические данные о течении некоторого исследуемого заболевания.

Найденный источник визуализируется для его последующей интерпретации экспертом. На рис. 4 изображен случай, когда в результате интерпретации сенсорно воспринимаемых данных (информационная среда) сформированы концепты нового знания (ментальная среда) и они представлены как индивидуальная информация (информационная среда) этого эксперта. Последующие процессы (комбинирование, интернализация, социализация и экстернализация) соответствуют спиральной модели.

Если вход и выход процесса принадлежат одной среде, то на рис. 4 она размещается именно в ней (см. процессы социализации, комбинирования и интерпретации). Если вход и выход процесса принадлежат двум разным средам, то она размещается на границе этих сред (см. процессы интернализации, экстернализации, визуализации и кодирования).

Принципиальное отличие предлагаемой модели от спиральной состоит в итерационном пополнении баз индивидуальных и коллективных знаний. Представление ее процессов и их стадий в средовой версии иерархии DIKW детально описано в работе [15].

## 5 Заключение

Рассмотренный вариант трансформации иерархии DIKW выполнен в рамках научной парадиг-

мы информатики, в которой ее предметная область делится на ментальную, информационную, цифровую, нейро- и ДНК-среду. В статье рассматривались только первые три среды, поэтому интерфейс «мозг—компьютер» отсутствует в рассмотренном варианте средовой версии иерархии DIKW. Описание ее варианта с позиционированием интерфейса «мозг—компьютер» заслуживает отдельной статьи.

Переход к принципу вертикального размещения сред позволил сопоставить отдельные известные интерфейсы и перечень отношений между компонентами иерархии DIKW. Такое сопоставление дало возможность увидеть местоположение отношений между сенсорно воспринимаемыми данными (информационная среда) и их ментальными образами (ментальная среда), которые не формализованы и не имеют компьютерной реализации (они обозначены знаком «?» на рис. 3).

Согласно Харелу, реализация таких отношений относится к задачам когнитивной сложности. Основной вопрос при решении подобных задач состоит в том, чтобы описать отношения между ментальными образами и знанием человека так, чтобы его представление подавалось алгоритмической обработке, применению систем и средств информатики [16, с. 402].

## Литература

1. Rowley J. The wisdom hierarchy: Representations of the DIKW hierarchy // J. Inf. Sci., 2007. Vol. 33. No. 2. P. 163–180. doi: 10.1177/0165551506070706.
2. Ackoff R. From data to wisdom // J. Appl. Systems Analysis, 1989. Vol. 16. P. 3–9.
3. Weinberger D. The problem with the data–information–knowledge–wisdom hierarchy // Harvard Bus. Rev., 2010. <https://hbr.org/2010/02/data-is-to-info-as-info-is-not>.
4. Зацман И. М. Теоретические основания компьютерного образования: среды предметной области информатики как основание классификации ее объектов // Системы и средства информатики, 2022. Т. 32. № 4. С. 77–89. doi: 10.14357/08696527220408.
5. Зацман И. М. Интерфейсы третьего порядка в информатике // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 82–89. doi: 10.14357/19922264190312.
6. Зацман И. М. Таблица интерфейсов информатики как информационно-компьютерной науки // Научно-техническая информация. Сер. 1: Организация и методика информационной работы, 2014. № 11. С. 1–15.
7. Зацман И. М. О научной парадигме информатики: верхний уровень классификации объектов ее предметной области // Информатика и её применения, 2022. Т. 16. Вып. 4. С. 108–114. doi: 10.14357/19922264220411.

8. Chen M., Ebert D., Hagen H., Laramée R., Van Liere R., Ma K.-L., Ribarsky W., Scheuermann G., Silver D. Data, information, and knowledge in visualization // IEEE Comput. Graph., 2009. Vol. 29. No. 1. P. 12–19. doi: 10.1109/MCG.2009.6.
9. Federico P., Wagner M., Rind A., Amor-Amoros A., Miksch S., Aigner W. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics // Conference on Visual Analytics Science and Technology Proceedings. — New York, NY, USA: IEEE, 2017. P. 92–103. doi: 10.1109/VAST.2017.8585498.
10. Zatsman I. A model of goal-oriented knowledge discovery based on human–computer symbiosis // 16th Forum (International) on Knowledge Asset Dynamics Proceedings. — Matera, Italy: Arts for Business Institute, 2021. P. 297–312.
11. Zatsman I., Khakimova A. New knowledge discovery for creating terminological profiles of diseases // 22nd European Conference on Knowledge Management Proceedings. — Reading, U.K.: Academic Publishing International Ltd., 2021. P. 837–846. doi: 10.34190/EKM.21.041.
12. Nonaka I. The knowledge-creating company // Harvard Bus. Rev., 1991. Vol. 69. No. 6. P. 96–104.
13. Nonaka I. A dynamic theory of organizational knowledge creation // Organ. Sci., 1994. Vol. 5. No. 1. P. 14–37. doi: 10.1287/orsc.5.1.14.
14. Нонака И., Такеучи Х. Компания — создатель знания / Пер. с англ. — М.: Олимп-бизнес, 2003. 384 с. (Nonaka I., Takeuchi H. The knowledge-creating company. — Oxford, NY, USA: Oxford University Press, 1995. 284 p.)
15. Zatsman I. Digital spiral model of knowledge creation and encoding its dynamics // 18th Forum (International) on Knowledge Asset Dynamics Proceedings. — Matera, Italy: Arts for Business Institute, 2023. P. 581–596.
16. Harel D. Algorithmics — the spirit of computing. — Reading, MA, USA: Addison-Wesley, 1987. 514 p.

Поступила в редакцию 21.06.23

## TRANSFORMATION OF THE ACKOFF'S HIERARCHY IN THE SCIENTIFIC PARADIGM OF INFORMATICS

I. M. Zatsman

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The DIKW (data, information, knowledge, and wisdom) hierarchy, which was published in 1989 by Russell Ackoff, is considered. In it, wisdom is at the top of the hierarchy followed by knowledge, information, and, at the very bottom, data. It was originally intended that the DIKW hierarchy could be used to describe the relationships between its four components. However, the problem of describing the mutual transformations of two neighboring components, especially for knowledge and information, turned out to be very difficult to solve within the DIKW hierarchy. The complexity of its solution lies in the fact that the DIKW hierarchy implies the generation of knowledge as an outcome of the process of filtering neighboring information but the means of implementing this process were not defined by Ackoff. It is also impossible to describe the semantic interpretation of data, since they are not directly adjacent to knowledge in the DIKW hierarchy which implies the existence of relations between neighboring components only. The aim of the paper is to transform the DIKW hierarchy within the framework of the scientific paradigm of informatics which is based on the medium division of its subject domain into mental, informational, digital, and a number of other media. While Ackoff used the principle of vertical placement of the components, this paper instead proposes to correlate the interfaces and sign systems used in informatics with the relationships between the three components of the hierarchy: data, information, and knowledge. If one uses the principle of vertical placement not of the components but of the media of the subject domain of informatics, then it can be proposed an approach to solving the problem of describing the mutual transformations of the three components of the hierarchy comparing them with informatics interfaces and sign systems. Such a comparison will make it possible to see those pairs of components for which the interfaces are not currently formalized, do not have a computer implementation, and are performed by experts. The paper provides the example of a knowledge discovery technology that combines automatic and expert (nonformalized) technological stages.

**Keywords:** scientific paradigm of informatics; data; information; knowledge; DIKW hierarchy; knowledge discovery technology

**DOI:** 10.14357/19922264230315

**EDN:** UMVRRV

### Acknowledgments

The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (СКР “Informatics”) of FRC CSC RAS (Moscow).



## References

1. Rowley, J. 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. *J. Inf. Sci.* 33(2):163–180. doi: 10.1177/0165551506070706.
2. Ackoff, R. 1989. From data to wisdom. *J. Appl. Systems Analysis* 16(1):3–9.
3. Weinberger, D. 2010. The problem with the data-information-knowledge-wisdom hierarchy. *Harvard Bus. Rev.* Available at: <https://hbr.org/2010/02/data-is-to-info-as-info-is-not> (accessed July 11, 2023).
4. Zatsman, I. M. 2022. Teoreticheskie osnovaniya komp'yuternogo obrazovaniya: sredy predmetnoy oblasti informatiki kak osnovanie klassifikatsii ee ob"ektov [Theoretical foundations of digital education: Subject domain media of informatics as the base of its objects' classification]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(4):77–89. doi: 10.14357/08696527220408.
5. Zatsman, I. M. 2019. Interfeysy tret'ego poryadka v informatike [Third-order interfaces in informatics]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):82–89. doi: 10.14357/19922264190312.
6. Zatsman, I. M. 2014. A table of interfaces of informatics as computer and information science. *Scientific Technical Information Processing* 41(4):230–243. doi: 10.3103/S014768821404008X.
7. Zatsman, I. 2022. O nauchnoy paradigme informatiki: verkhniy uroven' klassifikatsii ob"ektov ee predmetnoy oblasti [On the scientific paradigm of informatics: The classification high level of its objects]. *Informatika i ee Primeneniya — Inform. Appl.* 16(4):73–79. doi: 10.14357/19922264220411.
8. Chen, M., D. Ebert, H. Hagen, R. Laramée, R. van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver. 2009. Data, information, and knowledge in visualization. *IEEE Comput. Graph.* 29(1):12–19. doi: 10.1109/MCG.2009.6.
9. Federico, P., M. Wagner, A. Rind, A. Amor-Amoros, S. Miksch, and W. Aigner. 2017. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. *Conference on Visual Analytics Science and Technology Proceedings*. New York, NY: IEEE. 92–103. doi: 10.1109/VAST.2017.8585498.
10. Zatsman, I. 2021. A model of goal-oriented knowledge discovery based on human–computer symbiosis. *16th Forum (International) on Knowledge Asset Dynamics Proceedings*. Matera, Italy: Arts for Business Institute. 297–312.
11. Zatsman, I., and A. Khakimova. 2021. New knowledge discovery for creating terminological profiles of diseases. *22nd European Conference on Knowledge Management Proceedings*. Reading, U.K.: Academic Publishing International Ltd. 837–846. doi: 10.34190/EKM.21.041.
12. Nonaka, I. 1991. The knowledge-creating company. *Harvard Bus. Rev.* 69(6):96–104.
13. Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. *Organ. Sci.* 5(1):14–37. doi: 10.1287/orsc.5.1.14.
14. Nonaka, I., and H. Takeuchi. 1995. *The knowledge-creating company*. Oxford, NY: Oxford University Press. 284 p.
15. Zatsman, I. 2023. Digital spiral model of knowledge creation and encoding its dynamics. *18th Forum (International) on Knowledge Asset Dynamics Proceedings*. Matera, Italy: Arts for Business Institute. 581–596.
16. Harel, D. 1987. *Algorithmics — the spirit of computing*. Reading, MA: Addison-Wesley. 514 p.

Received June 21, 2023

## Contributor

**Zatsman Igor M.** (b. 1952) — Doctor of Science in technology, head of department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; izatsman@yandex.ru

**Агаларов Явер Мирзабекович** (р. 1952) — кандидат технических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Бегишев Вячеслав Олегович** (р. 1988) — кандидат физико-математических наук, доцент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

**Борисов Андрей Владимирович** (р. 1965) — доктор физико-математических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Босов Алексей Вячеславович** (р. 1969) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Вакуленко Василий Васильевич** (р. 1995) — инженер-исследователь Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Волканов Дмитрий Юрьевич** (р. 1979) — кандидат физико-математических наук, доцент кафедры автоматизации систем вычислительных комплексов факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

**Горбунов Сергей Юрьевич** (р. 2000) — магистрант факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; математик Московского центра фундаментальной и прикладной математики

**Горшенин Андрей Константинович** (р. 1986) — доктор физико-математических наук, доцент, главный научный сотрудник, руководитель отдела Федерального исследовательского центра «Информатика и управление» Российской академии наук; доцент кафедры математической статистики факуль-

тета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

**Грушо Александр Александрович** (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Грушо Николай Александрович** (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Дюкова Анастасия Петровна** (р. 1995) — аспирант Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Дюкова Елена Всеволодовна** (р. 1945) — доктор физико-математических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Забжайло Михаил Иванович** (р. 1956) — доктор физико-математических наук, профессор, главный научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Зацман Игорь Моисеевич** (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Зейфман Александр Израилевич** (р. 1954) — доктор физико-математических наук, заведующий кафедрой Вологодского государственного университета; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; главный научный сотрудник Вологодского научного центра Российской академии наук, старший научный сотрудник Мос-

ковского центра фундаментальной и прикладной математики Московского государственного университета имени М. В. Ломоносова

**Инькова Ольга Юрьевна** (р. 1965) — доктор филологических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; сотрудник Университета Женева

**Королев Денис Олегович** (р. 1999) — студент Института компьютерных наук и технологий Санкт-Петербургского политехнического университета Петра Великого

**Кружков Михаил Григорьевич** (р. 1975) — старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Кудрявцев Алексей Андреевич** (р. 1978) — кандидат физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; старший научный сотрудник Московского центра фундаментальной и прикладной математики Московского государственного университета имени М. В. Ломоносова

**Малашенко Юрий Евгеньевич** (р. 1946) — доктор физико-математических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Малеев Олег Геннадьевич** (р. 1971) — кандидат технических наук, доцент Института компьютерных наук и технологий Санкт-Петербургского политехнического университета Петра Великого

**Маслов Александр Русланович** (р. 1998) — аспирант кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

**Масляков Глеб Олегович** (р. 1996) — аспирант Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Назарова Ирина Александровна** (р. 1966) — кандидат физико-математических наук, научный сотрудник

Федерального исследовательского центра «Информатика и управление» Российской академии наук  
**Сатин Яков Александрович** (р. 1978) — кандидат физико-математических наук, доцент Вологодского государственного университета

**Смирнов Дмитрий Владимирович** (р. 1984) — бизнес-партнер по ИТ департамента безопасности ПАО «Сбербанк России»

**Сопин Эдуард Сергеевич** (р. 1987) — кандидат физико-математических наук, доцент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Тимонина Елена Евгеньевна** (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Торшин Иван Юрьевич** (р. 1972) — кандидат физико-математических наук, кандидат химических наук, старший научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Усов Илья Андреевич** (р. 1997) — аспирант Вологодского государственного университета

**Шестаков Олег Владимирович** (р. 1976) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; ведущий научный сотрудник Московского центра фундаментальной и прикладной математики

**Шоргин Всеволод Сергеевич** (р. 1978) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

---

## Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

---

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.
5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или ЛАТЭХ (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 10 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на *русском и английском языках*:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах: [http://www.ipiran.ru/journal/issues/2013\\_07\\_01/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp) и [http://www.ipiran.ru/journal/issues/2013\\_07\\_01\\_eng/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp);
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний. Если источник имеет DOI и/или EDN, то их необходимо указывать.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/bgn/> бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

**Описание статьи из журнала:**

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaistvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

**Описание статьи из электронного журнала:**

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

**Описание статьи из продолжающегося издания (сборника трудов):**

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal'–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

**Описание материалов конференций:**

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

**Описание книги (монографии, сборники):**

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

**Описание переводной книги** (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — New York, NY, USA: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. New York: Wiley. 521 p.

**Описание неопубликованного документа:**

Laturov, A. R., M. M. Khasanov, and V. A. Baikov. 2004 (unpubl.). *Geologiya i dobycha (NGT GiD)* [Geology and production (NGT GiD)]. Certificate on official registration of the computer program No. 2004611198.

**Описание интернет-ресурса:**

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

**Описание диссертации или автореферата диссертации:**

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. Moscow. D.Sc. Diss. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 23 p.

**Описание ГОСТа:**

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

**Описание патента:**

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.

11. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- использовать attach (присоединение);
- в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.

12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

**Адрес редакции журнала «Информатика и её применения»:**

Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05

e-mail: [ieep@frccsc.ru](mailto:ieep@frccsc.ru) (Стригина Светлана Николаевна)

<http://www.ipiran.ru/journal/issues/>

---

## Requirements for manuscripts submitted to Journal “Informatics and Applications”

---

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
- mathematical methods for studying complex systems and processes;
- information systems and networks;
- information technologies; and
- architecture and software of computational complexes and networks.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

***Agreement on the transfer of rights to publish:***

*“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.*

*Author(s) signature(s): (name(s), address(es), date).*

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or L<sup>A</sup>T<sub>E</sub>X (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 10 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site: <http://www.ipiran.ru/journal/template.doc>.

8. The articles should enclose data both in **Russian and English**:

- title;
- author’s name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format (see site):

[http://www.ipiran.ru/journal/issues/2013\\_07\\_01/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp) and

[http://www.ipiran.ru/journal/issues/2013\\_07\\_01\\_eng/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp);

- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;
  - indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.  
Important! Keywords must not be sentences;
  - Acknowledgments.
9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.net/ru/bgn/>).
- Please take into account the following examples of Russian references appearance:
- Article in journal:**  
Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.
- Journal article in electronic format:**  
Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).
- Article from the continuing publication (collection of works, proceedings):**  
Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental’noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem”* [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]. 593:125–130.
- Conference proceedings:**  
Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursoberegayushchie tekhnologii nedropol’zovaniya i povysheniya neftegazooitdachi”* [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]. Moscow. 267–272.
- Books and other monographs:**  
Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.
- Dissertation and Thesis:**  
Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovary informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 23 p.
- State standards and patents:**  
GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch’yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.  
Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel’nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.
- References in Latin transcription are presented in the original language.  
References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.
10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
- the journal title and author’s name in the “Subject” field;
  - an article and additional materials have to be attached using the “attach” function;
  - an electronic version of the article should contain the file with the text and a separate file with figures.
12. “Informatics and Applications” journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

**Editorial Board address:**

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia  
Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05  
e-mail: [iiep@frccsc.ru](mailto:iiep@frccsc.ru) (to Svetlana Strigina)  
<http://www.ipiran.ru/english/journal.asp>