Информатика и её применения

Том 18 Выпуск 2 Год 2024

СОДЕРЖАНИЕ

О функторном представлении оптимизируемых динамических мультиагентных систем
Н. С. Васильев **2**

Рынок с марковской скачкообразной волатильностью V:
пополнение рынка деривативами
А. В. Борисов **9**

Нижняя граница погрешности оценивания случайного параметра при заданном количестве информации
М. М. Ланге, А. М. Ланге **17**

Вероятностная модель затухания мощности сигнала в сценариях 3GPP TR 38.901 развертывания сети 5G
Е. Д. Макеева, И. А. Кочеткова, С. Я. Шоргин **25**

Метод оценки характеристик систем 5G/6G «новое радио» с учетом макро- и микромобильности пользователей
Д. Ю. Острикова, Е. С. Голос, В. А. Бесчастный, Е. А. Мачнев, В. С. Шоргин, Ю. В. Гайдамака **32**

Об однопороговом управлении очередью в системе массового обслуживания с нетерпеливыми заявками
Я. М. Агаларов **40**

О порождении синтетических признаков на основе опорных цепей и произвольных метрик в рамках топологического подхода к анализу данных.
Часть 2. Экспериментальная апробация на задачах фармакоинформатики
И. Ю. Торшин **47**

Выявление причинно-следственных связей при покрытии причин
А. А. Грушо, Н. А. Грушо, М. И. Забежайло, В. В. Кульченков, Е. Е. Тимонина **54**

Применение разложения изображения с помощью дискретного вейвлет-преобразования для построения архитектуры шумоподавляющей нейронной сети
А. С. Коваленко **60**

О применении генеративных моделей в системе электронного обучения математическим дисциплинам
А. В. Босов, А. В. Иванов **72**

Трансформации объектов первого и второго порядка в лексикографической информационной системе
И. М. Зацман **82**

Об авторах **92**

Правила подготовки рукописей **94**
Requirements for manuscripts **97**

Технический редактор *Л. Кокушкина* Художественный редактор *М. Седакова*
Подписано в печать 20.06.24. Формат 60 x 84 / 8
Бумага офсетная. Печать цифровая. Усл.-печ. л. 11,39. Уч.-изд. л. 10,5. Тираж 100 экз.
Заказ № 1217
Издательство «ТОРУС ПРЕСС», Москва 121614, ул. Крылатская, 29-1-43
Отпечатано в НИПКЦ «Восход-А» с готовых файлов
Москва 123103, наб. Новикова-Прибоя, д. 3, корп. 2

О ФУНКТОРНОМ ПРЕДСТАВЛЕНИИ ОПТИМИЗИРУЕМЫХ ДИНАМИЧЕСКИХ МУЛЬТИАГЕНТНЫХ СИСТЕМ

Н. С. Васильев[1]

Аннотация: Топос функторов выбран в качестве компьютерного инструмента синтеза динамических игр многих лиц. Задаваемая шкала упорядочивает объекты, отвечающие сопутствующим статическим подыграм. Последние служат состояниями динамической мультиагентной системы (ДМАС). Исходная динамическая игра и все статические подзадачи представляются в моноидальной категории бинарных отношений. Под рациональным решением игры понимается равновесие. Композициональное строение оптимизируемой ДМАС выражено в форме динамического результирующего отношения (ДРО) игры. Поиску равновесия отвечает максимизация ДРО. Это делается методом Беллмана, обобщенным на задачи оптимального управления, поставленные в форме отношений. Программная реализация предложенного подхода может быть основана на нейросетевых вычислениях ввиду согласованности архитектур применяемых графов отношений и нейросетей.

Ключевые слова: категория функторов; композициональность; моноидальная категория; обратный образ; динамическое отношение игры; статическая подыгра; отношение предпочтения; динамическое результирующее отношение; рациональное решение; морфизм Беллмана

DOI: 10.14357/19922264240201 **EDN:** CLMBXC

1 Введение

В ДМАС агенты принимают решения в каждом состоянии системы в зависимости от располагаемой ими информации о поведении других участников конфликта [1–4]. Воздействие стратегий на ДМАС распределено во времени и связано с выбором ситуаций в череде сопутствующих статических игровых подзадач.

Процесс изменения состояний ДМАС традиционно задают с помощью дифференциальных или итерационных уравнений, в которые входят управляющие воздействия игроков. От этого описания динамики приходится отходить даже в антагонистических дифференциальных играх с полной информированностью игроков. Управляемую систему представляют дифференциальным уравнением в контингенциях [3], т. е. применяют многозначные отображения. В стохастических постановках также возникает дополнительная неопределенность, связанная с прогнозированием будущих состояний системы [4].

Зачастую целесообразно отказаться от функционального описания динамики игры и перейти на язык отношений [5]. Благодаря этому значительно расширяется круг приложений [4, 6–9], приобретается свойство композициональности моделей ДМАС, удобное для модификации игровых задач. Этот подход поддерживается компьютерной алгеброй теории категорий [10, 11]. Кроме того, графовая структура отношений и сетевая структура игровой задачи допускают эффективную нейросетевую программную реализацию [7, 8, 12, 13].

Категорный подход охватывает общий случай динамических игр многих лиц с разнообразными классами допустимых стратегий игроков [5, 11]. В качестве состояний динамической системы рассматриваются сопутствующие статические игры, связанные динамическим отношением.

Традиционное моделирование игровой задачи проводится средствами категории множеств SET. Естественным обобщением классического подхода становится формализация игровой операции, выполняемая на языке моноидальной категории бинарных отношений REL. Применяемый аппарат позволяет единообразно выражать и модифицировать интересы агентов, проводить эквивалентные преобразования игровой задачи, описывать поиск рационального решения [5]. С помощью введения ДРО игры исследование ДМАС может быть сведено к последовательной максимизации ДРО.

Итак, вместо многошагового процесса игры предлагается перейти к представлению динамики задачи с помощью функтора $\tau : S \to \mathrm{SET}_T$, который преобразует выбираемую шкалу S в категорию SET_T монады (T, η, ψ), $T : A \to 2^A$ [11]. Объектами монады служат конечные множества, а морфизмами — отношения между ними. Равенство $\mathrm{SET}_T =$

[1]Московский государственный технический университет имени Н. Э. Баумана, nik8519@yandex.ru

$=$ REL означает, что функторная модель ДМАС строится на языке отношений. Единообразно выражаются правила игры, интересы участников, классы стратегий агентов, динамика задачи и даже алгоритм ее решения. Строгость функтора вложения $\mathrm{SET} \to \mathrm{SET}_T$ обеспечивает преемственность новой формы представления задачи и классической.

Функтор τ порождает динамическое отношение игры $R = R(\tau)$, регламентирующее смену состояний — решаемых в текущий момент времени статических подыгр $\Gamma(X)$ с множеством допустимых ситуаций X. Динамическое отношение определяет последовательность смены состояний $\Gamma(X_0), \ldots, \Gamma(X_T)$ и вводит отношение предшествования подыгр $\Gamma(X) \overset{R}{\prec} \Gamma(Y)$. В соответствии с ним определена преемственность ситуаций $x \overset{R}{\prec} y$, причем выбор любой пары ситуаций $x \in X$ и $y \in Y$, относящихся к разным состояниям системы, возможен, только если выполнено включение $(x, y) \in R$. Таким образом, из решений подзадач $\Gamma(X_0), \ldots, \Gamma(X_T)$ строится последовательность $x_0, \ldots, x_T$ выбираемых ситуаций подобно тому, как это делается в многошаговых играх с помощью уравнений.

Интересы агентов в задаче заданы в форме отношений предпочтения для терминального состояния системы $\Gamma(X^T)$:

$$\tilde{\rho}^{kT} \in \mathrm{REL}\left(X^T, X^T\right), \quad k \in K. \tag{1}$$

Правила функционирования ДМАС формулируются во всех подыграх $\Gamma(X_i)$. Во-первых, определены цели игроков — максимизация отношений предпочтения $\rho_i^k \in \mathrm{REL}\,(X_i, X_i)$, $k \in K(X_i)$, получаемых из заданных (1) с помощью вычисления обратных образов морфизмов $\tilde{\rho}^{kT}$ относительно динамического отношения R. Во-вторых, агенты выбирают классы допустимых стратегий посредством графа коммуникаций и функции, размечающей его дуги. Тем самым задается сетевая структура в статических подыграх [12, 13]. Разметка графа детализирует информацию, которой обмениваются партнеры при совершении ходов и формировании коалиций [5]. Стратегии агентов в исходной динамической задаче включают процедуру принятия решений в каждом состоянии системы.

Далее изучено композициональное строение динамического результирующего отношения игры, позволяющее искать рациональное решение, т. е. такое, которое оптимизирует функционирование ДМАС. Поиск проводится методом динамического программирования, обобщенным на задачи оптимального управления, поставленные в форме отношений.

2 Постановка задачи

В качестве шкалы S выберем категорию

$$S : L \underset{\text{in}}{\overset{\text{out}}{\rightrightarrows}} M. \tag{2}$$

Динамику мультиагентной системы зададим функтором $\tau : S \to \mathrm{SET}_T$, которому отвечает ориентированный граф динамического отношения Γ_R, $R = R(\tau)$. Граф должен быть ациклическим и связным. В его вершинах находятся множества допустимых ситуаций статических подыгр $\Gamma(X_i)$, а дуги отвечают отношениям $R_{ij} \in \mathrm{REL}\,(X_i, X_j)$.

Определение 2.1. Скажем, что ситуация $x_i \in X_i$ непосредственно предшествует ситуации $x_j \in X_j$, если и только если $(x_i, x_j) \in R_{ij}$. Отношением предшествования состояний назовем транзитивное замыкание $\overline{R} = \overline{R}_{ij}$ динамического отношения.

Последовательная смена состояний ДМАС происходит в соответствии с отношением непосредственного предшествования $\Gamma(X_i) \overset{R_{ij}}{\prec} \Gamma(X_j)$. Каждому состоянию соответствует некоторая статическая подыгра. Процесс функционирования системы складывается из последовательного выбора игроками ситуаций $x_0, \ldots, x_T$, где x_0 и x_T — решения начальной и терминальной подыгр. Итак, имеем множество всех допустимых ситуаций исходной динамической задачи вида

$$\tilde{X} = \\ = \left\{\tilde{x} = (x_0, \ldots, x_T) : (\forall_i x_i \in X_i) \wedge x_0 \overset{R}{\prec} \cdots \overset{R}{\prec} x_T\right\}.$$

Правила проведения каждой подыгры $\Gamma(X_i)$ задаются следующим функтором и функцией:

$$g_i : S \to \mathrm{SET}\,; \quad h_i : E \to \overline{2}\,. \tag{3}$$

Игроки стремятся по возможности максимизировать свои отношения предпочтения $\rho_l^k : X_l \to X_l$, $k \in K_l$. (Иначе рассматриваются противоположные отношения $(\rho_i^k)^{\mathrm{op}}$.) Функторы (2) и (3) определяют сетевую структуру игры. Шкала S выделяет множество дуг E, по которым участники $k \in K(X_i)$ статической игры осуществляют коммуникацию. Функциями in и out в (2) вводят порядок ходов. Функция h_i из формулы (3) осуществляет разметку дуг графа коммуникаций γ_i, определяя данные, которыми обмениваются игроки при своих ходах [5]. Непустые сообщения могут содержать либо выбранную стратегию-константу, либо стратегию-функцию [2, 5]. Предполагается также, что никто из агентов не блефует.

Понятие рационального поведения игроков зависит от сетевой структуры (3) статических игр [5, 12, 13]. Будем применять принцип равновесия как к исходной динамической задаче, так и ко всем по отдельности подыграм $\Gamma(X_i)$.

Пример 2.1. Рассмотрим функтор (2), которому отвечает граф

$$\Gamma_R : \boxed{X_1} \xrightarrow{R_{12}} \boxed{X_2} \xrightarrow{R_{23}} \cdots \xrightarrow{R_{n-1,n}} \boxed{X_n}. \qquad (4)$$

Функции out и in определяют начало $X_i \in M$ и конец $X_j \in M$ каждой дуги, обозначенной стрелкой $R_{i,i+1} \in L$. В моменты времени $i = 1, 2, \ldots, n$ решаются игры $\Gamma(X_i)$.

Схема (4) свойственна многошаговым операциям.

Пример 2.2. Пусть в позиционной многошаговой игре двух лиц известна следующая динамическая система, начальное условие и фазовое ограничение:

$$\begin{cases} y^1_{t+1} = y^1_t + u^1_t\,; \\ y^2_{t+1} = y^2_t - u^2_t\,; \end{cases}$$
$$y^1_0 = 0\,,\ \ y^2_0 = 0\,;$$
$$\left|y^1_t - y^2_t\right| \le 2\,,\ \ t = \overline{0,T}\,.$$

Агенты $k = 1, 2$ стремятся по возможности максимизировать функции выигрыша

$$J_1 = -\sum_{t=1}^{T-1} y_t u^1_t; \quad J_2 = \sum_{t=1}^{T-1} y_t u^2_t,$$

где

$$y_t = y^1_t - y^2_t, \quad u^k_t \in U_0 \equiv \{-1, 0, 1\}.$$

Интересы игроков $\tilde{\rho}^k \in \mathrm{REL}(\tilde{X}, \tilde{X})$ (1) заданы функционалами J_k. В статических подыграх $\Gamma(X_t)$, $X_t = \cup y_t X(y_t)$, $X_0 = U^2_0$, с помощью функций $f^1 = -y_t u^1_t$, $f^2 = y_t u^2_t$ введем отношения предпочтения игроков ρ^k_t. Видно, что интересы участников подыгр противоположны $\rho^2_t = \rho^{1\mathrm{op}}_t$, а множества допустимых ситуаций расслоены в зависимости от позиций $y_t = y^1_t - y^2_t$, $|y_t| \le 2$, в которых может пребывать динамическая система. Ситуация $(u^1_t, u^2_t) \in X(y_t)$ допустима, только если ее выбор не нарушает заданное фазовое ограничение.

Во все моменты времени отношения предпочтения $\rho^k_t(y_t) : X(y_t) \to X(y_t)$, $k = 1, 2$, одинаковы: $\rho^k_t = \rho^k_1$, $X_t = X_1$, $t = \overline{1, T-1}$, причем

$$\rho^k_t \triangleq \bigcup_{y_t} \rho^k_t(y_t).$$

Вычисления показывают, что при $t = \overline{1, T-1}$ имеем

$$X_t = \coprod_{y_t=-2}^{2} X(y_t),\ X(\pm 2) = \{(\pm 1, \pm 1)\},$$
$$X(\pm 1) = \{(\pm 1, 0), (0, \pm 1)\},$$
$$X(0) = \{(1, -1), (0, 0), (-1, 1)\}\,;$$

$$\rho^1_t(0) : (0, -1) \sim (-1, 0) \sim (0, 0),$$
$$\rho^1_t(1) = \{((0, -1), (-1, 0))\},$$
$$\rho^1_t(-1) = \{((-1, 0), (0, -1))\}\,,$$
$$\rho^1_1(\pm 2) = \{(\pm 1, \pm 1)\}.$$

Динамическое отношение игры, равное $R_{t,t+1} : X_t \to X_{t+1}$, $t = \overline{1, T-1}$, связывает допустимую текущую $u_m \equiv (u^1_m, u^2_m) \in X(y_m)$ и выбираемую следующую $u_{m+1} \in X(y_{m+1})$ ситуации в состояниях системы $\Gamma(X_m)$, $m = t, t+1$. Таким образом, отношение непосредственного предшествования ситуаций определено включением

$$u_t \xrightarrow{R_{tt+1}} u_{t+1},\ u_{t+1} \in X(y_{t+1}),\ y_{t+1} = y_t + \left(u^1_t + u^2_t\right).$$

Пусть сетевая структура (3) такова, что во всех играх $\Gamma(X_t)$ один из участников сообщает другому свое управляющее воздействие u^k_t. Тогда рациональное решение игры состоит в выборе ситуации, стабилизирующей траекторию системы $y_t \equiv 0$, а многошаговая игра имеет седловые точки $\forall_t u^1_t = -u^2_t$.

При поиске рационального решения игровых задач применяются вспомогательные морфизмы, строящиеся из исходных (1) с помощью операций алгебры отношений $A = (\circ, \cup, \cap, {}^{\mathrm{op}}, \times; \sigma, \varnothing)$ [5, 10, 11]. Получение каким-либо игроком дополнительной информации уменьшает неопределенность выбора подходящей стратегии (см. пример 2.2). Выбор сетевой структуры игры (3) изменяет отношения предпочтений участников ρ. Игроки руководствуются сужениями $\rho|_A = \rho \cap A^2$, $A \subset X$, исходных отношений ρ на некоторые, вполне определенные подмножества ситуаций.

При условии сделанных предположений граф динамического отношения содержит минимальные и максимальные элементы — вершины X^0 и X^T (см. пример 2.1). Процесс функционирования ДМАС начинается с решения статических подыгр $\Gamma(X^0)$ и заканчивается подзадачами $\Gamma(X^T)$. Они названы начальным и терминальным состояниями системы соответственно. В отличие от примера 2.2, в рассматриваемой постановке задаются терминальные отношения предпочтения агентов (1).

Рациональное поведение игроков заключается в том, что они стремятся максимизировать свои

предпочтения на множестве всех допустимых ситуаций:

$$\forall_k \tilde{\rho}^{kT} \to \underset{\tilde{X}}{\mathrm{MAX}}\,. \tag{5}$$

В задаче (1), (2), (3), (5) требуется найти ситуацию равновесия [2, 5].

Существование равновесия во многом зависит от свойств морфизмов ρ_i^k, $k \in K$. Наложим одно из таких требований. Пусть партнеры предлагают некоторому игроку k принять решение в ситуации s, для которой $\rho_i^k|_{s_k} = \varnothing$, $s \in X_i$. Предполагается, что возникшую неопределенность выбора каждый игрок $k \in K$ разрешит с помощью изменения своего отношения предпочтения, положив $\rho_i^k|_{s_k} = \{s\}$. (Игрок соглашается с выбором ситуации s.) Это требование выполняется для рефлексивных отношений.

3 Эквивалентные преобразования динамических игр

Категорное представление допускает применение моноидальных операций для преобразования игровой задачи [11]. Морфизмы $R : \mathrm{REL}\,(Y) \to \to \mathrm{REL}\,(X)$ представляют собой функторы, сравнивающие однообъектные подкатегории в категории REL. Образ морфизма $\rho \in \mathrm{REL}\,(Y, Y)$ относительно R будем записывать как композицию $R \circ \rho \in \in \mathrm{REL}\,(X, X)$.

Определение 3.1. Обратным образом (или кообразом) отношения $r : X \to X$ относительно морфизма $R : Y \to X$ назовем стрелку $r' : Y \to Y$, превращающую следующий квадрат в коммутативный (рис. 1).

Ввиду того что морфизм $R^{\mathrm{op}} \in \mathrm{REL}\,(X, Y)$ сохраняет композиции стрелок, он сам становится функтором $R^{\mathrm{op}} : \mathrm{REL}\,(X) \to \mathrm{REL}\,(Y)$. Так как $\forall_{ij}(x, y) \in R_{ij}^{\mathrm{op}} \Leftrightarrow (y, x) \in R_{ij}$, то назовем R^{op} противоположным к $\{R_{ij}\}$ динамическим отношением. Его можно изобразить, изменив направления всех дуг в графе Γ_R (вертикальных стрелок на рис. 1).

Пусть вершины X_1 и X_l графа динамического отношения Γ_R связаны некоторым путем $L = = (X_1, \ldots, X_l)$. Взяв композицию $R_L = R_{l-1} \circ \cdots \cdots \circ R_1$ морфизмов $R_i \in \mathrm{REL}\,(X_i, X_{i+1})$ вдоль цепи $L = (X_1, \ldots, X_l)$, можно «опустить» произвольное бинарное отношение $r_l : X_l \to X_l$ с X_l на множество X_1 и построить его обратный образ $r_1 = R_L^{\mathrm{op}} \circ r_l$.

Пусть теперь множество $\{L = (X_1, \ldots, X_l)\}$ всех попарно различных путей, связывающих вершины X_1 и X_l, состоит более чем из одного элемента. Под обратным образом отношения $\rho_l : X_l \to X_l$, переносимого на множество X_1, будем понимать копроизведение

$$\rho_1' = R^{\mathrm{op}} \circ \rho_l = \coprod_{\{L\}} R_L^{\mathrm{op}} \circ \rho_l;$$

$$\rho_1' : \coprod_{\{L\}} X_1 \to \coprod_{\{L\}} X_1\,. \tag{6}$$

«Поднятием» морфизма ρ_1 вдоль путей $\{L\}$ можно построить образ $\rho_L = R \circ \rho_1$. Это отношение $\coprod_{\{L\}} R_L \circ \rho_1$ на копроизведении объектов $\coprod_{\{L\}} X_l$ с числом сомножителей, равным мощности набора $\{L\}$.

По формуле (6) заданные в терминальных состояниях системы морфизмы (1) переносятся во все остальные состояния. Так вводятся отношения предпочтения игроков ρ_i^k в играх $\Gamma(X_i)$:

$$\forall_{ik} \rho_i^k = R^{\mathrm{op}} \circ \tilde{\rho}_i^{kT}.$$

Замечание 3.1. Общее определение кообраза (6) можно заменить двойственной конструкцией, основанной на произведении отношений [11]:

$$\rho'' = R^{\mathrm{op}} \circ \rho = \prod_{\{L\}} R_L^{\mathrm{op}} \circ \rho\,;\ \rho = R \circ \rho'' = \prod_{\{L\}} R_L \circ \rho''. \tag{7}$$

Получим морфизмы (7) вида

$$\rho : \prod_{\{L\}} X_1 \to \prod_{\{L\}} X_1,\ \rho'' : \prod_{\{L\}} X_l \to \prod_{\{L\}} X_l.$$

Пример 3.1. Воспользуемся формулой (6) применительно к ДМАС, представленной на рис. 2.

Кообразом $R^{\mathrm{op}} \circ \rho_4$ отношения $\rho_4 : X_4 \to X_4$ служит морфизм

$$\rho_1' : X_1 \coprod X_1 \to X_1 \coprod X_1;$$

$$\rho_1' = \left(R_{13}^{\mathrm{op}} \circ R_{34}^{\mathrm{op}} \coprod R_{12}^{\mathrm{op}} \circ R_{24}^{\mathrm{op}}\right) \circ \rho_4.$$

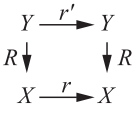


Рис. 1 Образ $r = R \circ r'$ и обратный образ $r' = R^{\mathrm{op}} \circ r$ относительно функтора R

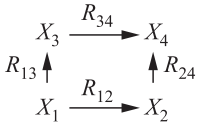


Рис. 2 Граф Γ_R динамического отношения R: $R_{ij} \in \in \mathrm{REL}\,(X_i, X_j)$

Опуская ρ_4 на множества X_2 и X_3, получим $\rho_2' = R_{24}^{\mathrm{op}} \circ \rho_4$, $\rho_3' = R_{34}^{\mathrm{op}} \circ \rho_4$ соответственно. Поднятием отношения $\rho_1' : X_1 \to X_1$ на множество X_4 строится образ $R \circ \rho_1'$ как

$$\rho_4 : X_4 \coprod X_4 \to X_4 \coprod X_4;$$
$$\rho_4 = \Big(R_{24} \circ R_{12} \coprod R_{34} \circ R_{13}\Big) \circ \rho_1'.$$

Конструкция (7) дает структуру

$$\rho_4 : X_4 \times X_4 \to X_4 \times X_4;$$
$$\rho_4 = \Big(R_{24} \circ R_{12} \prod R_{34} \circ R_{13}\Big) \circ \rho_1''.$$

Формулы (6) и (7) позволяют строить эквивалентные модели игры. Свойство универсальности объектов $X_2 \times X_3$ и $X_2 \coprod X_3$, выражаемое посредством коммутативных диаграмм [11], однозначно определяет новое динамическое отношение с соответствующими морфизмами. Например, вместо графа из рис. 2 можно работать с более простой графовой структурой, представленной в любой из следующих форм:

$$\boxed{X_1} \xrightarrow{\langle R_{12}, R_{13}\rangle} \boxed{X_2 \times X_3} \xrightarrow{\left\langle R_{24}^{\mathrm{op}}, R_{34}^{\mathrm{op}}\right\rangle^{\mathrm{op}}} \boxed{X_4}; \quad (8)$$

$$\boxed{X_1} \xrightarrow{[R_{12}, R_{13}]^{\mathrm{op}}} \boxed{X_2 \coprod X_3} \xrightarrow{[R_{24}, R_{34}]} \boxed{X_4}. \quad (9)$$

Возможность эквивалентного перехода от произвольного графа Γ_R (см. рис. 2) к цепи (4) (см. (8) и (9)) обоснована в теореме 3.1.

Теорема 3.1. *Граф всякого динамического отношения приводится к виду* (4).

4 Результирующее отношение динамической игры

Введением результирующего отношения игра сводится к проблеме оптимального управления, поставленной в форме отношений. Для ее решения можно применить обобщенный метод динамического программирования. Напомним [5], что в любой статической игре $\Gamma(Y)$ существует результирующее отношение P_Y. Оно выражает композициональное свойство игры, учитывающее интересы всех участников операции, сетевую структуру их взаимодействия и рациональность поведения. Решение игры $\Gamma(Y)$ сводится к поиску максимальных элементов $P_Y \to \mathrm{MAX}_Y$. Благодаря этому, можно по-новому определить состояния исходной ДМАС. Вместо игр $\Gamma(X_i)$ будем рассматривать оптимизационные задачи $(\tilde{X}_i, P_{\tilde{X}_i})$, $P_{\tilde{X}_i} \in \in \mathrm{REL}\,(\tilde{X}_i, \tilde{X}_i)$. Иначе говоря, теперь во всех состояниях системы решение принимает единственный агент.

Определение 4.1. Пусть $(\tilde{Y}_1, P_{\tilde{Y}_1}) \overset{R}{\prec} (\tilde{Y}_2, P_{\tilde{Y}_2})$; $\tilde{y}_1, \tilde{y}_2 \in \in \tilde{Y}_1$. Ситуация $\tilde{y}_2$ называется более перспективной по сравнению с $\tilde{y}_1$, если выполнено свойство

$$(\tilde{y}_1, \tilde{y}_2) \in \left(R^{\mathrm{op}} \circ P_{\tilde{Y}_2}\right) \circ P_{\tilde{Y}_1}. \quad (10)$$

В каждом состоянии ДМАС игрокам целесообразно использовать более перспективные ситуации и из них формировать рациональное решение задачи (1), (2), (3), (5). В этом заключается принцип Беллмана.

В случае динамического отношения с графом (4) (см. теорему 3.1) рассмотрим следующую итерационную схему построения «оптимизирующих» морфизмов $\{\tilde{P}_{\tilde{X}_k}\}$:

$$\tilde{P}_{\tilde{X}_T} = P_{X_T},\ \tilde{P}_{\tilde{X}_{T-k}} = \left(R^{\mathrm{op}} \circ \tilde{P}_{\tilde{X}_{T-k+1}}\right) \circ P_{\tilde{X}_{T-k}},$$
$$k = \overline{1, T}. \quad (11)$$

Назовем (11) уравнениями Беллмана в форме отношений.

Определение 4.2. Динамическим результирующим отношением называется семейство морфизмов Беллмана $\{\tilde{P}_{\tilde{X}_k},\ k = \overline{1, T}\}$ из уравнений (11).

Теорема 4.1. *Во всякой ДМАС существует результирующее отношение.*

Доказательство. Построение ДРО проведем методом математической индукции. На начальном шаге, в терминальном состоянии системы, ДРО совпадает с результирующим отношением $\tilde{P}_{X^T} \triangleq P_{X^T}$, $\tilde{X}^T = X^T$, статической игры $\Gamma(X^T)$. Опустим этот морфизм на все множества $\tilde{X}^{T-1}$, для которых имеет место непосредственное предшествование $\Gamma(\tilde{X}^{T-1}) \overset{R}{\prec} \Gamma(\tilde{X}^T)$. Композиция морфизмов (10), $\tilde{Y}_1 = \tilde{X}^{T-1}$, $\tilde{Y}_2 = \tilde{X}^T$, определяет компоненту ДРО $\tilde{P}_{\tilde{X}^{T-1}}$, отвечающую состоянию системы $(\tilde{X}^{T-1}, P_{\tilde{X}^{T-1}})$. Пусть отношение $\tilde{P}_{\tilde{X}_i}$ уже построено. Тогда опять по формуле (10) его можно продолжить на все состояния $\Gamma(\tilde{X}_j) \overset{R_{ij}}{\prec} \Gamma(\tilde{X}_i)$ исходной ДМАС, т. е. построить очередные морфизмы $\tilde{P}_{\tilde{X}_j}$. Процесс завершается в начальных состояниях системы.

5 Поиск рационального решения игры

Для решения задачи (1), (2), (3), (5) предложим следующее обобщение метода динамического программирования. Сначала из системы уравнений Беллмана (11) найдем ДРО игры. Затем последовательно для моментов времени $1, \ldots, T$ вычислим следующие множества максимальных элементов отношения Беллмана:

$$X_1^* = \mathrm{ARGMAX}\ \tilde{P}_{\tilde{X}_1};\ \ X_2^* = \underset{RX_1^*}{\mathrm{ARGMAX}}\ \tilde{P}_{\tilde{X}_2}; \ldots$$
$$\ldots; X_T^* = \underset{RX_{T-1}^*}{\mathrm{ARGMAX}}\ \tilde{P}_{\tilde{X}_T}. \quad (12)$$

Теорема 5.1. *Рациональному решению динамической игровой задачи* (1)–(5) *отвечает ситуация* $\tilde{x}^* = (x_1^*, \ldots, x_T^*)$; $x_s^* \in X_s^*$, $s = \overline{1,T}$, *найденная методом Беллмана* (11), (12).

Пример 5.1. Графом Γ_R (4), $n = 2$, динамического отношения задана ДМАС

$$R : (x_1, x_2, x_3) \to (x_3, x_1, x_1 \vee x_2) \cup (x_2, x_3, x_1 \wedge x_2);$$
$$x = (x_1, x_2, x_3) \in \underline{8} \simeq \{0,1\}^3.$$

Терминальная задача представляет собой игру Гермейера трех лиц $\Gamma^{2,3}$ [2, 5], где $X^T = \underline{8}$ — бинарный куб. Отношения предпочтений агентов $\tilde{\rho}_2^k \in \in \mathrm{REL}\,(\underline{8}, \underline{8})$ равны:

$$\tilde{\rho}_2^1 = \{01, 05, 40, 34, 32, 76\};$$
$$\tilde{\rho}_2^2 = \{02, 10, 24, 32, 35, 45, 64, 67\};$$
$$\tilde{\rho}_2^3 = \{20, 40, 35, 31, 64, 76\}.$$

Найдем интересы игроков $\rho_1^k = R^{\mathrm{op}}(\tilde{\rho}_2^k)$ в начальном состоянии системы $\Gamma^{1,3}$, $X^0 = \underline{8}$:

$$\rho_1^1 = \{01, 02, 03, 04, 06, 12, 13, 14, 16, 20, 21, 62, 63, 64, 65, 72, 73, 74, 75\};$$

$$\rho_1^2 = \{03, 04, 05, 13, 14, 15, 20, 21, 23, 26, 32, 40, 41, 42, 52, 61, 63, 64, 65, 71, 73, 74, 75, 76\};$$

$$\rho_1^3 = \{20, 21, 30, 31, 40, 41, 50, 51, 61, 62, 63, 64, 71, 72, 73, 74, 76\}.$$

Вычислим результирующие отношения статических подыгр $\Gamma(X^T), \Gamma(X^0)$ [5]:

$$P_{X^T} = \tilde{\rho}_1^3|_{x_3} \circ \tilde{\rho}_1^2 \circ \left(\tilde{\rho}_1^1 \cup \tilde{\rho}_1^{2G}\right),$$
$$\rho_1^{2G} = \tilde{\rho}_1^2|_{\mathrm{MIN}\tilde{\rho}_1^2|_{x_1}} \Rightarrow P_{X^T} = \{40, 05, 42, 35, 30, 70\},$$
$$P_{X^0} = \rho_1^1|_{x_1} \circ \rho_1^2|_{x_2} \circ \rho_1^3|_{x_3} = \{41, 53, 61, 73\}.$$

Из уравнений Беллмана (11) найдем искомое ДРО игры:

$$\tilde{P}_{X^T} = \{40, 05, 42, 35, 30, 70\},\ \tilde{P}_{X^0} = \{40, 00, 60, 70\}.$$

По формулам (12) и теореме 5.1 рациональное решение игры равно $\tilde{x}^* = (0, 0)$.

6 Заключение

Функторная модель стала наследником традиционной формы представления ДМАС. Обладая композициональной структурой, она создает условия для модификации игровой задачи и выполнения эквивалентных преобразований средствами алгебры теории категорий. Доказано существование результирующего отношения динамической игры многих лиц. Для задач оптимального управления в форме динамических отношений предложено обобщение метода Беллмана. С помощью ДРО этим методом строится рациональное решение задачи. Функторный подход поддерживается компьютерной алгеброй теории категорий. Сетевая архитектура применяемых морфизмов допускает эффективную нейросетевую программную реализацию, которую еще только предстоит осуществить.

Литература

1. *Моисеев Н. Н.* Элементы теории оптимальных систем. — М.: Наука, 1974. 526 с.
2. *Гермейер Ю. Б.* Игры с непротивоположными интересами. — М.: Наука, 1976. 326 с.
3. *Красовский Н. Н., Субботин А. И.* Позиционные дифференциальные игры. — М.: Наука, 1976. 456 с.
4. *Dockner E. J., Jorgensen S., Long N. V., Sorger G.* Differential games in economics and management science. — Cambridge: Cambridge University Press, 2000. 382 p. doi: 10.1017/CBO9780511805127.
5. *Васильев Н. С.* Композициональное представление структуры игры многих лиц в моноидальной категории бинарных отношений // Информатика и её применения, 2023. Т. 17. Вып. 2. С. 18–26. doi: 10.14357/19922264230203. EDN: GPMZTS.
6. *Dixit A. K., Natebuff B. J.* The art of strategy. — New York; London: W. W. Norton & Co., 2008. 446 p.
7. *Shoham Y., Leyton-Brown R.* Multiagent systems: Algorithmic, game-theoretic, and logical foundations. — Cambridge: Cambridge University Press, 2010. 532 p.
8. *Bai Q., Ren F., Fujita K., Znang M.* Multi-agent and complex systems. — Studies in computational intelligence ser. — Springer Singapore, 2016. Vol. 670. 210 p.
9. *Dixit A. K., Skeath S., Reiley W. W., Jr.* Games of strategy. — New York; London: W. W. Norton & Co., 2017. 880 p.
10. *Скорняков Л. А.* Элементы общей алгебры. — М.: Наука, 1983. 272 с.
11. *Маклейн С.* Категории для работающего математика / Пер. с англ. — М.: Физматлит, 2004. 352 с. (*Mac Lane S.* Categories for the working mathematician. — 2nd ed. — New York, NY, USA: Springer, 1998. 318 p.)
12. *Губко М. В.* Управление организационными системами с сетевым взаимодействием агентов. Обзор теории сетевых игр // Автоматика и телемеханика, 2004. № 8. С. 115–132.
13. Group formation in economics: Networks, clubs, and coalitions / Eds. G. Demange, M. Wooders. — Cambridge: Cambridge University Press, 2005. 475 p.

Поступила в редакцию 02.02.24

ON FUNCTOR REPRESENTATION OF OPTIMIZED DYNAMIC MULTIAGENT SYSTEMS

N. S. Vasilyev

N. E. Bauman Moscow State Technical University, 5-1 Baumanskaya 2nd Str., Moscow 105005, Russian Federation

Abstract: Functors' topoi is chosen as a computational tool for synthesizing dynamic multiagent systems (DMAS). The scale orders the objects as multiagent system states to solve attendant static subgames in them. The initial dynamic game and all static subproblems are represented in the monoidal category of binary relations. Players' preference relations might be maximized in DMAS. The game rational solution is understood as equilibrium. The compositional structure of the optimized DMAS can be described in the form of the game dynamic resulting relation (DRR). Players' rational behavior search is reduced to DRR subsequent maximization. For this purpose, the Bellman's method is generalized to solve control problems stated in the form of relations. The program implementation of the approach can be based on neural networks due to the consistency of the architectures of the applied relation graphs and neural networks.

Keywords: functor category; compositionality; monoidal category; opposite image; game dynamic relation; static subgame; preference relation; dynamic resulting relation; rational solution; Bellman morphism

DOI: 10.14357/19922264240201 **EDN:** CLMBXC

References

1. Moiseev, N. N. 1975. *Elementy teorii optimal'nykh sistem* [Elements of optimal systems theory]. Moscow: Nauka. 527 p.
2. Germeyer, Yu. B. 1976. *Igry s neprotivopolozhnymi interesami* [Games with nonopposing interests]. Moscow: Nauka. 326 p.
3. Krasovskiy, N. N., and A. I. Subbotin. 1974. *Pozitsionnye differentsial'nye igry* [Positional differential games]. Moscow: Nauka. 456 p.
4. Dockner, E. J., S. Jorgensen, N. V. Long, and G. Sorger. 2000. *Differential games in economics and management science*. Cambridge: Cambridge University Press. 382 p. doi: 10.1017/CBO9780511805127.
5. Vasilyev, N. S. 2023. Kompozitsional'noe predstavlenie struktury igry mnogikh lits v monoidal'noy kategorii binarnykh otnosheniy [Multiplayers' games compositional structure in the monoidal category of binary relations]. *Informatika i ee Primeneniya — Inform. Appl.* 17(2):18–26. doi: 10.14357/19922264230203. EDN: GPMZTS.
6. Dixit, A. K., and B. J. Nalebuff. 2008. *The art of strategy*. New York, London: W. W. Norton & Co. 446 p.
7. Shoham, Y., and R. Leyton-Brown. 2010. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press. 532 p.
8. Bai, Q., F. Ren, K. Fujita, and M. Znang. 2016. *Multiagent and complex systems*. Studies in computational intelligence ser. Springer Singapore. 210 p.
9. Dixit, A. K., S. Skeath, and D. H. Reiley, Jr. 2017. *Games of strategy*. New York, London: W. W. Norton & Co. 880 p.
10. Skornyakov, L. A. 1983. *Elementy obshchey algebry* [Elements of general algebra]. Moscow: Nauka. 272 p.
11. Mac Lane, S. 1998. *Categories for the working mathematician*. 2nd ed. New York, NY: Springer. 318 p.
12. Gubko, M. V. 2004. Control of organizational systems with network interaction of agents. II. Stimulation problems. *Automat. Rem. Contr.* 65(9):1470–1485. doi: 10.1023/B:AURC.0000041425.34118.7d. EDN: LFMUCG.
13. Demange, G., and M. Wooders, eds. 2005. *Group formation in economics: Networks, clubs, and coalitions*. Cambridge: Cambridge University Press. 475 p.

Received February 2, 2024

Contributor

Vasilyev Nikolai S. (b. 1952) — Doctor of Science in physics and mathematics, professor, N. E. Bauman Moscow State Technical University, 5-1 Baumanskaya 2nd Str., Moscow 105005, Russian Federation; nik8519@yandex.ru

РЫНОК С МАРКОВСКОЙ СКАЧКООБРАЗНОЙ ВОЛАТИЛЬНОСТЬЮ V: ПОПОЛНЕНИЕ РЫНКА ДЕРИВАТИВАМИ*

А. В. Борисов[1]

Аннотация: Пятая, заключительная часть цикла посвящена задаче пополнения рынка с марковской скачкообразной сменой режимов. Рынок включает в себя безрисковый банковский депозит с известной неслучайной процентной ставкой, а также набор базовых рисковых активов. Мгновенные процентные ставки и волатильности этих активов представляют собой функции скрытого фактора смены режимов, описываемого некоторым марковским скачкообразным процессом (МСП) с конечным множеством состояний. Целью статьи ставится пополнение предложенного рынка. Оно означает добавление некоторого набора новых финансовых инструментов таким образом, что выплаты по любому платежному требованию, сформированному на рынке, могут быть воспроизведены с помощью некоторого самофинансируемого портфеля, содержащего исходные и добавленные инструменты. Доказано, что для пополнения рынка к исходному множеству активов достаточно добавить деривативы европейского типа, построенные на уже представленных на рынке базовых рисковых активах. При этом число добавляемых деривативов совпадает с числом режимов рынка. Задача пополнения рынка имеет неединственное решение, и в статье приведено сравнение предложенного способа с уже существующим.

Ключевые слова: марковский скачкообразный процесс; портфель ценных бумаг; свойство самофинансирования; полнота рынка

DOI: 10.14357/19922264240202 **EDN:** DQFSDO

1 Введение

Предметом исследования заключительной заметки цикла [1–4] стала модель рынка, состоящего из безрискового банковского депозита и набора базовых рисковых активов. Мгновенная процентная ставка и волатильность этих активов зависят от режима рынка, описываемого некоторым ненаблюдаемым МСП. Рынки такого типа считаются неполными [5], так как содержат «неторгуемый» (*nontradable*) случайный процесс [6]. Этот факт усложняет теоретическое и практическое решение задач определения справедливых цен деривативов, хеджирования и оптимального инвестирования. Одним из направлений сокращения неопределенности в данных моделях стало оценивание режима рынка по имеющейся статистической информации: журналам заявок на покупку/продажу финансовых инструментов (базовых и производных), а также цен проведенных сделок. Предыдущие части цикла [1, 3, 4] содержали результаты исследований в этой области.

Другой способ сокращения данной неопределенности — пополнение рынка. Эта процедура означает добавление к уже имеющимся на рынке некоторых дополнительных инструментов, несущих информацию о неторгуемых процессах. Пополненный набор инструментов должен давать возможность воспроизводить выплаты по любым платежным требованиям, которые можно определить на рынке. Именно задаче пополнения рынка посвящена эта заметка.

Статья организована следующим образом. В разд. 2 приведено детальное описание модели исследуемого рынка. Раздел 3 содержит необходимые определения портфельных инвестиций, а также постановку задачи пополнения рынка. Раздел 4 — основной в данной работе: он представляет набор деривативов европейского типа от существующих рисковых активов, пригодный для пополнения рынка. В разд. 4 также приводится сравнение данного набора деривативов со скачкообразными финансовыми инструментами, предложенными в [7, 8] с аналогичной целью. Раздел 5 содержит выводы по проведенным исследованиям.

2 Исследуемая модель рынка

На вероятностном базисе с фильтрацией $(\Omega, \mathcal{F}, \mathcal{P}, \{\mathcal{F}_t\}_{t\in[0,T]})$ рассматривается модель финансового рынка, изначально состоящего из банковского вклада

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

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, aborisov@frccsc.ru

$$B_t = \exp\left(\int_0^t r_u\, d_u\right)$$

с детерминированной ставкой r_t и N базовых рисковых инструментов

$$S_t = \operatorname{col}(S_t^1, \ldots, S_t^N).$$

Цена S_t описывается стохастической дифференциальной системой (СДС)

$$dS_t = \operatorname{diag} S_t a(t, Z_t)\, dt + \operatorname{diag} S_t \sigma(t, Z_t)\, dw_t, \quad t \in (0, T],\ S_0 \sim p_0(s), \quad (1)$$

где $w_t \triangleq \operatorname{col}(w_t^1, \ldots, w_t^N)$ — N-мерный винеровский процесс, а случайные функции мгновенной процентной ставки a и внутренней волатильности σ имеют вид

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

с набором известных детерминированных функций $\{a^\ell(t)\}_{\ell=\overline{1,L}}$ и $\{\sigma^\ell(t)\}_{\ell=\overline{1,L}}$:

$$a^\ell(t) \triangleq \operatorname{col}(a_1^\ell(t), \ldots, a_N^\ell(t)); \sigma^\ell(t) = \|\sigma_{ij}^\ell(t)\|_{i,j=\overline{1,N}}.$$

В функциях $a(\cdot)$ и $\sigma(\cdot)$ $Z_t \triangleq \operatorname{col}(Z_t^1, \ldots, Z_t^L) \in \{e_1, \ldots, e_L\}$ — МСП с матрицей интенсивностей переходов $\Lambda(\cdot)$ и начальным распределением p_0^Z. Марковский скачкообразный процесс Z_t представляет собой решение СДС с мартингалом $M_t = \operatorname{col}(M_t^1, \ldots, M_t^L)$ в правой части [9]:

$$dZ_t = \Lambda^\top(t) Z_t\, dt + dM_t,\ \ t \in (0, T],\ Z_0 \sim p_0^Z. \quad (2)$$

Для вероятностного базиса с фильтрацией и модели (1), (2) выполнены следующие предположения.

1. С вероятностью 1 траектории процесса Z_t непрерывны справа и имеют конечные пределы слева [10].
2. $\mathcal{F}_t = \sigma\{S_0,\ w_u,\ Z_u :\ 0 \leqslant u \leqslant t\}$ для любых $t \in [0, T]$.
3. Для некоторого $\alpha > 0$ неравенство $\sigma^\ell(t)\left(\sigma^\ell(t)\right)^\top \geqslant \alpha I$ выполняется для всех $t \in [0, T]$ и $\ell = \overline{1, L}$.
4. $\mathcal{P}\{S_0 > 0\} = 1$.
5. $p_0^Z = \operatorname{col}(p_0^{Z,1}, \ldots, p_0^{Z,L})$: $\min_{1 \leqslant \ell \leqslant L} p_0^{Z,\ell} > 0$.
6. Начальные условия S_0 и Z_0 независимы.

Несмотря на технический характер, предположения 1–6 необременительны и имеют очевидный смысл. Предположения 1, 2 и 6 гарантируют непрерывность справа потока $\{\mathcal{F}_t\}$ — стандартное требование для правомерного применения математического аппарата стохастического анализа. Предположение 2 также означает, что вся случайность на рынке порождается только начальным условием S_0, процессами w и Z. Совместно с предположением 3 оно также обеспечивает совпадение фильтраций

$$\mathcal{F}_t \equiv \sigma\{S_u,\ Z_u :\ 0 \leqslant u \leqslant t\},\ \ t \in [0, T].$$

Предположение 4 с вероятностью 1 гарантирует положительность цен базовых рисковых активов в любой момент времени. Предположение 5 обеспечивает строгую положительность всех компонентов вектора распределения p_t^Z МСП Z при $t \in [0, T]$.

3 Задача пополнения рынка

Для корректной постановки задачи пополнения исследуемой модели рынка необходимо ввести ряд дополнительных определений [11]. Пусть $\mathcal{S}_t \triangleq \sigma\{S_u :\ 0 \leqslant u \leqslant t\}$ — естественный поток σ-алгебр, порожденный ценами базовых финансовых инструментов. Он задает структуру информации, доступной трейдерам для выработки торговых стратегий — портфелей ценных бумаг. Заметим, что для любого $t \in [0, T]$ в исследуемой модели рынка выполняется строгое включение $\mathcal{S}_t \subset \mathcal{F}_t$, причем поток σ-алгебр $\{\mathcal{S}_t\}$ не обладает свойством непрерывности справа в отличие от $\{\mathcal{F}_t\}$. С прикладной точки зрения это значит, что не все случайные процессы, влияющие на рынок, доступны трейдерам в форме цен тех или иных бумаг: имеется неторгуемый МСП Z_t. Данный процесс не является ценой какой-либо бумаги, но скрыто влияет на цены S_t исходных инструментов.

Портфелем ценных бумаг называется пара (ϖ, π), где

- $\varpi \triangleq \{\varpi_t\}_{t \in [0,T]}$, $\varpi_t \in \mathbb{R}$, — скалярный $\mathcal{S}_t$-предсказуемый случайный процесс, определяющий объем средств, вложенных в банковский депозит;
- $\pi \triangleq \{\pi_t\}_{t \in [0,T]}$, $\pi_t \in \mathbb{R}^N$, — N-мерный $\mathcal{S}_t$-предсказуемый случайный процесс $\pi_t = \operatorname{row}(\pi_t^1, \ldots, \pi_t^N)$, определяющий число рисковых ценных бумаг каждого наименования, входящих в данный момент времени в портфель.

При этом процесс $C_t \triangleq \varpi_t B_t + \pi_t S_t$ называется *капиталом* портфеля (ϖ, π).

Портфель C называется *самофинансируемым*, если равенство

$$C_t = \varpi_0 B_0 + \pi_0 S_0 + \int_0^t (\varpi_u dB_u + \pi_u dS_u)$$

выполняется для любого $t \in [0, T]$.

Рынок называется *полным*, если для любого платежного требования — $\mathcal{F}_T$-измеримой неотрицательной квадратично интегрируемой случайной величины q_T — существует такой самофинансируемый портфель (ϖ, π), что

$$C_T = \varpi_0 B_0 + \pi_0 S_0 + \int_0^T (\varpi_u\, dB_u + \pi_u\, dS_u) = \\ = q_T \ \ \mathcal{P}\text{-п. н.}$$

Полнота рынка означает, что любое платежное средство на нем может быть воспроизведено с помощью надлежащего выбора уже имеющихся финансовых инструментов. Следует также подчеркнуть, что данное определение полноты рынка предполагает возможность воспроизведения не только платежного требования «европейского типа», т. е. $q_T(\omega) = q_T(S_T(\omega))$, но и «американского», «азиатского» и прочих типов [12]:

$$q_T(\omega) = q_T(S_{[0,T]}(\omega)).$$

Известно, что модель (1), (2) описывает неполный рынок [8]. *Задача пополнения рынка* заключается в определении на исследуемом стохастическом базисе дополнительных финансовых инструментов таким образом, чтобы новая модель описывала полный рынок, а расширенный набор финансовых инструментов позволял воспроизводить любое платежное требование q_T.

4 Выбор множества пополняющих деривативов

Прежде всего отметим, что в случае $a^\ell(t) \equiv a(t)$ и $\sigma^\ell(t) \equiv \sigma(t)$ для любых $t \in [0, T]$, $\ell = \overline{1, L}$ модель (1), (2) превращается в классическую модель Блэка–Шоулза [11] и рынок становится полным. В этом частном случае и мгновенная процентная ставка, и волатильность рисковых активов становятся неслучайными и известными трейдерам. В общем же случае модели параметры $a^\ell(\cdot)$ и $\sigma^\ell(\cdot)$ для них скрыты. Идея пополнения рынка заключается в одновременном получении информации об МСП Z и построении дополнительных финансовых инструментов, позволяющих эту информацию использовать: по сути дела, торговать ею.

Один из вариантов пополнения рынка (1), (2) представлен в работах [7, 8]. Пусть J_t^ℓ — процесс, считающий число скачков МСП Z в состояние e_ℓ, произошедших на отрезке времени $[0, t]$. Данный процесс допускает следующее мартингальное разложение:

$$J_t^\ell = \int_0^t \sum_{i:\, i \neq \ell} \Lambda_{i\ell}(u) Z_u^i \, du + \int_0^t (1 - Z_{u-}^\ell)\, dM_u^\ell.$$

Для пополнения рынка предлагалось ввести дополнительно L (по числу возможных режимов рынка) скачкообразных финансовых инструментов, связанных с процессами J_t^ℓ. Их цены описывались векторным процессом $G_t \triangleq \operatorname{col}(G_t^1, \ldots, G_t^L)$, компоненты которого имеют вид

$$G_t^\ell = \exp\left(\int_0^t r_u\, du\right) \int_0^t (1 - Z_{u-}^\ell)\, dM_u^\ell, \ \ \ell = \overline{1, L}.$$

В [8] доказано утверждение, согласно которому для каждого платежного требования q_T существует самофинансируемый $\mathcal{F}_t$-предсказуемый портфель (ϖ, π, Π) $(\Pi_t \triangleq \operatorname{row}(\Pi_t^1, \ldots, \Pi_t^L))$, воспроизводящий q_T:

$$q_T = \varpi_0 B_0 + \pi_0 S_0 + \Pi_0 G_0 + \\ + \int_0^T (\varpi_u\, dB_u + \pi_u\, dS_u + \Pi_u\, dG_u) \ \ \mathcal{P}\text{-п. н.}$$

Подобное пополнение рынка представляется искусственным. Во-первых, введение на рынок скачкообразных инструментов предполагает доступность МСП Z точному наблюдению для всех трейдеров. Во-вторых, из определения новых инструментов следует, что цены G_t^ℓ могут с положительной вероятностью принимать отрицательные значения.

В данной работе предлагается провести пополнение рынка более «естественными» финансовыми инструментами, представляющими собой деривативы от уже имеющихся на рынке. Прежде всего заметим, что рынок (1), (2) — безарбитражный [5]. Это, в свою очередь, обеспечивает существование на измеримом пространстве $(\Omega, \mathcal{F})$ *мартингальной* вероятностной меры $\mathcal{Q}$ ($\mathcal{Q} \sim \mathcal{P}$), относительно которой процесс цен S_t описывается решением СДС

$$dS_t = r_t S_t\, dt + \operatorname{diag} S_t \sigma(t, Z_t)\, dw_t^{\mathcal{Q}}, \\ t \in (0, T], \ S_0 \sim p_0(s), \quad (3)$$

в которой $w_t^{\mathcal{Q}} \in \mathbb{R}^N$ — $\mathcal{F}_t$-согласованный винеровский процесс относительно $\mathcal{Q}$.

В силу неполноты рынка мартингальная мера $\mathcal{Q}$ не является единственной. Будем считать, что $\mathcal{Q}$ — одна из мартингальных мер, *преобладающая*, для которой помимо представления S_t в форме (3) выполняется следующее условие.

7. Процесс M_t является мартингалом относительно меры $\mathcal{Q}$.

Цена q_t платежного требования q_T в момент времени $t \in [0,T]$ служит дисконтированным условным средним

$$q_t = e^{-\int_t^T r_s\,ds}\mathsf{E}_{\mathcal{Q}}\left\{q_T|\mathcal{F}_t\right\}.$$

Процесс $\mu_t \triangleq \mathsf{E}_{\mathcal{Q}}\left\{q_T|\mathcal{F}_t\right\}$, будучи $\mathcal{F}_t$-согласованным $\mathcal{Q}$-мартингалом, допускает разложение [13]:

$$\mu_t = \mu_0 + \int\limits_0^t \xi_u\,dw_u^{\mathcal{Q}} + \int\limits_0^t \Xi_u\,dM_u, \tag{4}$$

в котором $\xi_t \triangleq \operatorname{row}(\xi_t^1,\ldots,\xi_t^N)$ и $\Xi_t \triangleq \operatorname{row}(\Xi_t^1,\ldots,\Xi_t^L)$, $t\in[0,T]$, — некоторые $\mathcal{F}_t$-предсказуемые интегранды.

Пусть $H(S_T)$ — некоторое платежное требование, определяющее дериватив «европейского типа», построенный на имеющихся базовых активах. Его цена F_t в момент времени $t\in[0,T]$ определяется формулой

$$F_t = e^{-\int_t^T r_s\,ds}\mathsf{E}_{\mathcal{Q}}\left\{H(S_T)|\mathcal{F}_t\right\}.$$

В предположении, что существует детерминированная функция цены $F(t,s,z):[0,T]\times\mathbb{R}^N\times\mathbb{S}^L$, такая что $\overline{F}_t = F(t,S_t(\omega),Z_t(\omega))$ $\mathcal{P}$-п. н., в [1] показано, что цена деривативa выражается следующим образом:

$$\overline{F}_t = \sum_{\ell=1}^L Z_t^\ell(\omega)F^\ell(t,S_t(\omega)).$$

Здесь вектор-функция $F(t,s) \triangleq \operatorname{col}(F^1(t,s),\ldots,F^L(t,s))$ представляет собой решение следующей системы дифференциальных уравнений в частных производных — обобщения классического уравнения Блэка–Шоулза (зависимость функций от своих аргументов опущена):

$$\left.\begin{aligned} F_t^\ell = rF^\ell - \sum_{j=1}^L \Lambda_{\ell j}F^j - \sum_{n=1}^N F^\ell_{s^n}s^n(r-a_n^\ell) - \\ -\frac{1}{2}\sum_{i,j=1}^N F^\ell_{s^i,s^j}s^is^jb_{ij}^\ell,\quad \ell=\overline{1,L},\quad t\in[0,T]; \\ F^\ell(T,s) = H(s). \end{aligned}\right\} \tag{5}$$

В данной системе использовано следующее обозначение:

$$\sigma^\ell(t)\left(\sigma^\ell(t)\right)^\top = b^\ell(t) = \|b_{ij}^\ell(t)\|_{i,j=\overline{1,L}}.$$

При этом процесс F_t допускает следующий стохастический дифференциал относительно мартингальной меры $\mathcal{Q}$:

$$dF_t = r_tF_t\,dt + \sum_{\ell=1}^L F^\ell(t,S_t)\,dM_t^\ell + \\ + \sum_{\ell=1}^L Z_t^\ell\nabla_sF^\ell(t,S_t)\operatorname{diag}(S_t)\,\sigma^\ell(t)\,dw_t^{\mathcal{Q}}, \tag{6}$$

где $\nabla_sF^\ell \triangleq \operatorname{row}\left(F^\ell_{s^1},\ldots,F^\ell_{s^N}\right)$.

Для пополнения рынка введем дополнительно L деривативов, соответствующих следующему платежному требованию: $H(S_T) \triangleq \operatorname{col}(H^1(S_T),\ldots,H^L(S_T))$ (T — общая дата исполнения всех требований). Объединим цены в векторный процесс $\overline{F}_t \triangleq \operatorname{col}(\overline{F}_t^1,\ldots,\overline{F}_t^L)$. Относительно $H(\cdot)$ сделано следующее предположение.

8. Матрица

$$\mathbf{F}(t,s) \triangleq \begin{bmatrix} F^{11}(t,s) & F^{12}(t,s) & \cdots & F^{1L}(t,s) \\ F^{21}(t,s) & F^{22}(t,s) & \cdots & F^{2L}(t,s) \\ \vdots & \vdots & \vdots & \vdots \\ F^{L1}(t,s) & F^{L2}(t,s) & \cdots & F^{LL}(t,s) \end{bmatrix}$$

невырождена почти везде на множестве $[0,T]\times\mathbb{R}_+^L$.

Заметим, что k-я строка матрицы $\mathbf{F}(t,s)$ $(F^{k1}(t,s),F^{k2}(t,s),\ldots,F^{kL}(t,s))$ представляет собой решение системы (5) с терминальным условием $F^{k\ell}(T,s)\equiv H^k(s)$ ($\ell=\overline{1,L}$) и эта строка полностью задает функцию цены k-го деривативa.

Легко видеть, что предположение 8 гарантирует, что МСП Z_t согласован с фильтрацией $\mathcal{S}_t\vee\overline{\mathcal{F}}_t$ (здесь $\overline{\mathcal{F}}_t \triangleq \sigma\{\overline{F}_u:\ 0\leqslant u\leqslant t\}$):

$$Z_t = \mathbf{F}^{-1}(t,S_t)\overline{F}_t \tag{7}$$

и, следовательно, при выполнении предположений 1–8 для любых $t\in[0,T]$ имеет место совпадение фильтраций $\mathcal{F}_t\equiv\mathcal{S}_t\vee\overline{\mathcal{F}}_t$. Поэтому стохастический дифференциал $\mathcal{Q}$-мартингала $w^{\mathcal{Q}}$ принимает вид:

$$dw_t^{\mathcal{Q}} = \underbrace{\sum_{\ell=1}^L\left(\mathbf{F}^{-1}(t,S_t)\overline{F}_t\right)^\ell(\sigma^\ell)^{-1}(t)\operatorname{diag}^{-1}(S_t)}_{\triangleq\gamma_t}\times \\ \times\left(dS_t - r_tS_t\,dt\right). \tag{8}$$

Из (6) и (8) следует, что

$$d\overline{F}^k_t = r_t\overline{F}^k_t dt + \sum_{\ell=1}^{L} F^{k\ell}(t,S_t)\,dM^\ell_t + \\ + \underbrace{\sum_{\ell=1}^{L}\left(\mathbf{F}^{-1}(t,S_t)\overline{F}_t\right)^\ell \nabla_s F^{k\ell}(u,S_u)}_{\triangleq\Gamma^k_t}\,(dS_t - r_tS_t\,dt)\,, \\ k=\overline{1,L},$$

и в векторном виде эволюция цен деривативов описывается с помощью следующей СДС (здесь $\Gamma_t = \mathrm{col}\,(\Gamma^1_t,\ldots,\Gamma^L_t)$):

$$d\overline{F}_t = r_t\overline{F}_t\,dt + \mathbf{F}_t\,dM_t + \Gamma_t\,(dS_t - r_tS_t\,dt)\,, \\ \overline{F}_0 \text{ — начальное условие.}$$

Таким образом, предположение 8 позволяет выразить мартингал M_t, участвующий в представлении МСП Z_t:

$$dM_t = \mathbf{F}^{-1}_t\left[d\overline{F}_t - r_t\overline{F}_t\,dt - \Gamma_t\,(dS_t - r_tS_t\,dt)\right]. \quad (9)$$

Вернемся к мартингалу μ_t (4), представляющему произвольное платежное требование, и сконструируем для него воспроизводящий портфель (π_t,Π_t,ϖ_t), обладающий свойством самофинансирования. Векторный процесс $\pi_t \triangleq \mathrm{row}\,(\pi^1_t,\ldots,\pi^N_t)$ описывает часть включенных в портфель базовых активов, $\Pi_t \triangleq \mathrm{row}\,(\Pi^1_t,\ldots,\Pi^L_t)$ выполняет ту же функцию для предложенных деривативов, а скалярный процесс ϖ_t определяет долю портфеля, находящуюся на банковском депозите. Выбор доли депозита в портфеле

$$\varpi_t = \mu_t - B^{-1}_t(\pi_tS_t + \Pi_t\overline{F}_t) \quad (10)$$

обеспечивает воспроизведение μ_t. Действительно, если C_t — текущий капитал портфеля, то

$$C_t = \varpi_tB_t + \pi_tS_t + \Pi_t\overline{F}_t = \\ = \left(\mu_t - B^{-1}_t(\pi_tS_t + \Pi_t\overline{F}_t)\right)B_t + \pi_tS_t + \Pi_t\overline{F}_t = \mu_tB_t.$$

Рассмотрим прибыль портфеля и преобразуем его, используя формулы (8)–(10) и интегрирование по частям:

$$\Delta_t \triangleq \int_0^t \varpi_ur_uB_u\,du + \int_0^t\left(\pi_u dS_u + \Pi_u\,d\overline{F}_u\right) = \\ = \int_0^t\left(\mu_u - B^{-1}_u(\pi_uS_u + \Pi_u\overline{F}_u)\right)r_uB_u\,du + \\ + \int_0^t\left(\pi_u\,dS_u + \Pi_u\,d\overline{F}_u\right) = \int_0^t\mu_u\,dB_u - \\ - \int_0^t\left(\pi_uS_u + \Pi_u\overline{F}_u\right)r_u\,du + \int_0^t\left(\pi_u\,dS_u + \Pi_u\,d\overline{F}_u\right) = \\ = \mu_tB_t - \mu_0 - \int_0^t B_u\,d\mu_u - \int_0^t\left(\pi_uS_u + \Pi_u\overline{F}_u\right)r_u\,du + \\ + \int_0^t\left(\pi_u\,dS_u + \Pi_u\,d\overline{F}_u\right) = \mu_tB_t - \mu_0 - \\ - \int_0^t\left(\pi_uS_u + \Pi_u\overline{F}_u\right)r_u\,du + \int_0^t\left(\pi_u\,dS_u + \Pi_u\,d\overline{F}_u\right) - \\ - \int_0^t B_u\left[\xi_u\gamma_u(dS_u - r_uS_u\,du) + \right. \\ \left. + \Xi_u\mathbf{F}^{-1}_u\left(d\overline{F}_u - r_u\overline{F}_u\,du - \Gamma_u(dS_u - r_uS_u\,du)\right)\right] = \\ = \mu_tB_t - \mu_0 + \int_0^t I^1_u\,du + \int_0^t I^2_u\,dS_u + \int_0^t I^3_u\,d\overline{F}_u,$$

где

$$I^1_t \triangleq r_t\left\{\left[B_t(\xi_t\gamma_t - \Xi_t\mathbf{F}^{-1}_t\Gamma_t) - \pi_t\right]S_t + \right. \\ \left. + \left[B_t\Xi_t\mathbf{F}^{-1}_t - \Pi_t\right]\overline{F}_t\right\};$$
$$I^2_t \triangleq \pi_t - B_t(\xi_t\gamma_t - \Xi_t\mathbf{F}^{-1}_t\Gamma_t);$$
$$I^3_t \triangleq \Pi_t - B_t\Xi_t\mathbf{F}^{-1}_t.$$

Легко проверить, что выбор долей

$$\pi_t = B_t(\xi_t\gamma_t - \Xi_t\mathbf{F}^{-1}_t\Gamma_t),\ \ \Pi_t = B_t\Xi_t\mathbf{F}^{-1}_t$$

гарантирует портфелю свойство самофинансирования:

$$\Delta_t = \mu_tB_t - \mu_0.$$

Таким образом, доказана

Теорема 1. *Если предположения* 1–8 *верны, то рынок* (1), (2) *может быть пополнен набором* L *производных финансовых инструментов*.

5 Заключение

Представленный способ пополнения рынка позволяет сделать следующие выводы.

1. Для пополнения использовались деривативы европейского типа, однако портфели, построенные на них и имеющихся базовых активах, позволяют воспроизводить платежные требования любого типа: американского, азиатского, бермудского и пр.

2. Предложенный в доказательстве портфель воспроизводит не только платежное требование в момент погашения T, но и всю траекторию цены этого требования на отрезке $[0, T]$.
3. Предложенные для пополнения инструменты выглядят более естественными, чем предложенные в [7, 8].

Вообще говоря, и в цитируемых работах, и в данной заметке используется единый подход. Изначально рынок содержит скрытый случайный процесс, недоступный наблюдению для участников рынка. Он порождает дополнительный риск, обладающий некоторой ценой. Далее на основе этого процесса строятся некоторые инструменты с наблюдаемыми ценами, которыми и предлагается пополнить рынок. В [7, 8] в качестве этих активов выступают мартингалы процессов, считающих скачки МСП в то или иное состояние.

Подобный выбор представляется не вполне естественным. Цены могут принимать отрицательные величины, МСП является скрытым, и с практической точки зрения не вполне понятен источник наблюдений этого процесса. Предложенный в данной работе способ пополнения выглядит более привлекательно: в качестве пополняющих используются «естественные» деривативы европейского типа, построенные на имеющихся на рынке базовых активах.

Результаты, представленные в данной заметке, имеют, на первый взгляд, лишь академический интерес. Действительно, при выполнении предположения 8 скрытый режим рынка Z_t может быть восстановлен точно с помощью элементарного алгебраического преобразования (7) без привлечения методов оптимальной нелинейной фильтрации [14]. Это означает, что наличие точно наблюдаемых в непрерывном времени цен деривативов исключает статистическую неопределенность процентной ставки и волатильности. В [3] приведена аргументация против наличия наблюдений цен базовых активов и деривативов в непрерывном времени, а также возможности наблюдать точные цены деривативов и в дискретные моменты времени. Эти факты подталкивают к разработке высокоточных методов оценивания состояний скрытого МСП по разнородным косвенным зашумленным наблюдениям. Подобные алгоритмы были предложены в [3, 4] для двух разных видов статистической информации.

Перспективным продолжением исследований в данной области представляется решение задачи хеджирования на рынках подобного типа с использованием концепции фильтрованного пространства.

Другое направление подразумевает проведение аналогичных исследований для моделей Халла–Уайта и Васичека с марковскими переключениями.

Литература

1. *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.
2. *Борисов А.* Рынок с марковской скачкообразной волатильностью II: алгоритм вычисления справедливой цены деривативов // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 18–24. doi: 10.14357/19922264230303. EDN: DNXJGB.
3. *Борисов А.* Рынок с марковской скачкообразной волатильностью III: алгоритм мониторинга цены риска по дискретным наблюдениям цен // Информатика и её применения, 2023. Т. 17. Вып. 4. С. 9–16. doi: 10.14357/19922264230402. EDN: OFYELT.
4. *Борисов А.* Рынок с марковской скачкообразной волатильностью IV: алгоритм мониторинга рыночной цены риска по потоку высокочастотных наблюдений базовых активов и деривативов // Информатика и её применения, 2024. Т. 18. Вып. 1. С. 26–32. doi: 10.14357/19922264240104. EDN: ZRQKIT.
5. *Criens D.* No arbitrage in continuous financial markets // Math. Financ. Econ., 2020. Vol. 14. P. 461–506. doi: 10.1007/s11579-020-00262-1.
6. *Fouque J., Papanicolaou G., Sircar K.* Derivatives in financial markets with stochastic volatility. — Cambrige, U.K.: Cambrige University Press, 2000. 218 p.
7. *Courcuera J., Nualart D., Schoutens W.* Completion of a Lévy market by power-jump assets // Financ. Stoch., 2005. Vol. 9. Iss. 1. P. 109–127. doi: 10.1007/s00780-004-0139-2.
8. *Zhang X., Elliott R., Siu T., Guo J.* Markovian regime-switching market completion using additional Markov jump assets // IMA J. Manag. Math., 2011. Vol. 23. Iss. 3. P. 283–305. doi: 10.1093/imaman/dpr018.
9. *Elliott R., Aggoun L., Moore J.* Hidden Markov models: Estimation and control. — New York, NY, USA: Springer, 2010. 382 p.
10. *Liptser R., Shiryayev A.* Theory of martingales; mathematics and its applications. — Amsterdam, The Netherlands: Springer, 2012. 806 p.
11. *Shiryayev A.* Essentials of stochastic finance: Facts, models, theory. — New Jersey, NJ, USA: World Scientific, 1999. 834 p.
12. *Wilmott P., Howison S., Dewynne J.* The mathematics of financial derivatives: A student introduction. – Cam-

brige, U.K.: Cambrige University Press, 1995. 317 p. doi: 10.1017/CBO9780511812545.

13. *Elliott R.* Double martingales // Z. Wahrscheinlichkeit., 1976. Vol. 34. P. 17–28. doi: 10.1007/BF00532686.

14. *Борисов А., Казанчян Д.* Фильтрация состояний марковских скачкообразных процессов по комплексным наблюдениям I: точное решение задачи // Информатика и её применения, 2021. Т. 15. Вып. 2. С. 12–19. doi: 10.14357/19922264210202. EDN: NKCTNS.

Поступила в редакцию 05.02.24

MARKET WITH MARKOV JUMP VOLATILITY V: MARKET COMPLETION WITH DERIVATIVES

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 final, fifth, part of the series is devoted to a replenishment procedure of the market with a Markov jump regime change. The market includes a riskless bank deposit with a known nonrandom interest rate and a set of underlying risky assets. The instantaneous interest rates and volatilities of the assets are the functions of the hidden regime change factor described by some finite state Markov jump process. The purpose of this article is to complete the investigated market. It means the market enlargement by a set of auxiliary financial instruments. The point is that any contingent claim declared in the market can be replicated with some self-financing portfolio containing the original and auxiliary instruments. For the market completion, it is enough to include European-style derivatives built on underlying risky assets already on the market. In this case, the number of added derivatives coincides with the number of market modes. The problem of replenishing the market has a nonunique solution and the article compares the proposed replenishment method with the existing one.

Keywords: Markov jump process; financial portfolio; self-financing property; market completeness

DOI: 10.14357/19922264240202 **EDN:** DQFSDO

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. 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.
2. Borisov, A. 2023. Rynok s'markovskoy skachkoobraznoy volatil'nost'yu II: algoritm vychisleniya spravedlivoy tseny derivativov [Market with Markov jump volatility II: Algorithm of derivative fair price calculation]. *Informatika i ee primeneniya — Inform. Appl.* 17(3):18–24. doi: 10.14357/19922264230303. EDN: DNXJGB.
3. Borisov, A. 2023. Rynok s markovskoy skachkoobraznoy volatil'nost'yu III: algoritm monitoringa tseny riska po diskretnym nablyudeniyam tsen aktivov [Market with Markov jump volatility III: Price of risk monitoring algorithm given discrete-time observations of asset prices]. *Informatika i ee primeneniya — Inform. Appl.* 17(4):9–16. doi: 10.14357/19922264230402. EDN: OFYELT.
4. Borisov, A. 2024. Rynok s markovskoy skachkoobraznoy volatil'nost'yu IV: algoritm monitoringa rynochnoy tseny riska po potoku vysokochastotnykh nablyudeniy bazovykh aktivov i derivativov [Market with Markov jump volatility IV: Price of risk monitoring algorithm given high frequency observation flows of assets prices]. *Informatika i ee primeneniya — Inform. Appl.* 18(1):26–32. doi: 10.14357/19922264240104. EDN: ZRQKIT.
5. Criens, D. 2020. No arbitrage in continuous financial markets. *Math. Financ. Econ.* 14:461–506. doi: 10.1007/s11579-020-00262-1.
6. Fouque, J., G. Papanicolaou, and K. Sircar. 2000. *Derivatives in financial markets with stochastic volatility*. Cambrige, U.K.: Cambrige University Press. 218 p.
7. Courcuera, J., D. Nualart, and W. Schoutens. 2005. Completion of a Lévy market by power-jump assets. *Financ. Stoch.* 9(1):109–127. doi: 10.1007/s00780-004-0139-2.
8. Zhang X., R. Elliott, T. Siu, and J. Guo. 2011. Markovian regime-switching market completion using additional Markov jump assets. *IMA J. Manag. Math.* 23(3):283–305. doi: 10.1093/imaman/dpr018.

9. Elliott, R., L. Aggoun, and J. Moore. 2010. *Hidden Markov models: Estimation and control*. New York, NY: Springer. 382 p.

10. Liptser, R., and A. Shiryayev. 2012. *Theory of martingales. Mathematics and its applications*. Amsterdam, The Netherlands: Springer. 806 p.

11. Shiryayev, A. 1999. *Essentials of stochastic finance: Facts, models, theory*. New Jersey, NJ: World Scientific. 834 p.

12. Wilmott, P., S. Howison, and J. Dewynne. 1995. *The mathematics of financial derivatives: A student introduction*. Cambrige, U.K.: Cambrige University Press. 317 p. doi: 10.1017/CBO9780511812545.

13. Elliott, R. 1976. Double martingales. *Z. Wahrscheinlichkeit.* 34:17–28. doi: 10.1007/BF00532686.

14. Borisov, A., and D. Kazanchyan. 2021. Fil'tratsiya sostoyaniy markovskikh skachkoobraznykh protsessov po kompleksnym nablyudeniyam I: tochnoe reshenie zadachi [Filtering of Markov jump processes given composite observations I: Exact solution]. *Informatika i ee primeneniya — Inform. Appl.* 15(2):12–19. doi: 10.14357/19922264210202. EDN: NKCTNS.

Received February 5, 2024

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@frccsc.ru

НИЖНЯЯ ГРАНИЦА ПОГРЕШНОСТИ ОЦЕНИВАНИЯ СЛУЧАЙНОГО ПАРАМЕТРА ПРИ ЗАДАННОМ КОЛИЧЕСТВЕ ИНФОРМАЦИИ

М. М. Ланге[1], А. М. Ланге[2]

Аннотация: Исследуется наименьшая средняя квадратичная погрешность оценивания случайного параметра плотности распределения по выборкам независимых наблюдений как функция средней взаимной информации в выборках относительно множества значений оценок. Рассматриваемая функция строится в форме обращения зависимости наименьшей средней взаимной информации от средней погрешности, которая представляет собой модификацию известной в теории информации функции скорость–погрешность (rate distortion function). Полученная зависимость наименьшей средней погрешности от количества используемой информации не зависит от вида оценки и дает нижнюю границу средней погрешности при фиксированных значениях количества информации. Такая зависимость определяет двухфакторный критерий качества решения, который позволяет сравнивать эффективность различных способов построения оценок в терминах избыточности их средней погрешности относительно нижней границы при заданной энтропии квантованных значений оценок.

Ключевые слова: плотность распределения; выборка наблюдений; оценка параметра; квадратичная погрешность; взаимная информация; функция скорость–погрешность; нижняя граница; избыточность

DOI: 10.14357/19922264240203 **EDN:** EFZGYW

1 Введение

Восстановление случайного параметра функции распределения вероятностей остается важной проблемой в задачах интеллектуальной обработки данных и, в частности, в задачах машинного обучения [1] и универсального кодирования источников [2]. Восстановление сводится к построению оценки параметра, качество которой должно удовлетворять допустимой погрешности. Как правило, оценка параметра, принимающего значения на непрерывном множестве, строится по выборке наблюдений, а качество оценки определяется средним значением квадратичной погрешности, которое должно уменьшаться с ростом размера выборки. Примеры таких оценок для двухпараметрических распределений Райса и Парето рассмотрены соответственно в работах [3, 4]. Оценки параметров гамма-экспоненциальных распределений, построенные на основе модификаций метода моментов, рассмотрены в работах [5, 6].

Известный метод оптимизации оценки параметра распределения базируется на минимизации средней квадратичной погрешности при фиксированном размере выборки наблюдений [7]. В этом случае наименьшая средняя погрешность может быть найдена либо путем непосредственного вычисления дисперсии апостериорной плотности оцениваемого параметра, либо с привлечением неравенства Рао–Крамера [8]. Зависимость указанной дисперсии от размера выборки позволяет варьировать величину наименьшей средней погрешности путем изменения размера выборки как эвристической меры количества информации, но не позволяет найти наименьшую среднюю погрешность оценки при заданной величине средней взаимной информации [9] между множеством выборок фиксированного размера и множеством значений оценок. Однако для вероятностной модели оценивания случайного параметра именно такая теоретико-информационная мера может быть полезна для построения нижней границы средней погрешности как функции среднего количества используемой информации. При этом наименьшее значение погрешности должно совпадать с границей Рао–Крамера, которая соответствует наибольшей средней взаимной информации, а наибольшее значение погрешности определяется дисперсией априорной плотности распределения оцениваемого параметра и соответствует нулевому значению взаимной информации.

В теории кодирования непрерывных сообщений с заданным критерием качества аналогичная зависимость наименьшей скорости кодирования от заданной средней погрешности определяется функцией скорость–погрешность [9].

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, lange_mm@mail.ru
[2]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, lange_am@mail.ru

Нижняя граница средней погрешности как функция средней взаимной информации позволяет оценить избыточность средней погрешности относительно нижней границы при различных методах построения оценок параметра с использованием моментов различных порядков, функций правдоподобия и др. [8], а также с использованием различных методов квантования [10]. Для заданного размера выборок минимум средней взаимной информации между выборками и значениями оценок параметра при ограничении средней погрешности может быть найден с использованием техники вычисления функции скорость–погрешность в схеме кодирования непрерывных сообщений, переданных по каналу с шумом [11]. Обращение полученной в результате предлагаемого подхода монотонной зависимости наименьшей средней взаимной информации от средней погрешности позволяет получить наименьшую среднюю погрешность как функцию средней взаимной информации. Такая функция не зависит от вида оценки и дает теоретико-информационную нижнюю границу средней погрешности оценивания параметра как функцию количества используемой информации.

В настоящей работе указанная теоретико-информационная граница найдена для средней квадратичной погрешности оценки параметра плотности распределения. По форме найденная граница аналогична границам вероятности ошибки, полученным ранее для дискретных моделей кодирования и классификации данных [12].

2 Формализация задачи

Пусть $p_{X|\Theta}(x|\theta)$, $x \in X$, $\theta \in \Theta$, — условная плотность распределения случайной величины на множестве значений X с неизвестным случайным параметром с априорной плотностью распределения $p_\Theta(\theta)$ на множестве значений Θ. Будем считать, что оценки $\hat\theta_n$ значений θ строятся по выборкам $x^n = (x_1, \ldots, x_n)$, содержащим n независимых наблюдений. Погрешность оценок измеряется в квадратичной мере $(\hat\theta_n - \theta)^2$, а множество значений оценок образует множество $\hat\Theta_n$. Предполагается, что множества Θ, X^n и $\hat\Theta_n$ обладают свойством марковости, при котором элементы каждого множества зависят только от элементов предыдущего множества.

В принятых обозначениях множество выборок X^n с условной по параметру плотностью $p_{X^n|\Theta}(x^n|\theta) = \prod_{k=1}^n p_{X|\Theta}(x_k|\theta)$, $x_k \in X$, и множество значений оценок $\hat\Theta_n$ с некоторой условной по выборке плотностью $q_{\hat\Theta_n|X^n}(\hat\theta_n|x^n)$ позволяют ввести среднюю погрешность [11]

$$E_{q_{\hat\Theta_n|X^n}}\left(X^n;\hat\Theta_n\right) = \int\limits_{X^n} p_{X^n}(x^n) \int\limits_{\hat\Theta_n} q_{\hat\Theta_n|X^n}\left(\hat\theta_n|x^n\right) \times$$
$$\times \int\limits_{\Theta} p_{\Theta|X^n}(\theta|x^n)\left(\theta-\hat\theta_n\right)^2 d\theta d\hat\theta_n dx^n \quad (1)$$

и среднюю взаимную информацию [9]

$$I_{q_{\hat\Theta_n|X^n}}\left(X^n;\hat\Theta_n\right) = \int\limits_{X^n} p_{X^n}(x^n) \times$$
$$\times \int\limits_{\hat\Theta_n} q_{\hat\Theta_n|X^n}\left(\hat\theta_n|x^n\right) \ln \frac{q_{\hat\Theta_n|X^n}\left(\hat\theta_n|x^n\right)}{q_{\hat\Theta_n}\left(\hat\theta_n\right)}\, d\hat\theta_n dx^n . \quad (2)$$

Здесь $p_{X^n}(x^n)$ и $q_{\hat\Theta_n}(\hat\theta_n)$ — безусловные плотности распределений на множествах X^n и $\hat\Theta_n$:

$$p_{X^n}(x^n) = \int\limits_{\Theta} p_\Theta(\theta) p_{X^n|\Theta}\left(x^n|\theta\right)\, d\theta\,;$$
$$q_{\hat\Theta_n}\left(\hat\theta_n\right) = \int\limits_{X^n} p_{X^n}\left(x^n\right) q_{\hat\Theta_n|X^n}\left(\hat\theta_n|x^n\right)\, dx^n\,;$$

$p_{\Theta|X^n}(\theta|x^n)$ — апостериорная плотность на множестве Θ:

$$p_{\Theta|X^n}(\theta|x^n) = \frac{p_\Theta(\theta) p_{X^n|\Theta}(x^n|\theta)}{p_{X^n}(x^n)}\,.$$

Функционалы (1) и (2) не зависят от функции вычисления оценки $\hat\theta_n$ по выборке x^n, но зависят от условной плотности распределения $q_{\hat\Theta_n|X^n}(\hat\theta_n|x^n)$. Указанные функционалы служат теоретико-информационными мерами средней погрешности и средней информации, которые используются в вероятностной модели, заданной парой стохастических преобразований

$$\Theta \xrightarrow{p_{X^n|\Theta}(x^n|\theta)} X^n \xrightarrow{q_{\hat\theta_n|x^n}(\hat\theta_n|x_n)} \hat\Theta_n.$$

Рассматриваемая вероятностная модель позволяет минимизировать среднюю взаимную информацию $I_{q_{\hat\Theta_n|X^n}}(X^n;\hat\Theta_n)$ по плотности $q_{\hat\Theta_n|X^n}(\hat\theta_n|x^n)$ при ограничении средней погрешности $E_{q_{\hat\Theta_n|X^n}}(X^n;\hat\Theta_n) \le \varepsilon$ допустимым значением $\varepsilon > 0$. При фиксированном размере выборки $n \ge 1$ и различных значениях ε такой условный минимум дает функцию

$$R_n(\varepsilon) =$$
$$= \min_{q_{\hat\Theta_n|X^n}:E_{q_{\hat\Theta_n|X^n}}(X^n;\hat\Theta_n)\le\varepsilon} I_{q_{\hat\Theta_n|X^n}}(X^n;\hat\Theta_n), \quad (3)$$

которая аналогична функции скорость–погрешность для модели кодирования независимых непрерывных сообщений, переданных по каналу с аддитивным гауссовым шумом [11]. Задача состоит в построении монотонно убывающей с ростом ε нижней границы $\underline{R}_n(\varepsilon) \le R_n(\varepsilon)$. Тогда обратная функция $\underline{R}_n^{-1}(I)$ дает нижнюю границу средней погрешности при значении средней взаимной информации $I_{q_{\hat{\Theta}_n|X^n}}(X^n;\hat{\Theta}_n) = I$.

Последующее изложение включает построение нижней границы функции $R_n(\varepsilon)$ (разд. 3) и пример вычисления найденной границы в случае оценивания среднего значения гауссовой плотности (разд. 4). Заключение содержит краткие выводы и перспективы дальнейших исследований.

3 Нижняя граница функции $R_n(\varepsilon)$

Вычисление функции (3) базируется на преобразованиях функционалов (1) и (2). Рассмотрим среднюю погрешность оценки $\hat{\theta}_n$

$$\int\limits_{\Theta} p_{\Theta|X^n}(\theta|x^n)\left(\theta-\hat{\theta}_n\right)^2 d\theta = \int\limits_{\Theta} p_{\Theta|X^n}(\theta|x^n)\left((\theta-\theta_n)+\left(\theta_n-\hat{\theta}_n\right)\right)^2 d\theta \quad (4)$$

и, требуя

$$\int\limits_{\Theta} p_{\Theta|X^n}(\theta|x^n)\,(\theta-\theta_n)^2\,d\theta \to \min_{\theta_n},$$

получим оптимальную оценку в форме математического ожидания

$$\theta_n(x^n) = \int\limits_{\Theta} p_{\Theta|X^n}(\theta|x^n)\theta\,d\theta\,.$$

С учетом оценки $\theta_n(x^n)$ средняя погрешность (4) преобразуется к виду

$$\int\limits_{\Theta} p_{\Theta|X^n}(\theta|x^n)\left(\theta-\hat{\theta}_n\right)^2 d\theta = \sigma_n^2(x^n) + \left(\theta_n(x^n)-\hat{\theta}_n\right)^2, \quad (5)$$

где $\sigma_n^2(x^n)$ — дисперсия апостериорной плотности $p_{\Theta|X^n}(\theta|x^n)$:

$$\sigma_n^2(x^n) = \int\limits_{\Theta} p_{\Theta|X^n}(\theta|x^n)\theta^2\,d\theta - \left(\int\limits_{\Theta} p_{\Theta|X^n}(\theta|x^n)\theta\,d\theta\right)^2.$$

Пусть оптимальная оценка, построенная на выборках $x^n \in X^n$, принимает значения на множестве Θ_n с плотностью распределения $p_{\Theta_n}(\theta_n)$. Множество X^n может быть представлено набором непересекающихся подмножеств $\{X_t^n\}$, где X_t^n — подмножество выборок x^n, которые дают значения $\theta_n(x^n)$ на отрезке размера Δ. Такое представление множества X^n порождает разбиение множества Θ_n на непересекающиеся кванты Θ_{nt} размера Δ. При $\Delta \to 0$ значения оценок для всех выборок $x^n \in X_t^n$ близки к величине $\theta_{nt} \in \Theta_{nt}$ и справедливы асимптотические равенства:

$$\int\limits_{X_t^n} p_{X^n}(x^n)\,dx^n \approx p_{\Theta_n}(\theta_{nt})\Delta;$$

$$q_{\hat{\Theta}_n|X^n}\left(\hat{\theta}_n|x^n\right) \approx q_{\hat{\Theta}_n|\Theta_n}\left(\hat{\theta}_n|\theta_{nt}\right).$$

Тогда замена интегралов на множестве X^n суммой интегралов по подмножествам X_t^n позволяет представить среднюю погрешность (1) и среднюю взаимную информацию (2) в терминах соответствующих функционалов, заданных на множествах Θ_n и $\hat{\Theta}_n$.

С учетом соотношения (5) средняя погрешность (1) преобразуется к виду

$$E_{q_{\hat{\Theta}_n|X^n}}\left(X^n;\hat{\Theta}_n\right) = \varepsilon_{n\_\min} + \int\limits_{\Theta_n} p_{\Theta_n}(\theta_n)\int\limits_{\hat{\Theta}_n} q_{\hat{\Theta}_n|\Theta_n}\left(\hat{\theta}_n|\theta_n\right)\left(\theta_n-\hat{\theta}_n\right)^2 d\hat{\theta}_n d\theta_n = \varepsilon_{n\_\min} + E_{q_{\hat{\Theta}_n|\Theta_n}}(\Theta_n;\hat{\Theta}_n), \quad (6)$$

где $\varepsilon_{n\_\min}$ — средняя погрешность оптимальных оценок на множестве выборок X^n:

$$\varepsilon_{n\_\min} = \int\limits_{X^n} p_{X^n}(x^n)\sigma_n^2(x^n)\,dx^n.$$

При этом средняя взаимная информация (2) принимает вид:

$$I_{q_{\hat{\Theta}_n|X^n}}\left(X^n;\hat{\Theta}_n\right) = I_{q_{\hat{\Theta}_n|\Theta_n}}\left(\Theta_n;\hat{\Theta}_n\right) = \int\limits_{\Theta_n} p_{\Theta_n}(\theta_n)\times \int\limits_{\hat{\Theta}_n} q_{\hat{\Theta}_n|\Theta_n}\left(\hat{\theta}_n|\theta_n\right)\ln\frac{q_{\hat{\Theta}_n|\Theta_n}\left(\hat{\theta}_n|\theta_n\right)}{q_{\hat{\Theta}_n}\left(\hat{\theta}_n\right)}\,d\hat{\theta}_n d\theta_n\,. \quad (7)$$

Поскольку функционалы $E_{q_{\hat{\Theta}_n|\Theta_n}}(\Theta_n;\hat{\Theta}_n)$ и $I_{q_{\hat{\Theta}_n|\Theta_n}}(\Theta_n;\hat{\Theta}_n)$ зависят от свободной условной плотности распределения $q_{\hat{\Theta}_n|\Theta_n}(\hat{\theta}_n|\theta_n)$, соотношения (6) и (7) позволяют переопределить функцию (3) в форме

$$R_n(\varepsilon) = \min_{q_{\hat{\Theta}_n|\Theta_n}:\, E_{q_{\hat{\Theta}_n|\Theta_n}}(\Theta_n;\hat{\Theta}_n)\le\varepsilon-\varepsilon_{n\_\min}} I_{q_{\hat{\Theta}_n|\Theta_n}}(\Theta_n;\hat{\Theta}_n), \quad (8)$$

где минимум от $I_{q_{\hat{\Theta}|\Theta_n}}(\Theta_n;\hat{\Theta})$ берется по всевозможным плотностям $q_{\hat{\Theta}_n|\Theta_n}(\hat{\theta}_n|\theta_n)$ при условии $E_{q_{\hat{\Theta}_n|\Theta_n}}(\Theta_n;\hat{\Theta}_n) \le \varepsilon - \varepsilon_{n\_\min}$.

Теорема 1. *Для функции, определенной в* (8), *справедлива монотонно убывающая с ростом ε неотрицательная нижняя граница*

$$R_n(\varepsilon) \ge \underline{R}_n(\varepsilon) = h(\Theta_n) - \frac{1}{2}\ln\left(2\pi e(\varepsilon-\varepsilon_{n\_\min})\right), \quad \varepsilon_{n\_\min} < \varepsilon \le \varepsilon_{\max}, \quad (9)$$

где $h(\Theta_n)$ — дифференциальная энтропия на множестве Θ_n:

$$h(\Theta_n) = -\int_{\Theta_n} p_{\Theta_n}(\theta_n)\ln p_{\Theta_n}(\theta_n)\,d\theta_n;$$

$\underline{R}_n(\varepsilon_{n\_\min}) \to \infty$; $\underline{R}_n(\varepsilon_{\max}) = 0$:

Доказательство. Построение границы (9) базируется на вычислении минимума вида (8) с помощью вариационного подхода, предложенного Шенноном и изложенного в монографии [9]. Опуская промежуточные выкладки, получим нижнюю границу для функции $R_n(\varepsilon)$ в форме

$$\underline{R}_n(\varepsilon) = h(\Theta_n) - h_s\left(\hat{\Theta}_n|\Theta_n\right), \quad (10)$$

где $h_s(\hat{\Theta}_n|\Theta_n)$ — условная дифференциальная энтропия на множествах $\hat{\Theta}_n$ и Θ_n:

$$h_s\left(\hat{\Theta}_n|\Theta_n\right) = -\int_{\Theta_n} p_{\Theta_n}(\theta_n) \times \int_{\hat{\Theta}_n} g^{(s)}_{\hat{\Theta}_n|\Theta_n}(\hat{\theta}_n|\theta_n)\ln g^{(s)}_{\hat{\Theta}_n|\Theta_n}(\hat{\theta}_n|\theta_n)\,d\hat{\theta}_n d\theta_n$$

с условной плотностью

$$g^{(s)}_{\hat{\Theta}_n|\Theta_n}\left(\hat{\theta}_n|\theta_n\right) = \frac{\exp\left(-s(\hat{\theta}_n-\theta_n)^2\right)}{\int_{\hat{\Theta}_n}\exp\left(-s(\hat{\theta}_n-\theta_n)^2\right)d\hat{\theta}_n}. \quad (11)$$

Значение параметра $s > 0$ плотности $g^{(s)}_{\hat{\Theta}_n|\Theta_n}(\hat{\theta}_n|\theta_n)$ следует из уравнения

$$\int_{\Theta_n} p_{\Theta_n}(\theta_n)\int_{\hat{\Theta}_n} g^{(s)}_{\hat{\Theta}_n|\Theta_n}\left(\hat{\theta}_n|\theta_n\right)\left(\theta_n-\hat{\theta}_n\right)^2 d\hat{\theta}_n d\theta_n = \varepsilon - \varepsilon_{n\_\min}.$$

Введение переменной $z_n = (\hat{\theta}_n - \theta_n)$, принимающей значения на интервале $(-\infty,\infty)$, преобразует плотность (11) к нормальному виду с нулевым средним и дисперсией $\sigma_s^2 = 1/(2s)$, где

$$s = \frac{1}{2}\left(\varepsilon-\varepsilon_{n\_\min}\right)^{-1} > 0\,.$$

Полученная нормальная плотность позволяет вычислить условную дифференциальную энтропию в форме

$$h_s\left(\hat{\Theta}_n|\Theta_n\right) = \frac{1}{2}\ln\left(2\pi e\sigma_s^2\right) = \frac{1}{2}\ln\left(2\pi e(\varepsilon-\varepsilon_{n\_\min})\right),$$

которая не зависит от значений $\theta_n \in \Theta_n$, имеет в точке $\varepsilon = \varepsilon_{n\_\min}$ разрыв типа $-\infty$ и монотонно возрастает с увеличением ε. Подстановка полученной условной дифференциальной энтропии в (10) дает границу в форме (9). Поскольку наибольшая средняя погрешность реализуется при нулевом значении средней взаимной информации, обеспечивая $\underline{R}_n(\varepsilon_{\max}) = 0$, значение $\varepsilon_{\max}$ не зависит от размера выборки и определяется дисперсией априорной плотности $p_\Theta(\theta)$. Теорема доказана.

Практическая ценность границы $\underline{R}_n(\varepsilon)$ состоит в возможности ее применения для вычисления избыточности средней погрешности квантованных значений оценок при фиксированных значениях количества информации, которая измеряется энтропией множества оценок. Особый интерес представляет избыточность погрешности для оценок, строящихся на достаточных статистиках и не зависящих от параметров априорных распределений, которые, как правило, не известны. Поэтому полезно рассмотреть множество оценок

$$\hat{\Theta}_n = \left\{\hat{\theta}_n(x^n),\ \forall\, x^n \in X^n\right\},$$

которые связаны с оптимальными оценками линейным преобразованием

$$\hat{\theta}_n(x^n) = \alpha_n\theta_n(x^n) + \beta_n \quad (12)$$

с коэффициентами $\alpha_n \ge 1$, $-\infty < \beta_n < \infty$. Предполагается, что с увеличением размера выборки $(n \to \infty)$ коэффициенты стремятся к предельным значениям $\alpha_n \to 1$ и $\beta_n \to 0$, что обеспечивает асимптотическую оптимальность оценок (12).

С учетом соотношения (12) равномерное квантование величин $\theta_n \in \Theta_n$ с шагом Δ соответствует равномерному квантованию величин $\hat{\theta}_n \in \hat{\Theta}_n$ с шагом $\alpha_n\Delta$. Тогда, согласно работе [10], при значениях $\Delta \to 0$ энтропия H_n и средняя погрешность $\underline{E}_n$ квантованных значений оценок из множества $\hat{\Theta}_n$ удовлетворяют соотношениям

$$H_n = h\left(p_{\hat{\Theta}}\right) - \ln(\alpha_n \Delta)\,; \tag{13}$$

$$E_n = \varepsilon_{n\_\min} + \frac{\alpha_n^2 \Delta^2}{12}\,. \tag{14}$$

Полагая $\underline{R}_n(\varepsilon) = H_n$, избыточность средней погрешности E_n относительно нижней границы можно определить величиной $r_n = E_n - \underline{R}_n^{-1}(H_n)$. С учетом соотношения дифференциальных энтропий $h(\hat{\Theta}_n) = h(\Theta_n) + \ln \alpha_n$ на множествах $\hat{\Theta}_n$ и Θ_n энтропия в (13) не зависит от α_n и равна $H_n = h(\Theta_n) - \ln \Delta$. Учитывая последнее замечание, из (13) и (14) получим

$$\frac{r_n}{E_n - \varepsilon_{n\_\min}} = 1 - \frac{6}{\pi e \alpha_n^2}\,. \tag{15}$$

При значении $\alpha_n = 1$ относительная избыточность (15) достигает наименьшего значения 0,297 и увеличивается с ростом α_n. В общем случае для значений $\Delta \to 0$ и $\alpha_n \geq 1$ соотношение (15) демонстрирует меньшую избыточность r_n средней погрешности E_n относительно нижней границы $\underline{R}_n^{-1}(H_n)$ по сравнению с традиционно используемой величиной $E_n - \varepsilon_{n\_\min}$. На практике величины H_n и E_n могут быть вычислены для заданных способов построения оценок и используемых методов их квантования.

4 Пример вычисления границы $\underline{R}_n(\varepsilon)$

Рассмотрим пример вычисления границы (9) для гауссовой модели, когда независимые наблюдения $x \in X$ имеют нормальную плотность распределения $p_{X|\Theta}(x|\theta)$ со случайным средним значением $\theta \in \Theta$ и известной дисперсией σ^2. Предполагается также, что значения θ имеют априорную нормальную плотность распределения $p_\Theta(\theta)$ со средним θ_0 и дисперсией σ_0^2. В этом случае апостериорное по выборке $x^n \in X^n$ распределение имеет нормальную плотность [7]

$$p_{\Theta|X^n}(\theta|x^n) = \frac{p_\Theta(\theta) p_{X^n|\Theta}(x^n|\theta)}{p_{X^n}(x^n)}$$

со средним значением

$$\theta_n = \frac{\sigma_0^2}{n\sigma_0^2 + \sigma^2} \sum_{k=1}^{n} x_k + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}\,\theta_0 \tag{16}$$

и дисперсией

$$\sigma_n^2 = \frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2}\,, \tag{17}$$

которая для рассматриваемой гауссовой модели не зависит от выборки x^n.

Пусть $\overline{X}_n$ — множество значений выборочного среднего $\overline{x}_n = (1/n)\sum_{k=1}^{n} x_k$. Известно, что случайная величина $(\overline{x}_n - \theta)/(\sigma/\sqrt{n})$ имеет нормализованную нормальную плотность распределения [13]. Тогда условная плотность $p_{\overline{X}_n|\Theta}(\overline{x}_n|\theta)$ является нормальной со средним значением θ и дисперсией σ^2/n. Поэтому свертка нормальных плотностей $p_{\overline{X}_n|\Theta}(\overline{x}_n|\theta)$ и $p_\Theta(\theta)$ дает для выборочного среднего $\overline{x}_n$ нормальную плотность

$$p_{\overline{X}_n}(\overline{x}_n) = \int_{-\infty}^{\infty} p_{\overline{X}_n|\Theta}(\overline{x}_n|\theta) p_\Theta(\theta)\, d\theta \tag{18}$$

со средним значением θ_0 и дисперсией $\sigma_0^2 + \sigma^2/n$ [13].

Согласно (16), плотности распределений значений θ_n и $\overline{x}_n$ удовлетворяют соотношению

$$p_{\Theta_n}(\theta_n) = p_{\overline{X}_n}(\overline{x}_n)\frac{d\overline{x}_n}{d\theta_n} = p_{\overline{X}_n}(\overline{x}_n)\frac{n\sigma_0^2 + \sigma^2}{n\sigma_0^2}\,,$$

которое дает соотношение дифференциальных энтропий

$$h(\Theta_n) = h\left(\overline{X}_n\right) + \ln\left(\frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\right). \tag{19}$$

Учитывая, что дифференциальная энтропия на множестве $\overline{X}_n$ значений выборочных средних с нормальной плотностью (18) равна

$$h(\overline{X}_n) = \frac{1}{2}\ln\left(2\pi e\left(\sigma_0^2 + \frac{\sigma^2}{n}\right)\right),$$

из (19) имеем

$$h(\Theta_n) = \frac{1}{2}\ln\left(2\pi e \frac{n\sigma_0^4}{n\sigma_0^2 + \sigma^2}\right).$$

Дисперсия (17) определяет наименьшую погрешность

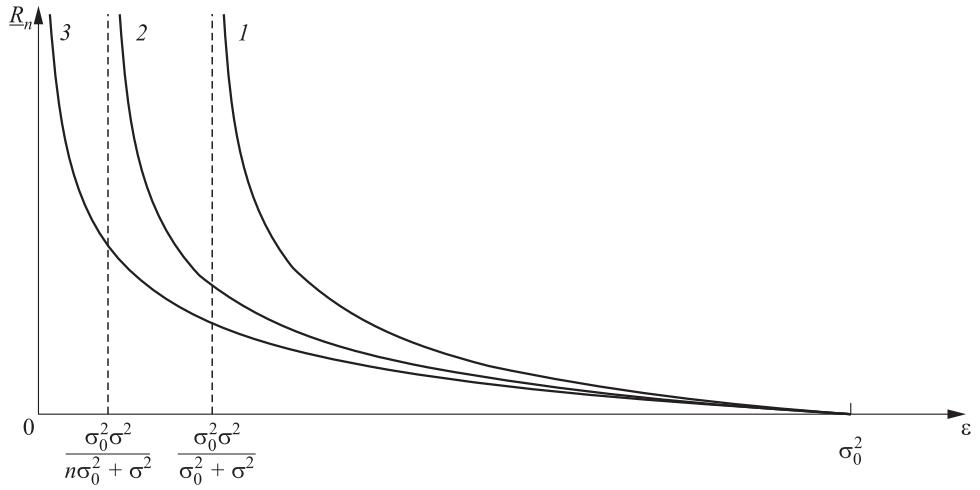
$$\varepsilon_{n\_\min} = \frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2}\,,$$

которая убывает с ростом размера выборки n.

Подстановка характеристик $h(\Theta_n)$ и $\varepsilon_{n\_\min}$ в (9) дает границу

$$\underline{R}_n(\varepsilon) = \\ = \frac{1}{2}\ln\left(\frac{n\sigma_0^4}{n\sigma_0^2 + \sigma^2}\right) - \frac{1}{2}\ln\left(\varepsilon - \frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}\right), \tag{20}$$

которая принимает нулевое значение в точке $\varepsilon_{\max} = \sigma_0^2$ при $n \geq 1$.



Поведение границы $\underline{R}_n(\varepsilon)$ для гауссовой модели с параметрами $n = 1$, $\sigma > 0$ (*1*), $n > 1$, $\sigma > 0$ (*2*) и $\sigma = 0$ (*3*)

Необходимо отметить, что при больших размерах выборки ($n \to \infty$) оценка

$$\hat{\theta}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$$

асимптотически оптимальна, поскольку связана с оценкой (16) преобразованием вида (12) с коэффициентами

$$\alpha_n = \frac{n\sigma_0^2 + \sigma^2}{n\sigma_0^2} \to 1\,; \;\; \beta_n = -\frac{\theta_0\sigma^2}{n\sigma_0^2} \to 0\,.$$

Избыточность средней погрешности квантованных оценок $\hat{\theta}_n$ при малом шаге квантования удовлетворяет соотношению (15).

Характер границы (20) при различных значениях параметров гауссовой модели показан на рисунке кривыми *1*–*3*. Кривые *1* и *2* соответствуют модели с одиночными наблюдениями ($n = 1$) и с выборками конечного размера ($n > 1$), когда $\sigma > 0$. Кривая *1* представляет нижнюю границу $\underline{R}_1(\varepsilon)$ функции скорость–погрешность для модели кодирования независимых гауссовых величин θ по сообщениям x на выходе канала с аддитивным гауссовым шумом $x - \theta$ [11]. Кривая *2* представляет нижнюю границу $\underline{R}_n(\varepsilon)$ средней взаимной информация как функцию средней квадратичной погрешности для модели оценивания математического ожидания θ гауссовой плотности по выборкам x^n. В случае $\sigma = 0$ кривая *3* представляет собой хорошо известную функцию скорость–погрешность

$$R(\varepsilon) = \frac{1}{2}\ln\left(\frac{\sigma_0^2}{\varepsilon}\right)$$

для модели кодирования независимых гауссовых величин θ с дисперсией σ_0^2 и допустимой погрешностью ε [9]. Следует отметить, что $\underline{R}_n(\varepsilon) \to R(\varepsilon)$, когда $n \to \infty$.

5 Заключение

В рамках вероятностной модели вычисления средней погрешности для оценок параметра по выборке независимых наблюдений получена аналитическая нижняя граница средней квадратичной погрешности как убывающая функция средней взаимной информации между множеством наблюдений и множеством возможных оценок. Граница получена в форме обращения известной в теории информации функции скорость–погрешность для модели, в которой наблюдаемые величины удовлетворяют заданной плотности распределения со случайным параметром. Полученная граница не зависит от метода построения оценки параметра, что позволяет использовать ее для сравнения эффективности различных методов восстановления параметра. При этом характеристикой эффективности любого метода служит избыточность средней квадратичной погрешности относительно нижней границы при количестве информации, которое задается энтропией множества оценок. Приведен пример вычисления границы для гауссовой модели, в которой независимые наблюдаемые величины удовлетворяют нормальной плотности с известной дисперсией и случайным средним значением с заданной нормальной плотностью распределения. Предложенный подход допускает обобщение для получения нижней границы средней квадратичной погрешности как функции количества используемой информации при оценивании векторного параметра, заданного набором независимых случайных величин.

Литература

1. *Bishop C. M.* Pattern recognition and machine learning. — New York, NY, USA: Springer, 2006. 746 p.
2. *Davisson L. D., McEliece R. I. Pursley M. B., Wallace M. S.* Efficient universal noiseless source codes // IEEE T. Inform. Theory, 1981. Vol. 27. No. 3. P. 269–279. doi: 10.1109/TIT.1981.1056355.
3. *Яковлева Т. В., Кульберг Н. С.* Методы математической статистики в решении задачи двухпараметрического анализа райсовского сигнала // Докл. Акад. наук. Сер. Математика, 2014. Т. 90. Вып. 3. С. 27–31.
4. *Вайчюлис М., Маркович Н. М.* Оценка параметров в суженном распределении Парето // Автоматика и телемеханика, 2021. Т. 82. Вып. 8. С. 85–107. doi: 10.31857/S0005231021080043.
5. *Кудрявцев А. А., Шестаков О. В., Шоргин С. Я.* Метод оценивания параметров изгиба, формы и масштаба гамма-экспоненциального распределения // Информатика и её применения, 2021. Т. 15. Вып. 3. С. 57–62. doi: 10.14357/19922264230308. EDN: IXMPXH.
6. *Кудрявцев А. А., Шестаков О. В.* Метод оценивания параметров гамма-экспоненциального распределения по выборке со слабо зависимыми компонентами // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 58–63. doi: 10.14357/19922264230308. EDN: PEXTVK.
7. *Duda R. O., Hart P. E., Stork D. G.* Pattern classification. — 2nd ed. — New York, NY, USA: John Wiley & Sons, 2001. 738 p.
8. *Боровков А. А.* Математическая статистика. Оценка параметров. Проверка гипотез. — М.: Наука, 1984. 472 с.
9. *Berger T.* Rate distortion theory. A mathematical basis for data compression. — Englewood Cliffs, NJ, USA: Prentice-Hall, 1971. 311 p.
10. *Gray R. M., Neuhoff D. L.* Quantization // IEEE T. Inform. Theory, 1998. Vol. 44. No. 6. P. 2325–2383. doi: 10.1109/18.720541.
11. *Dobrushin R. L., Tsybakov B. S.* Information transmission with additional noise // I. T. Inform. Theor., 1962. Vol. 8. No. 5. P. 293–304. doi: 10.1109/TIT.1962.1057738.
12. *Lange M. M., Lange A. M.* Information-theoretic lower bounds to error probability for the models of noisy discrete source coding and object classification // Pattern Recognition Image Analysis, 2022. Vol. 32. No. 3. P. 570–574. doi: 10.1134/S105466182203021X.
13. *Корн Г. А., Корн Т. М.* Справочник по математике для научных работников и инженеров. Определения, теоремы, формулы / Пер. с англ. — М.: Наука, 1970. 720 с. (*Korn G., Korn T.* Mathematical handbook for scientists and engineers. — New York – San Francisco – Toronto – London – Sydney: McGraw Hill Book Co., 1968. 1147 p.)

Поступила в редакцию 30.01.24

LOWER BOUND TO ESTIMATION DISTORTION OF A RANDOM PARAMETER FOR A GIVEN AMOUNT OF INFORMATION

M. M. Lange and A. M. Lange

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

Abstract: Given probability distribution density with an unknown value of a random parameter, a minimum of the average square distortion for the parameter estimates via the samples of random values as a function of the average mutual information between the samples and the estimates is investigated. This function is produced by inverting a modified rate distortion function as the dependency of the minimal values of the average mutual information on the appropriate values of the average distortion. The obtained smallest average square distortion as the function of the average mutual information is independent on an estimation form and this function yields the lower bound to the average distortion for the fixed values of the amount of information. The above relation is the bifactor fidelity decision criterion that allows one to compare various estimation functions by their efficiency in terms of the average distortion redundancy relative to the lower bound when the entropy of the quantized estimates is fixed.

Keywords: probability distribution density; data sample; parameter estimate; square distortion; mutual information; rate distortion function; lower bound; redundancy

DOI: 10.14357/19922264240203 **EDN:** EFZGYW

References

1. Bishop, C. M. 2006. *Pattern recognition and machine learning*. New York, NY: Springer. 746 p.
2. Davisson, L. D., R. I. McEliece, M. B. Pursley, and M. S. Wallace. 1981. Efficient universal noiseless source codes. *IEEE T. Inform. Theory* 27(3):269–279. doi: 10.1109/TIT.1981.1056355.
3. Yakovleva, T. V., and N. S. Kulberg. 2014. Methods of mathematical statistics in two-parameter analysis of Rician signals. *Dokl. Math.* 90(3):675–679. doi: 10.1134/S1064562414070060. EDN: UFVVGL.

4. Vaičiulis, M., and N. M. Markovich. 2021. Estimating the parameters of a tapered Pareto distribution. *Automat. Rem. Contr.* 82(8):1358–1377. doi: 10.1134/S000511792108004X.
5. Kudryavtsev, A. A., O. V. Shestakov, and S. Ya. Shorgin. 2021. Metod otsenivaniya parametrov izgiba, formy i masshtaba gamma-eksponentsial'nogo raspredeleniya [A method for estimating bent, shape and scale parameters of the gamma-exponential distribution]. *Informatika i ee Primeneniya — Inform. Appl.* 15(3):57–62. doi: 10.14357/19922264230308. EDN: IXMPXH.
6. Kudryavtsev, A. A., and O. V. Shestakov. 2023. Metod otsenivaniya parametrov gamma-eksponentsial'nogo raspredeleniya po vyborke so slabo zavisimymi komponentami [A method for estimating parameters of the gamma-exponential distribution from a sample with weakly dependent components]. *Informatika i ee Primeneniya — Inform. Appl.* 17(3):58–63. doi: 10.14357/19922264230308. EDN: PEXTVK.
7. Duda, R., P. Hart, and D. Stork. 2001. *Pattern classification*. 2nd ed. New York, NY: John Wiley and Sons. 738 p.
8. Borovkov, A. A. 1984. *Matematicheskaya statistika. Otsenka parametrov. Proverka gipotez* [Mathematical statistics. Parameter estimation. Hypothesis testing]. Moscow: Nauka. 472 p.
9. Berger, T. 1971. *Rate distortion theory. A mathematical basis for data compression*. Englewood Cliffs, NJ: Prentice-Hall. 311 p.
10. Gray, R. M., and D. L. Neuhoff. 1998. Quantization. *IEEE T. Inform. Theory* 44(6):2325–2383. doi: 10.1109/18.720541.
11. Dobrushin, R. L., and B. S. Tsybakov. 1962. Information transmission with additional noise. *IRE T. Inform. Theor.* 8(5):293–304. doi: 10.1109/TIT.1962.1057738.
12. Lange, M. M., and A. M. Lange. 2022. Information-theoretic lower bounds to error probability for the models of noisy discrete source coding and object classification. *Pattern Recognition Image Analysis* 32(3):570–574. doi: 10.1134/S105466182203021X.
13. Korn, G., and T. Korn. 1968. *Mathematical handbook for scientists and engineers*. New York – San Francisco – Toronto – London – Sydney: McGraw Hill Book Co. 1147 p.

Received January 30, 2024

Contributors

Lange Mikhail M. (b. 1945) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center "Computer Sciences and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; lange_mm@mail.ru

Lange Andrey M. (b. 1979) — Candidate of Science (PhD) in physics and mathematics, scientist, Federal Research Center "Computer Sciences and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; lange_am@mail.ru

ВЕРОЯТНОСТНАЯ МОДЕЛЬ ЗАТУХАНИЯ МОЩНОСТИ СИГНАЛА В СЦЕНАРИЯХ 3GPP TR 38.901 РАЗВЕРТЫВАНИЯ СЕТИ 5G*

Е. Д. Макеева[1], И. А. Кочеткова[2], С. Я. Шоргин[3]

Аннотация: Сети пятого (5G) и последующих поколений будут использовать терагерцевый диапазон радиочастот, что обеспечит сверхвысокую скорость передачи данных. Однако при этом возможны потери сигнала при прохождении через препятствия. Поэтому становится крайне важным моделирование распространения сигнала с помощью стохастической геометрии и использование актуальных моделей затухания сигнала. Модели для описания затухания сигнала для различных сценариев развертывания сети 5G в виде эмпирических формул содержатся в спецификации 3GPP TR 38.901. Тем не менее обычно для построения моделей стохастической геометрии используются упрощенные виды формул. В статье представлена функция распределения (ФР) затухания мощности сигнала при случайном расположении пользователей в соответствии со сценариями, описанными в 3GPP TR 38.901. На численных примерах показано, что разница значений с упрощенной формулой значительна и может привести к занижению оценки пропускной способности сети.

Ключевые слова: беспроводная сеть; 5G; 3GPP TR 38.901; мощность затухания сигнала; прямая видимость; непрямая видимость; стохастическая геометрия

DOI: 10.14357/19922264240204 **EDN:** EKLCAP

1 Введение

Сети пятого и последующих поколений будут использовать терагерцевый диапазон радиочастот, чтобы обеспечить сверхвысокую скорость передачи данных и пропускную способность. Однако использование миллиметровых волн связано со сложностями из-за потери сигнала при прохождении препятствий. Таким образом, для обеспечения производительности сетей 5G становится крайне важным моделирование распространения сигнала. Формула Шеннона–Хартли с формулой Фрииса задают пропускную способность канала

$$C = B\log_2\left(1 + \frac{P_t G_t G_r}{(N+I)\mathrm{PL}}\right),$$

где B — полоса пропускания канала; P_t — мощность передающей антенны; G_t — коэффициент усиления передающей антенны; G_r — коэффициент усиления приемной антенны; N — мощность шума; I — мощность интерференции; PL — мощность затухания сигнала (path loss, PL) на расстоянии от передающей антенны до приемной антенны [1]. Пропускная способность канала уже далее используется в управлении занятием радиоресурсов базовой станции (БС) для соблюдения необходимого качества обслуживания пользователей по требуемой скорости передачи данных.

Ввиду того что пользователи находятся на разных расстояниях от БС, значения мощностей затухания сигнала будут случайными. Как показано в работе [2], для учета влияния на пропускную способность канала случайного положения пользователей в соте применяется стохастическая геометрия. Рассмотреть совместное занятие радиоресурсов и случайный характер поведения пользователей позволяет модель на основе аппарата ресурсных систем массового обслуживания [3, 4]. Такие модели применяются для исследования различных сценариев развертывания сетей 5G [5, 6], например при анализе совместного обслуживания трафика со сверхнизкой задержкой и широкополосного трафика [7].

Модели для описания мощности PL затухания сигнала для разных сценариев отражены в спецификации 3GPP TR 38.901 [8]. И если при проведении имитационного моделирования исследователи по большей части полностью реализуют эти модели [9], то при построении моделей стохастической геометрии зачастую применяется упрощенный вид

*Публикация выполнена в рамках проекта № 025319-2-000 Системы грантовой поддержки научных проектов РУДН.

[1]Российский университет дружбы народов имени Патриса Лумумбы; Институт проблем управления имени В. А. Трапезникова Российской академии наук, elena-makeeva-96@mail.ru

[2]Российский университет дружбы народов имени Патриса Лумумбы; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, kochetkova-ia@rudn.ru

[3]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru

формул. В обзоре [2] рассмотрены различные виды функциональной зависимости затухания мощности сигнала от расстояния между пользователем и БС, которые применяют исследователи. Например, для простоты расчетов в работе [6] используются упрощенные формулы без учета кусочно-заданного вида функции для прямой видимости и максимума нескольких величин мощностей PL для непрямой видимости при построении ФР.

В данной статье получена ФР затухания мощности сигнала при случайном расположении пользователей в соответствии со сценариями 3GPP TR 38.901 развертывания сети 5G. Использованы формулы из этой спецификации, где приведены зависимости PL от расстояния между пользователем и БС. В данной статье закон распределения пользователей в соте взят произвольный, а для численного анализа — в соответствии с типовыми рекомендованными значениями параметров сценариев.

2 Затухание сигнала как функция от параметров сценария 3GPP

При исследовании распространения сигнала необходимо учитывать множество параметров сети, таких как частота, основные характеристики местности, высота принимающей и передающей антенн, конфигурация антенн и другие факторы. Для упрощения расчетов мощности PL затухания сигнала стандартом 3GPP TR 38.901 [8] были выделены основные сценарии развертывания сети 5G: макросота в городе (urban macro, UMa), микросота в городе (urban micro, UMi), макросота в сельской местности (rural macro, RMa), точка доступа внутри помещения (indoor hotspot, InH) и крытая фабрика (indoor factory, InF), — и путем экспериментов были получены эмпирические модели затухания сигнала для них. На основе этих моделей и в предположении случайного характера поведения пользователей в данном разделе получена ФР мощности затухания сигнала с учетом особенностей, описанных в данной спецификации.

Рассмотрим общее описание предлагаемых сценариев (рис. 1). Пусть передающая антенна БС расположена на высоте h_{BS}, использует несущую частоту f_c и создает покрытие радиуса R. Пользовательские устройства (ПУ) находятся на высоте h_{UT}, а проекция расстоянии от ПУ до БС составляет d.

В зависимости от своего расположения ПУ может находиться в зоне прямой видимости (line-of-sight, LOS) с устойчивым уровнем сигнала или вне этой зоны (non-line-of-sight, NLOS). Если ПУ расположено на расстоянии d, то вероятность того, что ПУ находится в зоне прямой видимости, представляет собой кусочно-заданную функцию:

$$\mathrm{Pr}_{\mathrm{LOS}}(d) = \begin{cases} \mathrm{Pr}_1^{\mathrm{LOS}}(d), & 0 = r_0 \le d < r_1; \\ \mathrm{Pr}_2^{\mathrm{LOS}}(d), & r_1 \le d < r_2; \\ \dots & \dots \\ \mathrm{Pr}_I^{\mathrm{LOS}}(d), & r_{I-1} \le d \le r_I = R, \end{cases} \tag{1}$$

где радиусы R_i определяют границы интервалов.

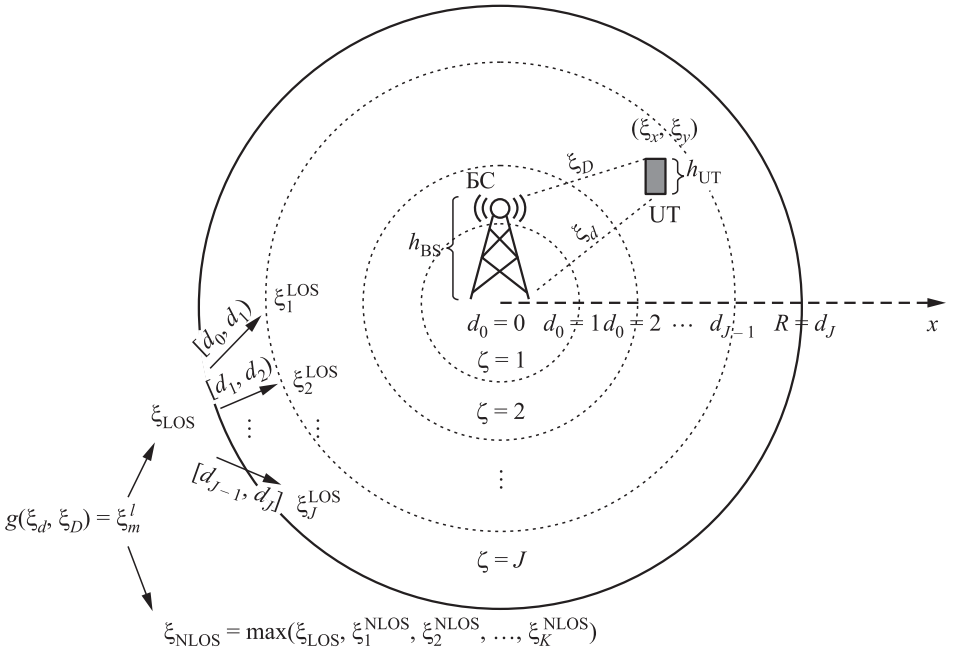


Рис. 1 Схема системной модели

Тогда мощность $\mathrm{PL}(d)$ затухания сигнала примет вид:

$$\mathrm{PL}\,(d) = \mathrm{PL}_{\mathrm{LOS}}(d)\mathrm{Pr}_{\mathrm{LOS}}(d) + \\ + \mathrm{PL}_{\mathrm{NLOS}}(d)\,[1 - \mathrm{Pr}_{\mathrm{LOS}}(d)]\,. \quad (2)$$

Мощность затухания сигнала в условиях прямой видимости LOS описывается кусочно-заданной функцией

$$\mathrm{PL}^{\mathrm{LOS}}(d) = \\ = \begin{cases} \mathrm{PL}_1^{\mathrm{LOS}}(d), & 0 = d_0 \le d < d_1; \\ \mathrm{PL}_2^{\mathrm{LOS}}(d), & d_1 \le d < d_2; \\ \dots & \dots \\ \mathrm{PL}_J^{\mathrm{LOS}}(d), & d_{J-1} \le d \le d_J = R, \end{cases} \quad (3)$$

где d_j — границы интервалов (break point distance), а в условиях непрямой видимости NLOS представляет собой максимум

$$\mathrm{PL}_{\mathrm{NLOS}}(d) = \\ = \max\left(\mathrm{PL}^{\mathrm{LOS}}(d), \mathrm{PL}_1^{\mathrm{NLOS}}(d), \dots, \mathrm{PL}_K^{\mathrm{NLOS}}(d)\right). \quad (4)$$

Каждая из компонент функций для случаев LOS и NLOS имеет схожую структуру:

$$\mathrm{PL}_m^l(d)[\mathrm{dB}] = \alpha_m^l[\mathrm{dB}] + \beta_m^l[\mathrm{dB}]\log_{10} D(d), \\ \mathrm{PL}_m^l(d) = \alpha_m^l \cdot D^{\beta_m^l}(d), \\ l = \begin{cases} \text{“LOS”}, & m = j = \overline{0,J}; \\ \text{“NLOS”}, & m = k = \overline{0,K}, \end{cases} \quad (5)$$

где $D(d) = \sqrt{d^2 + (h_{\mathrm{BS}} - h_{\mathrm{UT}})^2}$ — расстояние от ПУ до БС в трехмерном пространстве; α и β — коэффициенты модели затухания сигнала — константы для каждого из сценариев 3GPP TR 38.901.

3 Функция распределения затухания сигнала при случайном расположении пользователей

Примем теперь, что расстояние между ПУ и БС — случайная величина (СВ) ξ_d со значениями d и ФР $F_{\xi_d}(x)$. Тогда расстояние от ПУ до БС в трехмерном пространстве ξ_D будет функцией от СВ ξ_d с ФР

$$F_{\xi_D}(x) = \Pr\left(\xi_D \le x\right) = \\ = \Pr\left(\sqrt{\xi_d^2 + (h_{\mathrm{BS}} - h_{\mathrm{UT}})^2} \le x\right) = \\ = \Pr\left(\xi_d \le \sqrt{x^2 - (h_{\mathrm{BS}} - h_{\mathrm{UT}})^2}\right) = \\ = F_{\xi_d}\left(\sqrt{x^2 - (h_{\mathrm{BS}} - h_{\mathrm{UT}})^2}\right).$$

Случайная величина ξ_m^l — компонента функции затухания сигнала — зависит от СВ ξ_D и по формуле (5) имеет ФР

$$F_{\xi_m^l}(x) = \Pr\left(\xi_m^l \le x\right) = \Pr\left(\alpha_m^l (\xi_D)^{\beta_m^l} \le x\right) = \\ = \Pr\left(\xi_D \le \left(\frac{x}{\alpha_m^l}\right)^{1/\beta_m^l}\right) = F_{\xi_D}\left(\left(\frac{x}{\alpha_m^l}\right)^{1/\beta_m^l}\right) = \\ = F_{\xi_d}\left(\sqrt{\left(\frac{x}{\alpha_m^l}\right)^{2/\beta_m^l} - (h_{\mathrm{BS}} - h_{\mathrm{UT}})^2}\right), \\ l = \begin{cases} \text{“LOS”}, & m = j = \overline{0,J}; \\ \text{“NLOS”}, & m = k = \overline{0,K}. \end{cases}$$

Для затухания сигнала в условиях прямой видимости ФР СВ ξ_{LOS} по формуле (3) примет вид:

$$F_{\xi_{\mathrm{LOS}}}(x) = \Pr\left(\xi_{\mathrm{LOS}} \le x\right) = \\ = \sum_{j=1}^{J} \Pr\left(\xi_{\mathrm{LOS}} \le x \mid d_{j-1} \le \xi_d < d_j\right) \times \\ \times \Pr\left(d_{j-1} \le \xi_d < d_j\right) = \\ = \sum_{j=1}^{J} F_{\xi_j^{\mathrm{LOS}}}(x)\left[F_{\xi_d}(d_j) - F_{\xi_d}(d_{j-1})\right], \quad (6)$$

а для непрямой видимости ФР СВ ξ_{NLOS} по формуле (4) и с учетом [10] запишем как

$$F_{\xi_{\mathrm{NLOS}}}(x) = \Pr\left(\xi_{\mathrm{NLOS}} \le x\right) = \\ = \Pr\left(\max\left(\xi_{\mathrm{LOS}},\, \xi_1^{\mathrm{NLOS}}, \dots, \xi_K^{\mathrm{NLOS}}\right) \le x\right) = \\ = \Pr\left(\xi_{\mathrm{LOS}} \le x,\, \xi_1^{\mathrm{NLOS}} \le x, \dots, \xi_K^{\mathrm{NLOS}} \le x\right) = \\ = \Pr\left(\xi_{\mathrm{LOS}} \le x\right) \prod_{k=1}^{K} \Pr\left(\xi_k^{\mathrm{NLOS}} \le x\right) = \\ = F_{\xi_{\mathrm{LOS}}}(x) \prod_{k=1}^{K} F_{\xi_k^{\mathrm{NLOS}}}(x). \quad (7)$$

Наконец, ФР СВ ξ_{PL} затухания сигнала по формуле (2) запишем следующим образом:

$$F_{\xi_{\mathrm{PL}}}(x) = \Pr\left(\xi_{\mathrm{PL}} \le x\right) = \\ = \Pr\left(\xi_{\mathrm{LOS}}\xi_{\mathrm{Pr}_{\mathrm{LOS}}} + \xi_{\mathrm{NLOS}}\left[1 - \xi_{\mathrm{Pr}_{\mathrm{LOS}}}\right] \le x\right),$$

где $\xi_{\mathrm{Pr}_{\mathrm{LOS}}}$ — СВ вероятности расположения ПУ в зоне прямой видимости (1). Функция распределения $F_{\xi_{\mathrm{PL}}}(x)$ будет приближенно представлять собой свертку.

4 Численный анализ

В спецификации 3GPP TR 38.901 указаны основные диапазоны значений параметров для сценариев развертывания сетей 5G. Рассмотрим сценарии макросоты UMa и микросоты UMi в городе со следующим набором исходных данных: радиус действия БС $R = 5000$ м, центральная частота $f_c = 6$ ГГц, высота ПУ $h_{\mathrm{UT}} = 1{,}5$ м, высоты БС для UMa $h_{\mathrm{BS}} = 25$ м и для UMi $h_{\mathrm{BS}} = 10$ м. Предположим, что пользователи распределены равномерно в области действия БС радиусом R.

Согласно упрощенным формулам, подобным тем, что описаны в работе [6], ФР мощности PL затухания сигнала в условиях прямой и непрямой видимости могут быть представлены как

$$F_{\xi_{\mathrm{LOS}}}(x) = F_{\xi_1^{\mathrm{LOS}}}(x)\,;$$

$$F_{\xi_{\mathrm{NLOS}}}(x) = F_{\mathrm{Opt}}(x) = F_{\xi_d}\left(\sqrt{\left(\frac{x}{\alpha_{\mathrm{Opt}}}\right)^{2/\beta_{\mathrm{Opt}}} - (h_{\mathrm{BS}} - h_{\mathrm{UT}})^2}\right).$$

Коэффициенты модели затухания сигнала α_m^l [dB] и β_m^l [dB] для сценариев UMa и UMi, согласно [8], представлены в таблице.

В рамках данного численного анализа покажем графики ФР моделей PL затухания сигнала для сценариев UMa и UMi в условиях прямой и непрямой видимости по формулам (6) и (7) и упрощенным формулам. Графики с полученными результатами представлены на рис. 2 и 3. Во всех случаях график ФР по упрощенным формулам идет выше, что представляет собой верхнюю оценку. Однако при дальнейших расчетах с использованием упрощенных формул пропускная способность и максимальное число обслуженных пользователей в соте оказываются занижены. Таким образом, авторы рекомендуют при использовании модели затухания сигнала как компоненты, например в ресурсных системах массового обслуживания для анализа беспроводных сетей, использовать формулы (6) и (7).

Коэффициенты модели затухания сигнала для UMa и UMi

Зона	[dB]	UMa	UMi
LOS	α_1^{LOS}	$28 + 20\log_{10} f_c$	$32{,}4 + 20\log_{10} f_c$
	β_1^{LOS}	22	21
	α_2^{LOS}	$28 + 20\log_{10} f_c - 9\log_{10}\left(d_1^2 + (h_{\mathrm{BS}} - h_{\mathrm{UT}})^2\right)$	$32{,}4 + 20\log_{10} f_c - 9{,}5\log_{10}\left(d_1^2 + (h_{\mathrm{BS}} - h_{\mathrm{UT}})^2\right)$
	β_2^{LOS}	40	40
NLOS	α_1^{NLOS}	$13{,}54 + 20\log_{10} f_c - 0{,}6(h_{\mathrm{UT}} - 1{,}5)$	$22{,}4 + 21{,}3\log_{10} f_c - 0{,}3(h_{\mathrm{UT}} - 1{,}5)$
	β_1^{NLOS}	39,08	35,3
	α_{Opt}	$32{,}4 + 20\log_{10} f_c$	$32{,}4 + 20\log_{10} f_c$
	β_{Opt}	30	31,9

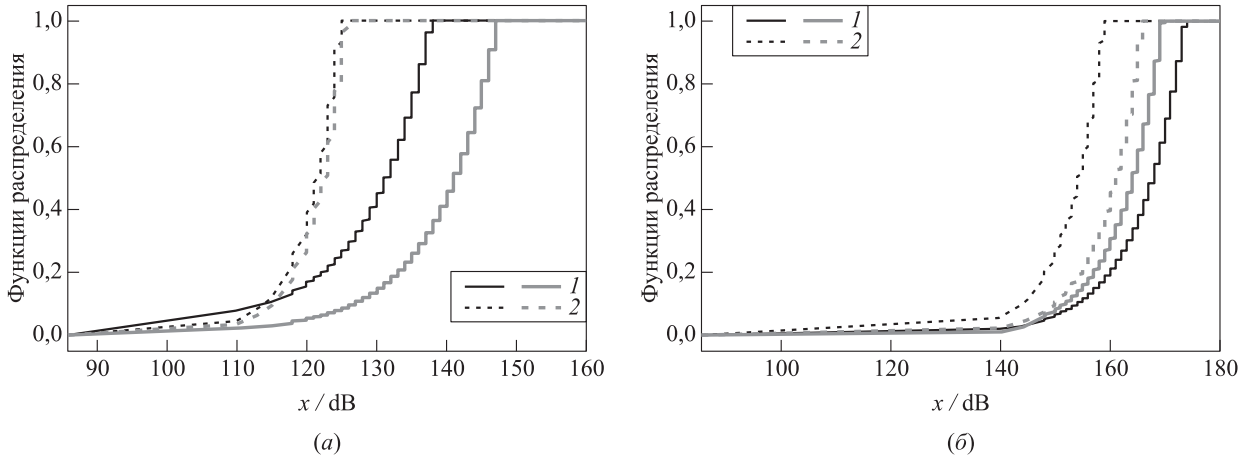


Рис. 2 Функции распределения PL для LOS (*а*) и NLOS (*б*) для сценариев UMa (черные кривые) и UMi (серые кривые): *1* — расчет по формулам (6) для LOS и (7) для NLOS; *2* — расчет по упрощенным формулам

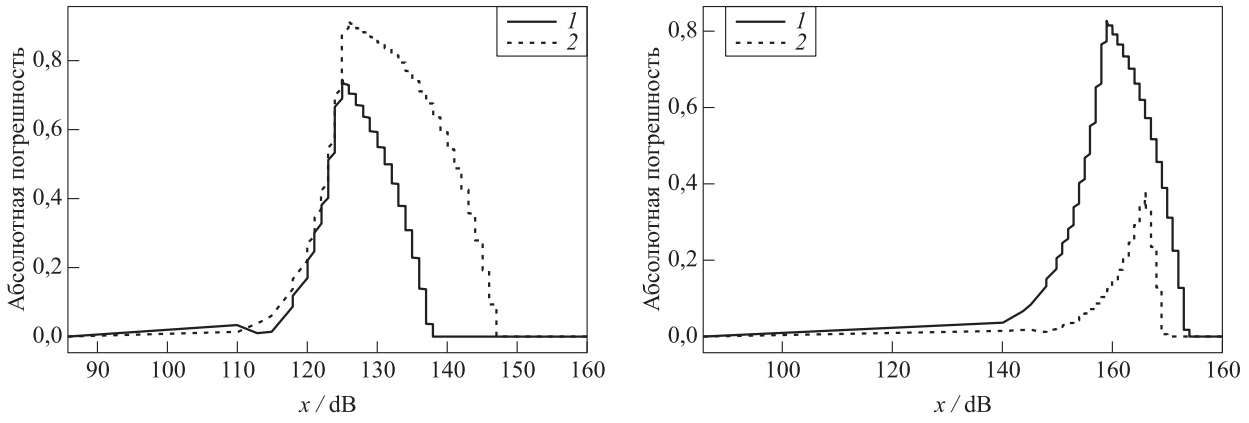


Рис. 3 Разница значений ФР PL для LOS (*а*) и NLOS (*б*): *1* — UMa; *2* — UMi

5 Заключение

В статье исследована модель затухания сигнала по формулам сценариев 3GPP TR 38.901. Была получена функция распределения (ФР) мощности затухания сигнала при случайном (произвольный закон) расположении пользователей в зоне покрытия БС. Она учитывает кусочно-заданный вид функции для LOS и максимум нескольких величин для NLOS. Проведен численный анализ для данных из спецификации 3GPP TR 38.90 для сценариев макро- и микросот в городе для сравнения ФР, представленных в данной работе, и ФР, полученных с помощью упрощенных формул. Результат анализа показал, что ФР по упрощенным формулам дает оценку сверху, что может понижать точность расчетов пропускной способности канала. Отметим, что авторы статьи не ставили перед собой задачу аналитического сравнения двух ФР, а хотели бы обратить внимание на несложный вид полученных формул, которые рекомендуют для использования как компоненту в ресурсных системах массового обслуживания при моделировании беспроводных сетей 5G/6G. Задачей дальнейшего исследования станет разработка ресурсной системы массового обслуживания с учетом случайного расположения пользователей в соте через представленную ФР для оценки схемы приоритетного обслуживания узкополосного трафика и прерывания обслуживания широкополосного трафика в сети 5G.

Литература

1. *Молчанов Д. А., Бегишев В. О., Самуйлов К. Е., Кучерявый Е. А.* Сети 5G/6G: архитектура, технологии, методы анализа и расчета. — М.: РУДН, 2022. 516 с.
2. *Hmamouche Y., Benjillali M., Saoudi S., Yanikomeroglu H., Renzo M. D.* New trends in stochastic geometry for wireless networks: A tutorial and survey // P. IEEE, 2021. Vol. 109. No. 7. P. 1200–1252. doi: 10.1109/JPROC.2021.3061778.
3. *Наумов В. А., Самуйлов К. Е.* О связи ресурсных систем массового обслуживания с сетями Эрланга // Информатика и её применения, 2016. Т. 10. Вып. 3. С. 9–14. doi: 10.14357/19922264160302.
4. *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.
5. *Маркова Е. В., Гольская А. А., Дзантиев И. Л., Гудкова И. А., Шоргин С. Я.* Сравнительный анализ показателей эффективности модели беспроводной сети межмашинного взаимодействия, работающей в рамках двух политик разделения радиоресурсов // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 108–116. doi: 10.14357/19922264190115.
6. *Moltchanov D. A., Sopin E. S., Begishev V. O., Samuylov A. K., Koucheryavy Y. A., Samouylov K. E.* A tutorial on mathematical modeling of 5G/6G millimeter wave and terahertz cellular systems // IEEE Commun. Surv. Tut., 2022. Vol. 24. No. 2. P. 1072–1116. doi: 10.1109/COMST.2022.3156207.
7. *Кочеткова И. А., Кущазли А. И., Харин П. А., Шоргин С. Я.* Модель схемы приоритетного доступа трафика URLLC и eMBB в сети пятого поколения в виде ресурсной системы массового обслуживания // Информатика и её применения, 2021. Т. 15. Вып. 4. С. 87–92. doi: 10.14357/19922264210412.
8. 3GPP TR 38.901. Study on channel model for frequencies from 0.5 to 100 GHz, 2024. Release 17.1.0.
9. *Bolla R., Bruschi R., Lombardo C., Mohammadpour A., Trivisonno R., Poe W. Y.* A 5G multi-gNodeB simulator for ultra-reliable 0.5–100 GHz communication in indoor Industry 4.0 environments // Comput. Netw., 2023. Vol. 237. Art. No. 110103. doi: 10.1016/j.comnet.2023.11010.
10. *Вентцель Е. С., Овчаров Л. А.* Теория вероятностей и ее инженерные приложения. — М.: Юстиция, 2018. 480 с.

Поступила в редакцию 15.03.24

STOCHASTIC PATH LOSS MODEL IN 5G NETWORK DEPLOYMENT SCENARIOS: A STUDY BASED ON 3GPP TR 38.901

E. D. Makeeva[1,2], I. A. Kochetkova[1,3], and S. Ya. Shorgin[3]

[1]RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation
[2]V. A. Trapeznikov Institute of Control Science of the Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation
[3]Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The fifth-generation (5G) and beyond networks will utilize radio frequencies in the terahertz spectrum, enabling extremely high data transmission rates. However, signal loss may occur when signals pass through obstacles, making it crucial to simulate signal propagation using stochastic geometry and apply up-to-date models for signal attenuation. The 3GPP TR 38.901 specification includes models that describe signal attenuation in various 5G network scenarios using empirical formulas. Nevertheless, simpler formulas are typically employed to create models based on stochastic geometry. The authors present the cumulative distribution function for path loss at random user locations according to the scenarios described in 3GPP TR 38.901. In numerical examples, it is shown that the difference in values with the simplified formula can be significant and lead to underestimation of the network's capacity

Keywords: wireless network; 5G; 3GPP TR 38.901; path loss; line-of-sight (LOS); non-line-of-sight (NLOS); stochastic geometry

DOI: 10.14357/19922264240204 **EDN:** EKLCAP

Acknowledgments

The publication has been supported by the RUDN University Scientific Projects Grant System, project No. 025319-2-000.

References

1. Moltchanov, D. A., V. O. Begishev, K. E. Samouylov, and Y. A. Koucheryavy. 2022. *Seti 5G/6G: arkhitektura, tekhnologii, metody analiza i rascheta* [The 5G/6G networks: Architecture, technologies, analysis methods, and calculations]. Moscow: RUDN University. 516 p.
2. Hmamouche, Y., M. Benjillali, S. Saoudi, H. Yanikomeroglu, and M. D. Renzo. 2021. New trends in stochastic geometry for wireless networks: A tutorial and survey. *P. IEEE.* 109(7):1200–1252. doi: 10.1109/JPROC.2021.3061778.
3. Naumov, V. A., and K. E. Samouylov. 2016. O svyazi resursnykh sistem massovogo obsluzhivaniya s setyami Erlanga [On relationship between queuing systems with resources and Erlang networks]. *Informatika i ee Primeneniya — Inform Appl.* 10(3):9–14. doi: 10.14357/19922264160302.
4. Gorbunova, A. V., V. A. Naumov, Yu. V. Gaidamaka, and K. E. Samouylov. 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.
5. Markova, E. V., A. A. Golskaia, I. L. Dzantiev, I. A. Gudkova, and S. Ya. Shorgin. 2019. Sravnitel'nyy analiz pokazateley effektivnosti modeli besprovodnoy seti mezhmashinnogo vzaimodeystviya, rabotayushchey v ramkakh dvukh politik razdeleniya radioresursov [Comparative analysis of performance measures for a wireless machine-to-machine network model operating within two radio resource management policies]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):108–116. doi: 10.14357/19922264190115.
6. Moltchanov, D. A., E. S. Sopin, V. O. Begishev, A. K. Samuylov, Y. A. Koucheryavy, and K. E. 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.
7. Kochetkova, I. A., A. I. Kushchazli, P. A. Kharin, and S. Ya. Shorgin. 2021. Model' skhemy prioritetnogo dostupa trafika URLLC i eMBB v seti pyatogo pokoleniya v vide resursnoy sistemy massovogo obsluzhivaniya [Model for analyzing priority admission control of URLLC and eMBB communications in 5G networks as a resource queuing system]. *Informatika i ee Primeneniya — Inform. Appl.* 15(4):87–92. doi: 10.14357/19922264210412.
8. 3GPP TR 38.901. 2023. Study on channel model for frequencies from 0.5 to 100 GHz, Release 17.1.0.
9. Bolla, R., R. Bruschi, C. Lombardo, A. Mohammadpour, R. Trivisonno, and W. Y. Poe. 2023. A 5G multi-gNodeB

simulator for ultra-reliable 0.5–100 GHz communication in indoor Industry 4.0 environments. *Comput. Netw.* 37:110103. doi: 10.1016/j.comnet.2023.11010.

10. Ventzel, E. S. and L. A. Ovcharov. 2018. *Teoriya veroyatnostey i ee inzhenernye prilozheniya* [Probability theory and its engineering applications]. Moscow: Justice. 480 p.

Received March 15, 2024

Contributors

Makeeva Elena D. (b. 1996) — PhD student, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; junior scientist, V. A. Trapeznikov Institute of Control Science of the Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation; elena-makeeva-96@mail.ru

Kochetkova Irina A. (b. 1985) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; kochetkova-ia@rudn.ru

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, principal scientist, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; sshorgin@ipiran.ru

МЕТОД ОЦЕНКИ ХАРАКТЕРИСТИК СИСТЕМ 5G/6G «НОВОЕ РАДИО» С УЧЕТОМ МАКРО- И МИКРОМОБИЛЬНОСТИ ПОЛЬЗОВАТЕЛЕЙ*

Д. Ю. Острикова[1], Е. С. Голос[2], В. А. Бесчастный[3], Е. А. Мачнев[4], В. С. Шоргин[5], Ю. В. Гайдамака[6]

Аннотация: Оценка производительности сотовых систем 5G/6G «новое радио», как правило, проводится в предположении о статичном местоположении пользователей и идеально направленных антеннах, которые оправданы для случая регулярной синхронизации между пользовательским устройством (ПУ) и базовой станцией (БС). Однако в случае работы с имеющими высокую энергоэффективность ПУ с ограниченным функционалом (RedCap — Reduced Capability) БС реже получает информацию о качестве принимаемого устройством сигнала, которое меняется при перемещении ПУ. Это приводит к необходимости исследования динамики показателей эффективности систем с ПУ RedCap во времени. В статье для анализа спектральной эффективности в зависимости от расстояния между БС и ПУ и направленности антенны ПУ в произвольный момент времени используются инструменты стохастической геометрии и теории случайных блужданий. Численный эксперимент показал, что макромобильность оказывает существенное влияние на спектральную эффективность, влияние микромобильности меньше и проявляется только на коротких промежутках времени, при этом размер фазированной антенной решетки (ФАР) на стороне БС практически не влияет на полученный результат.

Ключевые слова: 5G «новое радио»; mmWave; sub-THz; микромобильность; макромобильность; спектральная эффективность

DOI: 10.14357/19922264240205 **EDN:** JCUFHS

1 Введение

Ожидается, что современные системы сотовой связи 5G и будущие 6G, работающие в диапазонах миллиметровых (mmWave, 30–100 ГГц) длин волн и субтерагерцевых (sub-THz, 100–300 ГГц) частот, смогут обеспечить исключительно высокую пропускную способность на участке беспроводного доступа [1, 2]. Это позволит внедрить множество новых приложений, включая передачу потокового видео в формате сверхвысокой четкости, приложения телемедицины, виртуальной и дополненной реальности [3]. Использование чрезвычайно высоких частот в диапазонах mmWave/sub-THz требует применения как на стороне БС, так и на стороне ПУ фазированных антенных решеток, работающих в режиме формирования луча, расширяя зону действия БС [4, 5]. Для своевременной компенсации ущерба от потери связи, характерной для систем с направленными лучами, применяются методы отслеживания луча, синхронизирующие направления лучей между ПУ и БС на малых временны́х интервалах [6].

Необходимость экономии ресурса аккумулятора ПУ привела к появлению устройств с ограниченным функционалом (RedCap) [7], для которых процедура управления радиоресурсами (*англ*. Radio Resource Management, RRM) подразумевает пропуск блоков сигналов синхронизации (*англ*. Synchronization Signal Blocks, SSB) для повышения энергоэффективности. Следствием более редкой синхронизации становится риск потери связи с подвижным ПУ из-за эффектов макро- и микромобильности [8, 9] на средних и больших интервалах времени. Под макромобильностью понимается перемещение пользователей внутри зоны покрытия, приводящее к изменению расстояния между ПУ и БС, а также к нарушению взаимно-

*Исследование выполнено за счет гранта Российского научного фонда № 23-79-10084, https://rscf.ru/project/23-79-10084.

[1]Российский университет дружбы народов им. Патриса Лумумбы, ostrikova-dyu@rudn.ru

[2]Российский университет дружбы народов им. Патриса Лумумбы, golos-es@rudn.ru

[3]Российский университет дружбы народов им. Патриса Лумумбы, beschastnyy-va@rudn.ru

[4]Российский университет дружбы народов им. Патриса Лумумбы, machnev-ea@rudn.ru

[5]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vshorgin@ipiran.ru

[6]Российский университет дружбы народов им. Патриса Лумумбы; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, gaydamaka-yuv@rudn.ru

го выравнивания направленных лучей ФАР БС и ПУ. Под микромобильностью понимается незначительное изменение положения ПУ в пространстве из-за вращения ПУ в руках пользователя, при котором луч ФАР ПУ отклоняется от направления на БС, но расстояние между ПУ и БС не меняется. Таким образом, к применяемым ранее моделям стохастической геометрии, предполагающим методы анализа при статическом расположении пользователей [10–13], необходимо добавить инструменты моделирования мобильности ПУ.

В данной работе использован подход к моделированию макро- и микромобильности ПУ на основе диффузионных процессов, предложенный в [14] при сравнении стратегий энергосбережения для устройств с ограниченным функционалом для приложений промышленной автоматизации. В отличие от [14], целью данной статьи ставится разработка метода оценки характеристик производительности соты 5G/6G «новое радио», в частности спектральной эффективности системы, с учетом перемещения пользователей внутри зоны покрытия при развертывании в городской среде.

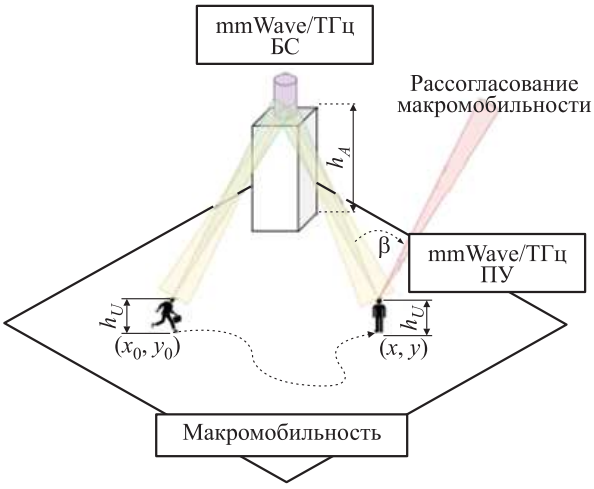
2 Системная модель

Для анализа влияния эффектов макро- и микромобильности на характеристики систем mmWave/sub-THz исследована модель соты сети 5G/6G «новое радио» [1], изображенная на рис. 1, *б*. Зона покрытия БС, расположенной в точке с координатами $(0, 0)$, имеет форму квадрата с длиной стороны L м. Подвижное ПУ, перемещающееся в зоне покрытия, в начальный момент времени t_0 расположено в точке (x_0, y_0), в произвольный момент $t > 0$ — в точке (x, y). Высоты БС и ПУ постоянны и равны h_A и h_U соответственно. Макромобильность при перемещении ПУ на средних и больших временны́х интервалах отражается в динамике мощности принимаемого ПУ сигнала из-за изменяющегося расстояния между БС и ПУ и рассогласования направленности лучей антенн ПУ и БС (угол β на рис. 1, *а* и 1, *б*). Микромобильность приводит только к рассогласованию направленности лучей (угол γ на рис. 1, *в*), при этом эффект наблюдается на малых временны́х интервалах.

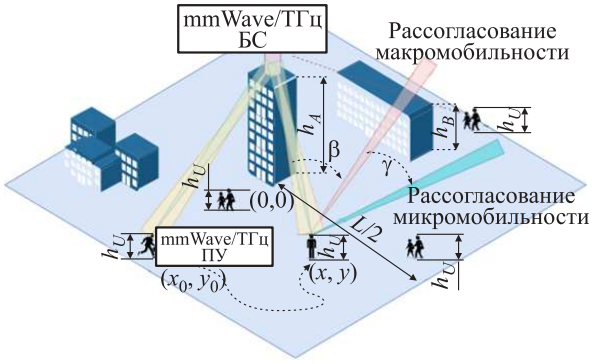
Для анализа спектральной эффективности в момент t используем выражение [15]:

$$S_E(t) = \log_2\left(1 + \frac{P_T G_A(t) G_U(t) L(y,t)}{N_0 + I_M}\right). \quad (1)$$

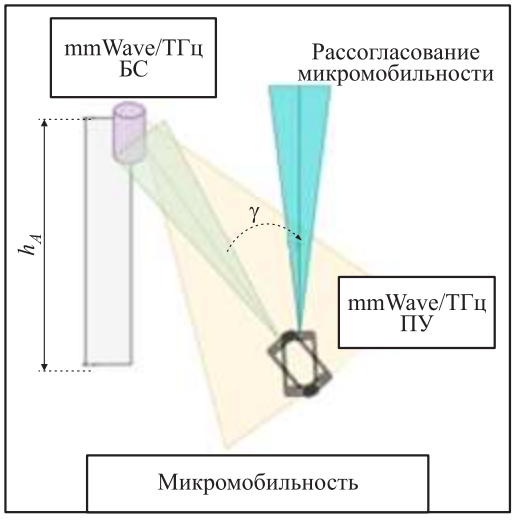
где P_T — мощность излучения антенны БС; $G_A(t)$ и $G_U(t)$ — коэффициенты усиления антенн БС и ПУ



(*а*)



(*б*)



(*в*)

Рис. 1 Эффекты макро- и микромобильности в системах «новое радио»

в момент t; $L(y, t)$ — коэффициент потери мощности принимаемого ПУ сигнала на расстоянии y от БС в момент t; N_0 — шум; I_M — интерференция.

Заметим, что значения функций $L(y,t)$, $G_A(t)$ и $G_U(t)$ вычисляются в зависимости от трех параметров, определяемых новым местоположением ПУ и отклонением оси ПУ от направления на БС в момент t, возникшими вследствие перемещения ПУ и вращения ПУ в руках пользователя, а именно: от расстояния $g(t)$ по оси y и от угловых отклонений $\beta_H(t)$, $\gamma_H(t)$ и $\beta_V(t)$, $\gamma_V(t)$ оси антенны ПУ от направления на БС в горизонтальной и вертикальной плоскостях соответственно. Таким образом, для оценки значений функций $L(y,t)$, $G_A(t)$ и $G_U(t)$, меняющихся во времени, необходимо выбрать метод моделирования мобильности ПУ.

В [14] для моделирования макромобильности использованы два независимых диффузионных процесса блуждания частицы по осям $0x$ и $0y$, ограниченные в $(0,L)$, с коэффициентами диффузии D_x и D_y. Для одномерного ограниченного диффузионного процесса плотность вероятности $p(x,t|x_0,t_0)$ найти частицу в точке x в момент t, учитывая, что она находилась в точке x_0 в момент t_0, подчиняется частному случаю второго закона диффузии Фика и описывается уравнением Фоккера–Планка с нулевым сносом

$$\frac{\partial p(x,t|x_0,t_0)}{\partial t} = \frac{D\partial^2 p(x,t|x_0,t_0)}{\partial x^2} \qquad (2)$$

с начальным условием $p(x,t|x_0,t_0) = \delta(x-x_0)$, где D — коэффициент диффузии, характеризующий скорость перемещения частицы [16]. Вид полученного методом разделения переменных решения уравнения (2) в указанных выше ограничениях показан в [14]:

$$p(x,t|x_0,t_0) = \frac{1}{L} + \frac{2}{L}\sum_{n=1}^{\infty}\left\{\exp\left[-\left(\frac{n\pi}{2}\right)^2\frac{t-t_0}{\tau}\right]\times \right.$$
$$\left.\times\cos\left(\frac{nx\pi}{L}\right)\cos\left(\frac{nx_0\pi}{L}\right)\right\},$$

где $\tau = L^2/D$ — время релаксации.

При условии старта перемещения ПУ в начальный момент t_0 из точки $(x_0,0)$ значения расстояния $y(t)$ и углов азимута $\beta_H(t)$ и элевации $\beta_V(t)$ отклонения оси ПУ вследствие макромобильности в произвольный момент t могут быть вычислены по следующим формулам [14]:

$$y(t) = \sqrt{[X(t)]^2 + (h_A - h_U)^2};$$
$$\beta_V(t) = \arccos\left[\frac{X(t)}{Y(t)}\right];$$
$$\beta_H(t) = \arctan\left(\frac{Y(t)}{X(t_0)}\right),$$

где состояния $X(t_0)$ и $X(t)$ независимого одномерного диффузионного процесса с коэффициентом диффузии D_x, ограниченного в $(0,L)$, представляют собой координату ПУ на оси $0x$ в начальный t_0 и текущий t моменты времени; состояние $Y(t)$ такого же процесса с коэффициентом диффузии D_y с началом движения в точке 0 — координату ПУ на оси $0y$ в момент t, при этом $X(t_0) = x_0$ и $Y(t_0) = 0$.

Микромобильность ПУ в приложениях X-VR моделируется также с помощью двух диффузионных процессов $G_H(t)$ и $G_V(t)$ с коэффициентами диффузии D_H и D_V с начальными точками $G_H(t_0) = G_V(t_0) = 0$. Состояния независимых процессов $G_H(t)$ и $G_V(t)$ определяют значения угла азимута $\gamma_H(t)$ и угла элевации $\gamma_V(t)$, т. е. отклонение оси антенны ПУ от направления на БС в момент t в горизонтальной и вертикальной плоскостях соответственно.

3 Численный анализ

В работе предполагается наличие планарных симметричных ФАР как на БС, так и на ПУ. Рассматриваются ПУ с двумя физическими антеннами, расположенными на противоположных сторонах устройства, диаграммы направленности антенн соответствуют стандарту 3GPP TR 37.977 [17]. Модель усиления антенны имеет вид:

$$G_U(\beta_H,\beta_V) = $$
$$= \varepsilon\rho_H\left(\beta_H + \gamma_H(t)\right)\rho_V\left(\beta_V + \gamma_V(t)\right), \qquad (3)$$

где ε — коэффициент усиления при идеальном выравнивании лучей от БС и от ПУ; $\rho_H(\beta_H)$ и $\rho_V(\beta_V)$ — функции направленности для угловых отклонений β_H и β_V оси антенны ПУ от направления на БС в горизонтальной и вертикальной плоскостях при макромобильности, $\beta_H, \beta_V \in [0,\pi]$; γ_H и γ_V — дополнительные угловые отклонения, возникающие при микромобильности в соответствующих плоскостях, $\gamma_H, \gamma_V \in [-\pi/2, \pi/2]$.

Заметим, что в предположении о симметричной конической антенне, т. е. $\beta_H = \beta_V$ и $\rho_H(\beta_H) = \rho_V(\beta_V) = \rho(\beta)$, параметры ε и $\rho(\beta)$ зависят от ширины 2D-луча α, определяемой числом антенных элементов N, и имеют следующий вид [18]:

$$\alpha = 2\arccos\frac{2{,}782}{N\pi};$$
$$\varepsilon = \frac{2}{1-\cos(\alpha/2)};$$
$$\rho(\beta) = \begin{cases} 1 - \dfrac{\beta}{\alpha}, & \beta \le \alpha; \\ 0 & \text{в противном случае.} \end{cases}$$

Системные параметры

Обозначение	Описание	Значения
L	Длина стороны квадратной зоны	100 м
h_U	Высота ПУ	1,5 м
h_A	Высота БС NR	4 м
P_T	Излучаемая мощность антенны БС	23 дБ
f_C	Несущая частота	28 ГГц
ζ	Коэффициент затухания сигнала	2,1
N_0	Шум	-174 дБи
I_M	Интерференция	3 дБи
$N_H \times N_V$	Число антенных элементов ФАР БС/ПУ	4×4, 8×8, 15×15
Δx, Δy	Средняя скорость перемещения ПУ вдоль осей	0,8 м/с
$\Delta\varphi$, $\Delta\theta$	Средняя скорость отклонения оси ПУ в горизонтальной и вертикальной плоскостях	0–4 град/с

Коэффициент $L(y,t)$ потери мощности принимаемого ПУ сигнала на расстоянии y [м] от БС в момент t вычисляется по формуле:

$$L(y,t) = 10^{2\lg f_C + 3{,}24} y^{-\zeta},$$

где f_C — несущая частота; ζ — коэффициент затухания сигнала.

Параметры среды распространения сигнала и системы по умолчанию представлены в таблице.

На рис. 2 показано влияние микромобильности на среднюю спектральную эффективность, вычисляемую с помощью (1), для различных типов приложений и двух точек (x_0, y_0) старта ПУ — начальных точек $(20, 0)$ и $(80, 0)$. Микромобильность, представленная в предложенной модели процессами изменения угловых отклонений $\gamma_H(t)$ и $\gamma_V(t)$, влияет на усиление антенны ПУ, как показано в (3). Средняя скорость отклонения осей ПУ $\Delta\varphi = \Delta\theta$ составляет 2 град/с для видеоприложений и 4 град/с для приложений VR. Средняя скорость перемещения ПУ из-за макромобильности составляет $\Delta x = \Delta y = 0{,}8$ м/с. По умолчанию $L = 100$ м, конфигурация антенны БС — 15×15 элементов.

Влияние макромобильности на среднюю спектральную эффективность показано на рис. 3 для старта ПУ из начальной точки $x_0 = 20$, $y_0 = 0$ Средняя скорость перемещения ПУ из-за макромобильности $\Delta x = \Delta y$ варьируется в пределах 0,8–2 м/с, средняя скорость отклонения оси ПУ из-за микромобильности $\Delta\varphi = \Delta\theta = 2$ град/с. По умолчанию $L = 100$ м, конфигурация антенны БС — 15×15 элементов.

Полученные результаты показывают, что резкие колебания спектральной эффективности вследствие микромобильности проявляются на малых временны́х интервалах в течение 1–3 с после выравнивания луча, тогда как макромобильность влияет на показатели средней спектральной эффективности в течение более длительного интервала

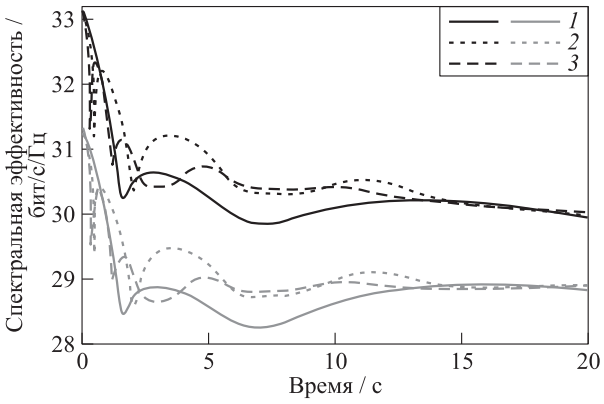


Рис. 2 Спектральная эффективность в зависимости от микромобильности: *1* — расчет по формуле (1); *2* — видеоприложения; *3* — приложения VR; черные кривые — $x_0 = 20$ м; серые кривые — $x_0 = 80$ м

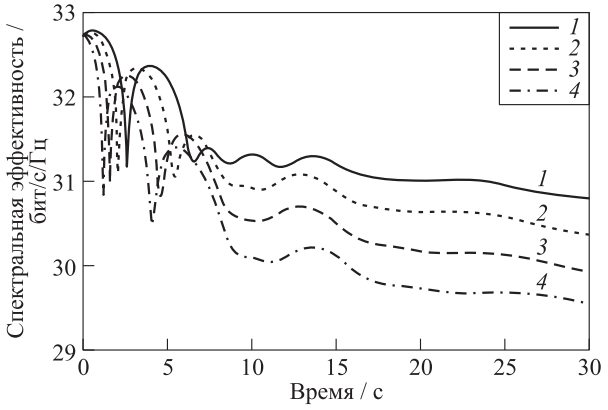


Рис. 3 Спектральная эффективность в зависимости от макромобильности: *1* — $D_M = 25$; *2* — 50; *3* — 100; *4* — $D_M = 200$

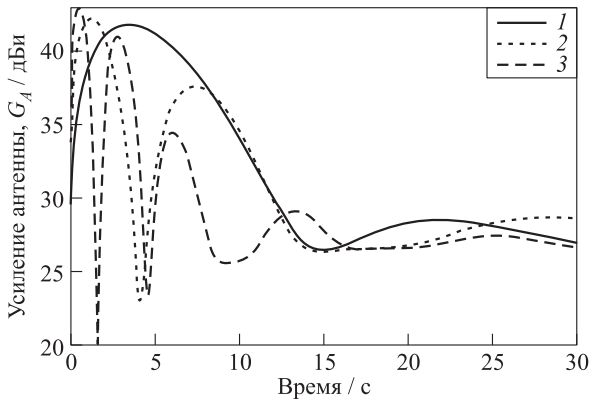


Рис. 4 Усиление на антенне БС в зависимости от конфигурации антенной решетки: *1* — 4×4; *2* — 8×8; *3* — 15×15

времени для всего диапазона исследованных скоростей перемещения ПУ. Активное перемещение ПУ при скорости 2 м/с ($D_M = D_x = D_y = 200$) приводит к более быстрому снижению спектральной эффективности по сравнению с низкой скоростью перемещения ПУ 0,8 м/с ($D_M = D_x = D_y = 25$), что связано с различающейся для разных скоростей ПУ динамикой среднего расстояния до БС.

Кроме расстояния между БС и ПУ спектральная эффективность зависит от конфигурации антенны, в частности от ширины угла основного лепестка диаграммы направленности антенны БС. Как показано на рис. 4, с течением времени происходит значительное падение коэффициента усиления антенны БС из-за рассинхронизации лучей в результате смещения ПУ от оси антенны БС в сторону границ лепестка, так что коэффициент снижается до усредненных значений около 25 дБи. При этом всплески на графике соответствуют попаданию ПУ в боковые лепестки антенны БС. Ожидаемо резче и быстрее коэффициент падает для антенны 15×15 элементов, формирующей более узкие лучи с более ярко выраженными боковыми лепестками. Заметим, что дополнительное усиление, получаемое при выравнивании антенны ПУ до идеальной направленности на БС, не вносит существенного вклада в спектральную эффективность из-за использования для ее оценки медленно возрастающей логарифмической функции.

4 Заключение

Перспективы внедрения новых устройств с ограниченным функционалом для систем 5G «новое радио» и новых механизмов энергосбережения на стороне ПУ, позволяющих пропускать циклы синхронизации с БС, мотивировали авторов на разработку аналитического аппарата для оценки спектральной эффективности в беспроводном канале в условиях сочетания макро- и микромобильности пользователей. Модель учитывает параметры диаграммы направленности антенн БС и ПУ, описанные в стандартах 3GPP, а также нарушение выравнивания лучей антенн во времени, вызванное вращением ПУ в руках пользователя при одновременном перемещении пользователя в зоне покрытия БС. Численные результаты показали, что разница между спектральной эффективностью, полученной в предположении о статичном местоположении пользователей и идеально направленных антеннах БС и ПУ, и зависящей от времени спектральной эффективностью незначительна для покрытия малых сот и находится в пределах примерно 5%–10% для широкого диапазона параметров системы. Это означает, что при необходимости связь может поддерживаться без идеального выравнивания антенн БС и ПУ с несколько ухудшенным качеством. Влияние диаграммы направленности антенны БС на изменения спектральной эффективности во времени весьма ограничено, что приводит к разнице до 5% между ФАР 4×4 элементов и 15×15 элементов. С точки зрения моделирования стоит отметить, что использование реалистичных диаграмм направленности излучения антенн критически важно при исследовании систем с макро- и микромобильностью.

Литература

1. *Holma H., Toskala A., Nakamura T.* 5G technology: 3GPP new radio. — Hoboken, NJ, USA: John Wiley & Sons, 2020. 536 p.
2. *Jiang W., Han B., Habibi M. A., Schotten H. D.* The road towards 6G: A comprehensive survey // IEEE Open J. Communications Society, 2021. Vol. 2. P. 334–366. doi: 10.1109/OJCOMS.2021.3057679.
3. *Saad W., Bennis M., Chen M.* A vision of 6G wireless systems: Applications, trends, technologies, and open research problems // IEEE Network, 2019. Vol. 34. No. 3. P. 134–142. doi: 10.1109/MNET.001.1900287.
4. *Zhang J., Ge X., Li Q., Guizani M., Zhang Y.* 5G millimeter-wave antenna array: Design and challenges // IEEE Wirel. Commun., 2016. Vol. 24. No. 2. P. 106–112. doi: 10.1109/MWC.2016.1400374RP.
5. *Guo Y. J., Ziolkowski R. W.* Advanced antenna array engineering for 6G and beyond wireless communications. — Hoboken, NJ, USA: John Wiley & Sons, 2021. 316 p.
6. 3GPP. NR; Physical layer; General description (Release 18). 3GPP TS 38.201 V18.0.0, 2023.
7. 3GPP. Study on support of reduced capability NR devices (Release 17): Technical Specification 38.875

V17.0.0, 2021. 136 p. https://www.3gpp.org/ftp/Specs/archive/38_series/38.875/38875-h00.zip.

8. *Stepanov N., Moltchanov D., Begishev V., Turlikov A., Koucheryavy Y.* Statistical analysis and modeling of user micromobility for THz cellular communications // IEEE T. Veh. Technol., 2021. Vol. 71. No. 1. P. 725–738. doi: 10.1109/TVT.2021.3124870.
9. *Moltchanov D., Gaidamaka Y., Ostrikova D., Beschastnyi V., Koucheryavy Y., Samouylov K.* Ergodic outage and capacity of terahertz systems under micromobility and blockage impairments // IEEE T. Wirel. Commun., 2021. Vol. 21. No. 5. P. 3024–3039. doi: 10.1109/TWC.2021.3117583.
10. *Haenggi M.* Stochastic geometry for wireless networks. — Cambridge, U.K.: Cambridge University Press, 2012. 298 p.
11. *Petrov V., Komarov M., Moltchanov D., Jornet J. M., Koucheryavy Y.* Interference and SINR in millimeter wave and terahertz communication systems with blocking and directional antennas // IEEE T. Wirel. Commun., 2017. Vol. 16. No. 3. P. 1791–1808. doi: 10.1109/TWC.2017.2654351.
12. *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. EDN: YAMDIL.
13. *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.
14. *Бесчастный В. А., Голос Е. С., Острикова Д. Ю., Мачнев Е. А., Шоргин В. С., Гайдамака Ю. В.* Анализ совместного использования стратегий энергосбережения для устройств 5G с ограниченным функционалом // Системы и средства информатики, 2023. Т. 33. № 4. С. 69–81. doi: 10.14357/08696527230407. EDN: KATMLB.
15. *Rappaport T. S.* Wireless communications: principles and practice. — 2nd ed. — Cambridge, U.K.: Cambridge University Press, 2024. 708 p. doi: 10.1017/9781009489843.
16. *Risken H.* Fokker–Planck equation. — Springer, 1996. 472 p.
17. 3GPP. Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA); Verification of radiated multi-antenna reception performance of User Equipment (UE). 3GPP TR 37.977 V17.0, 2022.
18. *Chukhno O., Chukhno N., Galinina O., Andreev S., Gaidamaka Y., Samouylov K., Araniti G.* A Holistic assessment of directional deafness in mmWave-based distributed 3D networks // IEEE T. Wirel. Commun., 2022. Vol. 21. No. 9. P. 7491–7505. doi: 10.1109/TWC.2022.3159086.

Поступила в редакцию 14.03.24

ASSESSING THE CHARACTERISTICS OF 5G/6G “NEW RADIO” SYSTEMS WITH USER’S MACRO- AND MICROMOBILITY

D. Yu. Ostrikova[1], E. S. Golos[1], V. A. Beschastnyi[1], E. A. Machnev[1], V. S. Shorgin[2], and Yu. V. Gaidamaka[1,2]

[1]RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation
[2]Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The performance of 5G/6G cellular “new radio” systems is typically evaluated using static user location and perfectly directional antennas which are justified in the case of regular synchronisation between the user equipment (UE) and the base station (BS). However, in the case of high energy-efficient UEs with limited RedCap functionality, BS is less likely to get information about the quality of the signal received by the device which changes when the UE moves. This leads to the need to investigate the dynamics of the performance indicators of systems with RedCap UEs over time. In the paper, tools of stochastic geometry and random walk theory are used to analyze the spectral efficiency depending on the distance between the BS and the UE and the directionality of the UE antenna at a random moment of time. A numerical experiment has shown that macromobility has a significant impact on the spectral efficiency, the impact of micromobility is smaller and appears only at short time intervals, while the size of the phased antenna array on the BS side practically does not affect the obtained result.

Keywords: 5G new radio; mmWave; sub-THz; micromobility; macromobility; spectral efficiency

DOI: 10.14357/19922264240205 **EDN:** JCUFHS

Acknowledgments

The reported study was funded by the Russian Science Foundation, project No. 23-79-10084.

References

1. Holma, H., A. Toskala, and T. Nakamura. 2020. *5G technology: 3GPP new radio*. New York, NY: John Wiley & Sons. 536 p.
2. Jiang, W., B. Han, M. A. Habibi, and H. D. Schotten. 2021. The road towards 6G: A comprehensive survey. *IEEE Open J. Communications Society* 2:334–366. doi: 10.1109/OJCOMS.2021.3057679.
3. Saad, W., M. Bennis, and M. Chen. 2019. A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE Network* 34(3):134–142. doi: 10.1109/MNET.001.1900287.
4. Zhang, J., X. Ge, Q. Li, M. Guizani, and Y. Zhang. 2016. 5G millimeter-wave antenna array: Design and challenges. *IEEE Wirel. Commun.* 24(2):106–112. doi: 10.1109/MWC.2016.1400374RP.
5. Guo, Y. J., and R. W. Ziolkowski. 2021. *Advanced antenna array engineering for 6G and beyond wireless communications*. Hoboken, NJ: John Wiley & Sons. 316 p.
6. 3GPP. 2023. NR; Physical layer; General description (Release 18). 3GPP TS 38.201 V18.0.0.
7. 3GPP. 2021. Study on support of reduced capability NR devices (Release 17): Technical Specification 38.875 V17.0.0. 136 p. Available at: https://www.3gpp.org/ftp/Specs/archive/38_series/38.875/38875-h00.zip (accessed May 17, 2024).
8. Stepanov, N. V., D. Moltchanov, V. Begishev, A. Turlikov, and Y. Koucheryavy. 2021. Statistical analysis and modeling of user micromobility for THz cellular communications. *IEEE T. Veh. Technol.* 71(1):725–738. doi: 10.1109/TVT.2021.3124870.
9. Moltchanov, D., Y. Gaidamaka, D. Ostrikova, V. Beschastnyi, Y. Koucheryavy, and K. Samouylov. 2021. Ergodic outage and capacity of terahertz systems under micromobility and blockage impairments. *IEEE T. Wirel. Commun.* 21(5):3024–3039. doi: 10.1109/TWC.2021.3117583.
10. Haenggi, M. 2012. *Stochastic geometry for wireless networks*. Cambridge, U.K.: Cambridge University Press. 298 p.
11. Petrov, V., M. Komarov, D. Moltchanov, J. M. 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(3):1791–1808. doi: 10.1109/TWC.2017.2654351.
12. 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. EDN: YAMDIL.
13. 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.
14. Beschastnyi, V. A., E. S. Golos, D. Yu. Ostrikova, E. A. Machnev, V. S. Shorgin, and Yu. V. Gaidamaka. 2023. Analiz sovmestnogo ispol'zovaniya strategiy energosberezheniya dlya ustroystv 5G s ogranichennym funktsionalom [Analysis of joint usage of energy conservation strategies for 5G devices with reduced capability]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(4):69–81. doi: 10.14357/08696527230407. EDN: KATMLB.
15. Rappaport, T. S. 2024. *Wireless communications: Principles and practice*. 2nd ed. Cambridge, U.K.: Cambridge University Press. 708 p. doi: 10.1017/9781009489843.
16. Risken, H. 1996. *Fokker–Planck equation*. Springer. 472 p.
17. 3GPP. 2022. Universal terrestrial radio access (UTRA) and evolved universal terrestrial radio access (E-UTRA); Verification of radiated multi-antenna reception performance of User Equipment (UE). 3GPP TR 37.977 V17.0.0 Release 17.
18. Chukhno, O., N. Chukhno, O. Galinina, S. Andreev, Y. Gaidamaka, K. Samouylov, and G. Araniti. 2022. A holistic assessment of directional deafness in mmWave-based distributed 3D networks. *IEEE T. Wirel. Commun.* 21(9):7491–7505. doi: 10.1109/TWC.2022.3159086.

Received March 14, 2024

Contributors

Ostrikova Daria Yu. (b. 1988) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; ostrikova-dyu@rudn.ru

Golos Elizaveta S. (b. 1998) — PhD student, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; 1142210130@rudn.ru

Beschastnyi Vitalii A. (b. 1992) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; beschastnyy-va@rudn.ru

Machnev Egor A. (b. 1996) — PhD student, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; 1042200071@rudn.ru

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

Gaidamaka Yuliya V. (b. 1971) — Doctor of Science in physics and mathematics, professor, Department of Probability Theory and Cyber Security, RUDN University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; gaydamaka-yuv@rudn.ru

ОБ ОДНОПОРОГОВОМ УПРАВЛЕНИИ ОЧЕРЕДЬЮ В СИСТЕМЕ МАССОВОГО ОБСЛУЖИВАНИЯ С НЕТЕРПЕЛИВЫМИ ЗАЯВКАМИ

Я. М. Агаларов[1]

Аннотация: Изложены результаты теоретического исследования управляемой системы массового обслуживания (СМО) типа $M/M/s$ с нетерпеливыми заявками и однопороговым управлением очередью. Ставится задача оптимизации однопорогового управления очередью, суть которой заключается в вычислении для длины очереди некоторого порогового значения, максимизирующего заданную целевую функцию. В исследуемой системе заявка покидает систему необслуженной, если время ожидания в очереди (или время обслуживания на приборе) превышает некоторый интервал времени случайной длины, распределенной по показательному закону с заданным параметром. В качестве показателя эффективности управления очередью (целевой функции) используется стоимостная функция, учитывающая потери в единицу времени из-за технического обслуживания системы, отклонения заявок на входе системы, ухода заявок до завершения обслуживания. Предложены метод решения задачи максимизации стоимостной целевой функции на множестве однопороговых управлений очередью и алгоритм гарантированного вычисления оптимального порога.

Ключевые слова: система массового обслуживания; нетерпеливые заявки; управление очередью

DOI: 10.14357/19922264240206 **EDN:** JZHAKU

1 Введение

Настоящая работа служит продолжением исследований, посвященных проблеме оптимизации порогового управления очередью в СМО с учетом стоимостных потерь из-за отклонения и задержек заявок, а также затрат на техническое обслуживание системы. Суть порогового управления очередью заключается в том, что для длины очереди задается одно или несколько пороговых значений, по достижении каждого из которых принимается соответствующее решение по сбросу нагрузки из очереди с целью повышения эффективности работы системы [1].

Ниже будем рассматривать оптимизационную задачу управления очередью для простейшей СМО, у которой ограничено время пребывания заявки в очереди или на приборе. Заявка покидает систему необслуженной, если время ожидания в очереди или на приборе превышает некоторую случайную величину с заданным средним значением. В простейшей модели системы такого типа предполагают, что заявки покидают очередь через случайные интервалы времени, распределенные по показательному закону, т. е. возникает поток уходящих из очереди с постоянной интенсивностью заявок. Таким образом, каждая заявка, находящаяся в очереди или на приборе, может покинуть систему, не дождавшись обслуживания, через случайный интервал времени, распределенный по показательному закону. Заявки в этом случае называют «нетерпеливыми», а СМО — системой с «нетерпеливыми» заявками. Такая СМО имеет четыре потока, влияющих на состояние системы: входной поток заявок, поток обслуженных заявок, поток заявок, покидающих очередь, не дождавшись начала обслуживания, и поток уходящих с приборов заявок, не дождавшихся завершения обслуживания. Так как поток уходящих заявок пуассоновский, то процесс, протекающий в системе под влиянием такого потока, будет марковским.

С увеличением порога длины очереди, с одной стороны, увеличивается поток заявок — потенциальных плательщиков за обслуживание, с другой — увеличиваются потери системы (из-за увеличения задержек заявок, ухода «нетерпеливых» заявок, затрат системы на хранение и обслуживание заявок). Возникает задача поиска значения порога длины очереди, максимизирующего доход системы.

Результаты теоретических и экспериментальных исследований по рассматриваемой в данной статье проблеме, изложенные в ранее опубликованных работах, получены для задачи оптимизации порогового управления очередью в одноканальных и многоканальных СМО с терпеливыми заявками (заявки не покидают систему до завершения обслуживания) (см., например, [2–8]). При исследовании СМО с управляемыми очередями методами математического моделирования, как правило, требует-

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, agglar@yandex.ru

ся предварительно решить подзадачу расчета вероятностно-временны́х характеристик (показателей) исследуемой системы [9–13] и использовать данную расчетную модель для разработки и исследования алгоритма управления очередью (очередями), что приводит к математической модели с более сложными функциональными взаимозависимостями параметров системы по сравнению с расчетной моделью. В научных публикациях, посвященных исследованию систем с нетерпеливыми заявками, отсутствуют результаты по оптимизации управления очередью, в основном в них рассмотрены задачи по расчету вероятностно-временны́х характеристик и оптимизации структуры таких систем [9, 10].

Ниже приводим метод и результаты теоретического исследования однопорогового управления очередью системы $M/M/s$ с «нетерпеливыми» заявками при стоимостном критерии оптимальности.

2 Постановка задачи

Рассматривается многоканальная СМО $M/M/s$ с управляемой очередью, в которой заявки, не дожидаясь завершения обслуживания, могут покинуть систему по истечении некоторого времени пребывания в очереди или на приборе. Предполагается, что время, через которое заявка покидает систему, — показательно распределенная случайная величина, при этом параметр распределения равен α_i ($\alpha_1 \le \alpha_2 \le \cdots$), если заявка i-я в очереди, а если на приборе, то параметр равен β. Поступившая извне заявка допускается в систему, если длина очереди в системе меньше, чем $h \ge 0$ — некоторая заданная величина (пороговое значение), иначе отклоняется и теряется. Допущенная в систему заявка занимает любой из свободных приборов, если такой есть, иначе становится в конец очереди. Будем считать, что заявки обслуживаются в порядке поступления. Отметим, что поведение такой системы описывается цепью Маркова, в которой состоянием считается число заявок в системе [9].

Введем обозначения:

λ — интенсивность входного потока;

μ — интенсивность обслуживания заявки на приборе;

s — число приборов в системе;

$h + s$ — объем накопителя;

C_0 — плата заявки, принятой в накопитель системы;

C_1 — стоимость потерь из-за отклонения заявки на входе системы;

C_2 — стоимость потерь из-за ухода i-й заявки, находящейся в очереди;

C_3 — стоимость потерь из-за ухода с прибора заявки, не дождавшейся завершения обслуживания;

$\pi_i^{(h)}$ — стационарная вероятность состояния i системы при пороговом значении h;

$\overline{S}^{(h)} = s - \sum_{i=1}^{S}(s-i)\pi_i^{(h)}$ — среднее число занятых приборов;

$Q(h)$ — доход системы в единицу времени при пороговом значении h.

В качестве целевой функции задачи оптимизации порогового управления рассматривается предельный средний доход системы в единицу времени, вычисляемый по формуле:

$$Q(h) = \lambda C_0\left(1 - \pi_{s+h}^{(h)}\right) - \lambda C_1 \pi_{s+h}^{(h)} - \\ - d\left(\overline{L}^{(h)}\right) - \beta C_3 \overline{S}^{(h)}. \quad (1)$$

Здесь $\lambda C_0(1 - \pi_{s+h}^{(h)})$ — средняя суммарная плата заявок, принимаемых в накопитель в единицу времени; $\lambda C_1 \pi_{s+h}^{(h)}$ — средние потери в единицу времени из-за отклонения заявок; $d(\overline{L}^{(h)})$ — средние потери в единицу времени из-за ухода заявок из очереди:

$$d\left(\overline{L}^{(h)}\right) = C_2 \sum_{i=1}^{h}\sum_{j=1}^{i} \alpha_j \pi_{i+S}^{(h)};$$

$\beta C_3 \overline{S}^{(h)}$ — средние потери в единицу времени из-за ухода с приборов заявок, не дождавшихся завершения обслуживания.

Ставится задача оптимизации порогового значения длины очереди, т. е. математическая задача вида

$$h^* = \arg\max_{0 \le h} Q(h)\,. \quad (2)$$

3 Метод решения и результаты

Для стационарных вероятностей состояний описанной в предыдущем разделе СМО справедливы равенства [9]:

$$\left.\begin{aligned} \pi_l^{(h)} &= \frac{\rho^l}{l!(1+\gamma)^i}\,\pi_0^{(h)} \text{ при } l \in \overline{1,s}\,, \\ \pi_{s+l}^{(h)} &= \frac{\rho^s}{s!(1+\gamma)^3}\prod_{j=1}^{l}\frac{\rho}{s(1+\gamma)+j\theta_j}\,\pi_0^{(h)} \\ &\qquad \text{при } l \in \overline{1,h}\,, \end{aligned}\right\} \quad (3)$$

где

$$\pi_0^{(h)} = \left[1 + \sum_{m=1}^{s} \frac{\rho^m}{m!(1+\gamma)^m} + \right.$$

$$\left. + \frac{\rho^s}{s!(1+\gamma)^s} \sum_{l=1}^{h} \prod_{j=1}^{l} \frac{\rho}{s(1+\gamma)+j\theta_j}\right]^{-1};$$

$$\rho = \frac{\lambda}{\mu}; \quad \gamma = \frac{\beta}{\mu}; \quad \theta_j = \frac{\alpha_j}{\mu}.$$

Покажем, что для стационарных вероятностей справедливы соотношения

$$\pi_l^{(h+1)} = \left(1 - \pi_{s+h+1}^{(h+1)}\right)\pi_l^{(h)},\ l = \overline{0, s+h}; \qquad (4)$$

$$\pi_{s+h+1}^{(h+1)} = \left(1 - \pi_{s+h+1}^{(h+1)}\right)\pi_{s+h}^{(h)} \frac{\rho}{s(1+\gamma)+(h+1)\theta_{h+1}}. \qquad (5)$$

Из (3) при $l = \overline{1, h}$ следует

$$\pi_{s+l}^{(h)} - \pi_{s+l}^{(h+1)} = \left(\frac{\rho^s}{s!(1+\gamma)^s}\right)^2 \prod_{j=1}^{l} \frac{\rho}{s(1+\gamma)+j\theta_j} \times$$

$$\times \prod_{j=1}^{h+1} \frac{\rho}{s(1+\gamma)+j\theta_j} \pi_0^{(h)}\pi_0^{(h+1)} = \pi_{s+l}^{(h)}\pi_{s+h+1}^{(h+1)}.$$

Точно так же, использовав (3) при $l = \overline{0, s}$, получаем равенство

$$\pi_l^{(h)} - \pi_l^{(h+1)} = \pi_l^{(h)}\pi_{s+h+1}^{(h+1)}.$$

Следовательно, равенства (4) справедливы. Аналогично, использовав (3) и (4), находим

$$\pi_{s+h+1}^{(h+1)} = \frac{\rho^s}{s!(1+\gamma)^2} \prod_{j=1}^{h+1} \frac{\rho}{s(1+\gamma)+j\theta_j} \pi_0^{(h+1)} =$$

$$= \frac{\rho^s}{s!(1+\gamma)^s} \frac{\rho}{s(1+\gamma)+(h+1)\theta_{h+1}} \times$$

$$\times \prod_{j=1}^{h} \frac{\rho}{s(1+\gamma)+j\theta_j} \left(1 - \pi_{s+h+1}^{(h+1)}\right)\pi_0^{(h)},$$

откуда следует (5).

Покажем, что имеет место равенство

$$Q(h) - Q(h+1) = \pi_{s+h+1}^{(h+1)} \left[Q(h) - G(h)\right], \qquad (6)$$

где

$$G(h) = (C_0 + C_1)(\mu+\beta)s + (C_0 + C_1)\alpha_{h+1}(h+1) - \sum_{j=1}^{h+1} \alpha_j C_2 - C_1\lambda - C_3\beta s. \qquad (7)$$

Использовав (1)–(7), получим:

$$Q(h) - Q(h+1) = \lambda C_0\left(1 - \pi_{s+h}^{(h)}\right) - \lambda C_1\pi_{s+h}^{(h)} - d\left(\overline{L}^{(h)}\right) - \beta C_3\overline{S}^{(h)} -$$

$$- \lambda C_0\left(1 - \pi_{s+h+1}^{(h+1)}\right) + \lambda C_1\pi_{s+h+1}^{(h+1)} + d\left(\overline{L}^{(h+1)}\right) +$$

$$+ \beta C_3\overline{S}^{(h+1)} = -\lambda C_0\pi_{s+h}^{(h)} - \lambda C_1\pi_{s+h}^{(h)} - d\left(\overline{L}^{(h)}\right) -$$

$$- \beta C_3\overline{S}^{(h)} + \lambda C_0\pi_{s+h+1}^{(h+1)} + \lambda C_1\pi_{s+h+1}^{(h+1)} +$$

$$+ \left(1 - \pi_{s+h+1}^{(h+1)}\right) d\left(\overline{L}^{(h)}\right) + C_2 \sum_{j=1}^{h+1} \alpha_j \pi_{s+h+1}^{(h+1)} +$$

$$+ \beta s C_3\pi_{s+h+1}^{(h+1)} + \beta C_3\left(1 - \pi_{s+h+1}^{(h+1)}\right)\overline{S}^{(h)} =$$

$$= -\lambda C_0\pi_{s+h}^{(h)} - \lambda C_1\pi_{s+h}^{(h)} + \lambda C_0\pi_{s+h+1}^{(h+1)} +$$

$$+ \lambda C_1\pi_{s+h+1}^{(h+1)} + C_2 \sum_{j=1}^{h+1} \alpha_j \pi_{s+h+1}^{(h+1)} + \beta s C_3\pi_{s+h+1}^{(h+1)} -$$

$$- \pi_{s+h+1}^{(h+1)} d\left(\overline{L}^{(h)}\right) - \beta C_3\pi_{s+h+1}^{(h+1)}\overline{S}^{(h)} =$$

$$= \pi_{s+h+1}^{(h+1)}\left[\lambda C_0 + \lambda C_1 + C_2 \sum_{j=1}^{h+1} \alpha_j - d\left(\overline{L}^{(h)}\right) - C_3\beta\overline{S}^{(h)} + \beta s C_3 - \lambda\frac{C_0 + C_1}{\pi_{s+h+1}^{(h+1)}}\pi_{s+h}^{(h)}\right] =$$

$$= \pi_{s+h+1}^{(h+1)}\left[\lambda C_0\left(1 - \pi_{s+h}^{(h)}\right) - \lambda C_1\pi_{s+h}^{(h)} + \lambda C_1 + \sum_{j=1}^{h+1} \alpha_j C_2 - d\left(\overline{L}^{(h)}\right) - C_3\beta\overline{S}^{(h)} + \lambda C_1\pi_{s+h}^{(h)} + \lambda C_0\pi_{s+h}^{(h)} + \beta s C_3 - \lambda\frac{C_0 + C_1}{\pi_{s+h+1}^{(h+1)}}\pi_{s+h}^{(h)}\right] =$$

$$= \pi_{s+h+1}^{(h+1)}\left[Q(h) + \lambda C_1 + \sum_{j=1}^{h+1} \alpha_j C_2 + \beta s C_3 - (C_0 + C_1)\lambda\frac{s(1+\gamma)+(h+1)\theta}{\rho}\right] =$$

$$= \pi_{s+h+1}^{(h+1)}\left[Q(h) + \lambda C_1 + C_2 \sum_{j=1}^{h+1} \alpha_j + \beta s C_3 - (C_0 + C_1)[s(\mu+\beta) + \alpha_{h+1}(h+1)]\right].$$

Значит, равенство (6) имеет место.

Так как верно равенство

$$\lambda\left(1-\pi_{s+h}^{(h)}\right)=\sum_{i=1}^{h}\sum_{j=1}^{i}\alpha_j\pi_i^{(h)}+(\beta+\mu)\overline{S}^{(h)},$$

то равенства для $Q(h)$ и $G(h)$ в (1) и (7) эквивалентны равенствам

$$Q(h)=[(C_0+C_1)(\beta+\mu)-\beta C_3]\,\overline{S}^{(h)}+$$
$$+(C_0+C_1-C_2)\sum_{i=1}^{h}\sum_{j=1}^{i}\alpha_j\pi_i^{(h)}-C_1\lambda;$$
$$G(h)=[(C_0+C_1)(\beta+\mu)-\beta C_3]\,s+$$
$$+(C_0+C_1)\,\alpha_{h+1}(h+1)-C_2\sum_{j=1}^{h+1}\alpha_j-C_1\lambda.$$

При $h=0$ последние равенства примут вид:

$$Q(0)=[(C_0+C_1(\,(\beta+\mu)-\beta C_3]\,\overline{S}^{(0)}-C_1\lambda;$$
$$G(0)=[(C_0+C_1)(\beta+\mu)-\beta C_3]\,s+$$
$$+(C_0+C_1-C_2)\,\alpha_1-C_1\lambda.$$

Далее всюду будем предполагать, что при условии $C_0+C_1-C_2<0$ выполняется и условие $(C_2/(C_0+C_1)-1)\alpha_{i+1}/(\alpha_{i+1}-\alpha_i)\geq i$ для всех $i\geq 1$. Обратим внимание, что функция $G(h)$ возрастает по h при $C_0+C_1-C_2>0$ и не возрастает, когда $C_0+C_1-C_2\leq 0$ и α_i такие, что условие $(C_2/(C_0+C_1)-1)\alpha_{i+1}/(\alpha_{i+1}-\alpha_i)\geq i$ для всех $i\geq 1$.

Воспользуемся теоремой 1 из работы [14]. Нетрудно заметить (см. равенство (6)), что функция $Q(h)$ при $C_0+C_1\leq C_2$ и функция $-Q(h)$ при $C_0+C_1>C_2$ удовлетворяют условиям теоремы 1 из [14]. Тогда из указанной теоремы непосредственно следует справедливость следующего утверждения.

Утверждение. *При выполнении предположения, введенного выше относительно параметров C_0, C_1, C_2 и α_i, $i\geq 1$, решение задачи* (2) *обладает следующими свойствами*:

(1) *если $C_0+C_1\leq C_2$, то $Q(h)$ — унимодальная функция (так как $G(h)$ не возрастает по h и $G(0)\leq Q(0)$), и если $[(C_0+C_1)(\beta+\mu)-\beta C_3](s-\overline{S}^{(0)})\leq(C_0+C_1-C_2)\alpha_1$, то $h^*=0$, иначе существует $0<h^*<\infty$;*

(2) *если $C_0+C_1>C_2$ и $[(C_0+C_1)(\beta+\mu)-\beta C_3](s-\overline{S}^{(0)})+(C_0+C_1-C_2)\alpha_1>0$, то $h^*=\infty$ и при этом $Q(h)$ монотонно возрастает по h (так как $G(h)$ возрастает по h и $G(0)>Q(0)$);*

(3) *если $C_0+C_1>C_2$ и $[(C_0+C_1)(\beta+\mu)-\beta C_3](s-\overline{S}^{(0)})+(C_0+C_1-C_2)\alpha_1\leq 0$, то функция $-Q(h)$ унимодальная (так как удовлетворяет условиям теоремы* 1 *из* [14]*) и при этом*

$$h^*=\begin{cases}0, & \text{если } Q(\infty)\leq Q(0);\\ \infty, & \text{если } Q(\infty)>Q(0)\end{cases}$$

(*так как $G(h)$ возрастает по h и $G(0)\leq Q(0)$*).

На рис. 1 и 2 проиллюстрировано поведение функций $Q(h)$ и $G(h)$ для двух наборов значений параметров рассматриваемой СМО:

(1) рис. 1: $\lambda=8$; $\mu=2$; $\alpha_i=0{,}5$; $i=\overline{1,h}$; $\beta=0{,}25$; $C_0=20$; $C_1=5$; $C_2=40$; $C_3=10$;

(2) рис. 2: $\lambda=8$; $\mu=1$; $\alpha_i=0{,}25$; $i=\overline{1,h}$; $\beta=0{,}125$; $C_0=20$; $C_1=3$; $C_2=10$; $C_3=15$.

Заметим, что в случае, изображенном на рис. 1, целевая функция достигает максимума при пороговом значении $h^*=18$, что согласуется с утверждением пункта 1, а в случае рис. 2 оптимальное значение порога $h^*=\infty$, что соответствует пункту 2.

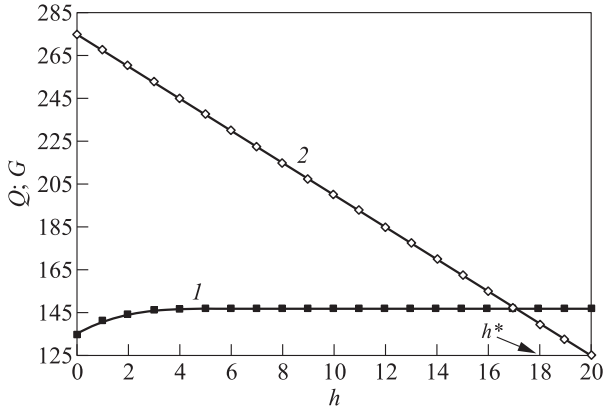


Рис. 1 Зависимости функций Q (*1*) и G (*2*) от порогового значения h: h^* — оптимальное пороговое значение длины очереди

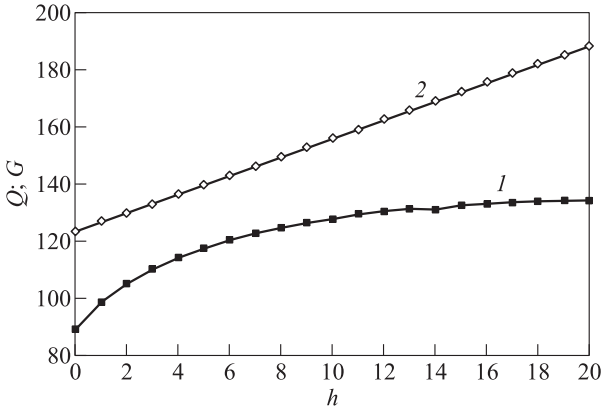


Рис. 2 Зависимости функций Q (*1*) и G (*2*) от порогового значения h

4 Заключение

Практическим результатом проведенных выше исследований стал следующий простой алгоритм оптимизации однопорогового управления очередью для рассмотренной выше модели СМО при выполнении условий утверждения относительно параметров C_0, C_1, C_2 и α_i, $i \geq 1$.

1. Если выполняется условие $C_0 + C_1 \leq C_2$, то
   (1) положить $h = 0$;
   (2) до тех пор пока выполняется неравенство $Q(h+1) > Q(h)$, полагать $h = h + 1$;
   (3) положить $h^* = h$.
2. Если выполняются неравенства $C_0 + C_1 > C_2$ и $[(C_0 + C_1)(\beta + \mu) - \beta C_3](s - \overline{S}^{(0)}) + (C_0 + C_1 - C_2)\alpha_1 > 0$, то положить $h^* = \infty$.
3. Если $C_0 + C_1 > C_2$ и $[(C_0 + C_1)(\beta + \mu) - \beta C_3](s - \overline{S}^{(0)}) + (C_0 + C_1 - C_2)\alpha_1 \leq 0$, то положить

$$h^* = \begin{cases} 0, & \text{если } Q(\infty) \leq Q(0); \\ \infty & \text{иначе.} \end{cases}$$

Обратим внимание, что при выполнении условий третьего пункта алгоритма справедливы неравенства $Q(0) \leq 0$ и $G(0) \leq Q(0)$ и на отрезке $[0, h^0]$, где h^0 — максимальное значение, такое что $Q(h^0) \geq G(h^0)$, функция $Q(h)$ не возрастает, а при $h \in [h^0, \infty)$ возрастает (так как в случае пункта 3 функция $-Q(h)$ унимодальная). Следовательно, если $C_0 + C_1 \geq C_2$ и выполняется условие $Q(0) \leq Q(1)$, то $h^* = \infty$.

Литература

1. *Floyd S., Jacobson V.* Random early detection gateways for congestion avoidance // IEEE ACM T. Network., 1993. Vol. 1. P. 397–413. doi: 10.1109/90.251892.
2. *Коновалов М. Г.* Об одной задаче оптимального управления нагрузкой на сервер // Информатика и её применения, 2013. Т. 7. Вып. 4. С. 34–43. doi: 10.14357/19922264130404. EDN: RRROXB.
3. *Konovalov M. G., Razumchik R. V.* Comparison of two active queue management schemes through the $M/D/1/N$ queue // Информатика и её применения, 2018. Т. 12. Вып. 4. С. 9–15. doi: 10.14357/19922264180402. EDN: VOGJOZ.
4. *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. EDN: YQHHGP.
5. *Агаларов Я. М., Ушаков В. Г.* Об унимодальности функции дохода системы массового обслуживания типа $G/M/s$ с управляемой очередью // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 55–61. doi: 10.14357/19922264190108. EDN: HYAODW.
6. *Коновалов М. Г., Разумчик Р. В.* Комплексное управление в одном классе систем с параллельным обслуживанием // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 54–59. doi: 10.14357/19922264190409. EDN: REESRH.
7. *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. EDN: FCYDUT.
8. *Агаларов Я. М.* Оптимизация схемы распределения буферной памяти узла пакетной коммутации // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 39–48. doi: 10.14357/19922264230306. EDN: XQLXCKV.
9. *Кирпичников Ф. П., Флакс Д. Б., Галямова К. Н.* Средняя длина очереди в системе массового обслуживания с ограниченным средним временем пребывания заявки в системе // Вестник Технологического университета, 2017. Т. 20. № 2. С. 81–84. EDN: XVFSTN.
10. *Савинов Ю. Г., Табакова Е. Д., Сафиуллов И. Д.* Оптимизация в СМО с нетерпеливыми заявками // Ученые записки УлГУ. Сер. Математика и информационные технологии, 2019. № 1. С. 92–98. EDN: OWOZYR.
11. *Мейханаджян Л. А., Разумчик Р. В.* Система массового обслуживания $\mathrm{Geo}/G/1/\infty$ с инверсионным порядком обслуживания и ресамплингом в дискретном времени // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 60–67. doi: 10.14357/19922264190410. EDN: LNIHGC.
12. *Милованова Т. А., Разумчик Р. В.* Однолинейная система массового обслуживания с инверсионным порядком обслуживания с вероятностным приоритетом, групповым пуассоновским потоком и фоновыми заявками // Информатика и её применения, 2020. Т. 14. Вып. 3. С. 26–34. doi: 10.14357/19922264200304. EDN: NOMSAM.
13. *Берговин А. К., Ушаков В. Г.* Исследование систем обслуживания со смешанными приоритетами // Информатика и её применения, 2023. Т. 17. Вып. 2. С. 57–61. doi: 10.14357/19922264230208. EDN: JULPWS.
14. *Агаларов Я. М.* Признак унимодальности целочисленной функции одной переменной // Обозрение прикладной и промышленной математики, 2019. Т. 26. Вып. 1. С. 65–66.

Поступила в редакцию 23.02.24

ON SINGLE-THRESHOLD QUEUE MANAGEMENT IN A QUEUING SYSTEM WITH IMPATIENT CUSTOMERS

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 results of a theoretical study of a managed queuing system of $M/M/k$ type with impatient customers and single-threshold queue management are presented. The task of optimizing single-threshold queue management is set, the essence of which is to calculate for the queue length a certain threshold value that maximizes a given objective function. In the system under study, a customer leaves the system unattended if the waiting time in the queue (or the service time on the device) exceeds a certain time interval of random length distributed according to an exponential law with a given parameter. A cost function is used as an indicator of the effectiveness of queue management (objective function) which takes into account the losses per unit of time due to system technical maintenance, rejection of customers at the entrance of the system, and leaving of customers until the end of the service. A method for solving the problem of maximizing the cost objective function on a set of single-threshold queue controls and an algorithm for guaranteed calculation of the optimal threshold are proposed.

Keywords: queuing system; impatient customers; queue management

DOI: 10.14357/19922264240206 **EDN:** JZHAKU

References

1. Floyd, S., and V. Jacobson. 1993. Random early detection gateways for congestion avoidance. *IEEE ACM T. Network*. 1:397–413. doi: 10.1109/90.251892.
2. Konovalov, M. G. 2013. Ob odnoy zadache optimal'nogo upravleniya nagruzkoy na server [About one task of overload control]. *Informatika i ee Primeneniya — Inform. Appl.* 7(4):34–43. doi: 10.14357/19922264130404. EDN: RRROXB.
3. Konovalov, M., and R. Razumchik. 2018. Comparison of two active queue management schemes through the $M/D/1/N$ queue. *Informatika i ee Primeneniya — Inform. Appl.* 12(4):9–15. doi: 10.14357/19922264180402. EDN: VOGJOZ.
4. 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. EDN: YQHHGP.
5. Agalarov, Ya. M., and V. G. Ushakov. 2019. Ob unimodal'nosti funktsii dokhoda sistemy massovogo obsluzhivaniya tipa $G/M/s$ s upravlyaemoy ochered'yu [On the unimodality of the income function of a type $G/M/s$ queueing system with controlled queue]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):55–61. doi: 10.14357/19922264190108. EDN: HYAODW.
6. Konovalov, M. G., and R. V. Razumchik. 2019. Kompleksnoe upravlenie v odnom klasse sistem s parallel'nym obsluzhivaniem [Mixed policies for online job allocation in one class of systems with parallel service]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):54–59. doi: 10.14357/19922264190409. EDN: REESRH.
7. 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. EDN: FCYDUT.
8. Agalarov, Ya. M. 2023. Optimizatsiya skhemy raspredeleniya bufernoy pamyati uzla paketnoy kommutatsii [Optimization of the buffer memory allocation scheme of the packet switching node]. *Informatika i ee Primeneniya — Inform. Appl.* 17(3):39–48. doi: 10.14357/19922264230306. EDN: QLXCKV.
9. Kirpichnikov, F. P., D. B. Flaks, and K. N. Galyamova. 2017. Srednyaya dlina ocheredi v sisteme massovogo obsluzhivaniya s ogranichennym srednim vremenem prebyvaniya zayavki v sisteme [The average queue length in a queuing system with a limited average time for the request to stay in the system]. *Vestnik Tekhnologicheskogo universiteta* [Bulletin of Technological University] 20(2):81–84. EDN: XVFSTN.
10. Savinov, Yu. G., E. D. Tabakova, and I. D. Safiullov. 2019. Optimizatsiya v SMO s neterpelivymi zayavkami [Optimization in the queuing system with impatient customers]. *Uchenyye zapiski UlGU. Ser. Matematika i informatsionnye tekhnologii* [Scientific Notes of UlSU. Ser. Mathematics and Information Technology] 1:92–98. EDN: OWOZYR.
11. Meykhanadzhyan, L. A., and R. V. Razumchik. 2019. Sistema massovogo obsluzhivaniya $\mathrm{Geo}/G/1/\infty$ s inversionnym poryadkom obsluzhivaniya i resamplingom v diskretnom vremeni [Discrete-time $\mathrm{Geo}/G/1/\infty$ LIFO queue with resampling policy]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):60–67. doi: 10.14357/19922264190410. EDN: LNIHGC.
12. Milovanova, T. A., and R. V. Razumchik. 2020. Odnolineynaya sistema massovogo obsluzhivaniya s inversionnym

poryadkom obsluzhivaniya s veroyatnostnym prioritetom, gruppovym puassonovskim potokom i fonovymi zayavkami [A single-server queueing system with LIFO service, probabilistic priority, batch Poisson arrivals, and background customers]. *Informatika i ee Primeneniya — Inform. Appl.* 14(3):26–34. doi: 10.14357/ 19922264200304. EDN: NOMSAM.

13. Bergovin, A. K., and V. G. Ushakov. 2023. Issledovanie sistem obsluzhivaniya so smeshannymi prioritetami [Analysis of the queueing systems with mixed priorities]. *Informatika i ee Primeneniya — Inform. Appl.* 17(2):57–61. doi: 10.14357/19922264230208. EDN: JULPWS.
14. 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 February 23, 2024

Contributor

Agalarov Yaver M. (b. 1952) — Candidate of Science (PhD) in technology, associate professor, leading scientist, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; agglar@yandex.ru

О ПОРОЖДЕНИИ СИНТЕТИЧЕСКИХ ПРИЗНАКОВ НА ОСНОВЕ ОПОРНЫХ ЦЕПЕЙ И ПРОИЗВОЛЬНЫХ МЕТРИК В РАМКАХ ТОПОЛОГИЧЕСКОГО ПОДХОДА К АНАЛИЗУ ДАННЫХ. ЧАСТЬ 2. ЭКСПЕРИМЕНТАЛЬНАЯ АПРОБАЦИЯ НА ЗАДАЧАХ ФАРМАКОИНФОРМАТИКИ*

И. Ю. Торшин[1]

Аннотация: Рассмотрение прецедентных отношений между признаками и таргетной переменной в виде наборов элементов булевой решетки указывает на возможность порождения синтетических признаков с использованием метрических функций расстояния. Сформулированы подходы к (1) оценке релевантности («информативности») метрик по отношению к решаемым задачам, (2) порождению и (3) отбору синтетических признаков, более информативных, чем исходные признаковые описания. Представленные результаты топологического анализа 2400 выборок данных «молекула–свойство» из ProteomicsDB позволили получить достаточно эффективные алгоритмы прогнозирования свойств молекул (ранговая корреляция в кросс-валидации — $0{,}90 \pm 0{,}23$). На данной выборке задач установлены метрики, которые наиболее часто порождают информативные синтетические признаки: максимальное уклонение Колмогорова, «косое» расстояние, метрики Lp, Реньи, фон Мизеса. Для решения изученного комплекса задач показано преимущество полиномных корректоров по сравнению с нейросетевыми и с корректорами типа «случайный лес».

Ключевые слова: топологический анализ данных; теория решеток; алгебраический подход Ю. И. Журавлёва; фармакоинформатика

DOI: 10.14357/19922264240207 **EDN:** OTXCUD

1 Введение

В первой части работы [1] принимается, что задано регулярное множество прецедентов

$$\mathbf{Q} = \{\mathrm{D}(x_i)|x_i \in \mathbf{X}\}$$

на решетке $L(T(\mathbf{X}))$, порожденное на основе множества исходных описаний объектов $\mathbf{X} = \{x_1, \ldots, x_{N_0}\}$. Для индивидуального объекта $x_i \in \mathbf{X}$ прецедентному соотношению между значениями признаками $\Gamma_k(x_i)$ и t-й таргетной переменной соответствует множество пар $\{(\{\Gamma_k^{-1}(\Gamma_k(x_i)), k = \overline{1,,n}\}, \Gamma_t^{-1}(\Gamma_t(x_i))), i = \overline{1,N_0}, \ k = \overline{1,n}, t = \overline{n+1,n+l}\}$, где l — число таргетных переменных. В рамках топологической теории распознавания прецедентное соотношение между множествами $\{\Gamma_k^{-1}(\Gamma_k(x_i))\}$ и $\Gamma_t^{-1}(\Gamma_t(x_i))$ моделируется как соответствующие массивы расстояний, порождаемые той или иной метрикой ρ_m: $L(T(\mathbf{X}))^2 \to [0\ldots1]$, $m = \overline{1,m_0}$. В [1] предложены способы «встраивания» в формализм полуэмирических расстояний на множествах $a \in L(T(\mathbf{X}))$, векторах $\vec{v}_\alpha[a] = (v_{\alpha_1}[a], v_{\alpha_2}[a], \ldots, v_{\alpha_i}[a], \ldots)$ и функциях $\hat{\phi}(x)\mathbf{\Gamma}_t(u)$.

Здесь для практического приложения формализма сформулированы подходы к исследованию свойств ρ_m, способы оценки релевантности функций ρ_m по отношению к решаемым задачам, способы порождения и отбора синтетических признаков, основанных на ρ_m. Представлены результаты экспериментальной апробации на задачах фармакоинформатики.

2 Об исследовании свойств функций расстояния ρ_m

Рабочая гипотеза настоящего исследования состоит в том, что для порождения более «информативных» признаков могут использоваться полуэмпирические функционалы расстояния на множествах, векторах, функциях [2]. Метрические свойства используемых функций расстояния ρ_m могут исследоваться аналитически или комбина-

*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).

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, tiy135@yahoo.com

торно с использованием аксиом метрики [3]. Для анализа свойств этих функционалов в топологической теории распознавания вводится следующее понятие.

Определение 1. Обобщенной оценочной функцией расстояния будем называть конструкцию вида

$$\rho(a,b) = f(g(v[a \vee b]) - g(v[a \wedge b])),$$

в которой f и g — функции, монотонные на соответствующих участках действительной оси; $v : L \to \to R^+$ — изотонная оценка, для которой выполнено условие оценки (**уО**: $\forall_L a, b : v[a] + v[b] = v[a \wedge b] + + v[a \vee b]$) и изотонности (**уИ**: $\forall_L a, b : a \supseteq b \Rightarrow v[a] \geq \geq v[b]$).

Теорема 1. *Функция расстояния ρ считается обобщенной оценочной функцией расстояния тогда и только тогда, когда $\rho(a,b) = \rho(a \vee b, a \wedge b)$, а термы от a и b в формуле для $\rho(a,b)$ представляют собой композицию монотонной функции и изотонной оценки.*

Необходимость следует из $a \vee b = (a \vee b) \vee (a \wedge b)$ и $a \wedge b = (a \vee b) \wedge (a \wedge b)$ при подстановке $a \vee b$ и $a \wedge b$ вместо a и b в определение 1. Эквивалентность $\rho(a,b)$ и $\rho(a \vee b, a \wedge b)$ указывает на то, что в выражение для вычисления ρ входят термы-функционалы, содержащие выражения $a \vee b$ и $a \wedge b$, взаимозаменяемые с a и b, т. е. термы вида $g'(a \vee b)$ и $g'(a \wedge b)$. По условию теоремы эти термы включают монотонную функцию от изотонной оценки, т. е. g' монотонна. Так как ρ — функция расстояния, то g'-термы не могут входить в выражение для ρ в виде произведения, суммы, отношения, степени или суммы, а только в виде разности, т. е.

$$\rho(a,b) = f\left(g'(a \vee b) - g'(a \wedge b)\right),$$

из чего следует достаточность. Теорема доказана.

Следствие 1. Для обобщенной оценочной ρ

$$\forall \ell \subseteq L(T(\mathbf{X})) : \Delta_{\vee\wedge}(\ell) \equiv 0,$$

$$\Delta_{\vee\wedge}(\ell) = \sum_{a,b \in \ell} |\rho(a,b) - \rho(a \vee b, a \wedge b)| \frac{2}{|\ell|/(|\ell|-1)}.$$

Следствие 2. Выберем «опорное» множество $a \in \in L(T(\mathbf{X}))$ и обобщенную оценочную ρ. При $f(x) = = g(x) = x$ $v_{a,\rho}[b] = \rho(a,b) = \rho(a \vee b, a \wedge b)$ — изотонная оценка.

Следует из того, что любая линейная комбинация изотонных оценок — изотонная оценка при условии положительной определенности (теорема 2 в [4]). Также проверяется прямой подстановкой $v_{a,\rho}[b]$ в уО и уИ.

Следствие 3. Расстояния Фреше–Никодима, Амана, Рэнда/Щекановского, Сокала–Сниса (варианты 1, 2 и 3), Рассела–Рао, Роджера–Танимото, Фейта, Тверского и Юле могут служить обобщенными оценочными функциями расстояния.

Следствие 4. Расстояния Симпсона, Брауна–Бланке, Андерберга и Говера-2 не входят в число обобщенных оценочных функций расстояния.

Теорема 1 со следствиями предоставляет аналитический и комбинаторный инструментарий для исследования свойств полуэмпирических функций расстояния. Если заданная ρ служит обобщенной оценочной функцией расстояния, то могут быть получены соответствующие аналитические выражения для функций f и g. Например, расстояние Сокала–Сниса-2

$$\rho(a,b) = 1 - \frac{|a \cap b|}{|a \cup b| + |a \Delta b|}$$

выступает обобщенным оценочным расстоянием с $f(x) = (e^x - 1)/(0{,}5e^x - 1)$ и $g(x) = \ln(x)$. При невозможности аналитической проверки свойства ρ как обобщенной оценочной могут быть изучены на подмножествах ℓ решетки $L(T(\mathbf{X}))$ посредством вычисления значений функционала $\Delta_{\vee\wedge}(\ell)$ (следствие 1).

3 О способах оценки релевантности метрик ρ_m по отношению к задаче классификации/прогнозирования

Биекция между множеством прецедентов $\mathbf{Q}$ и множеством исходных описаний объектов $\mathbf{X}$, существующая при выполнении условия регулярности по Журавлёву ($\forall \mathrm{x} \in \mathbf{X}, \mathrm{x} = D^{-1}(D(\mathrm{x}))$, гарантирует однозначность соответствия описаний x_i и q_i. Это делает возможным рассматривать прецедентные соотношения, заданные на $\mathbf{Q}$, в терминах множеств $\{\Gamma_k^{-1}(\Gamma_k(x_i))\}$ и $\Gamma_t^{-1}(\Gamma_t(x_i))$ с использованием расстояний ρ_m на подмножествах множества $\mathbf{X}$ [1].

Пусть таргетный класс объектов $\mathbf{c}_{\boldsymbol{\alpha}}$ задан посредством α-го значения t-й переменной $\lambda_{t\alpha} \in \mathrm{I}_t$, $t = \overline{n+1, n+l}$, как $\mathbf{c}_{\boldsymbol{\alpha}} = \Gamma_t^{-1}(\lambda_{t\alpha})$. В случае числовой переменной за $\mathbf{c}_{\boldsymbol{\alpha}}$ может приниматься каждый из элементов $u(\lambda_{t\alpha})$ цепи A_t. Так как $L(T(\mathbf{X}))$ булева, то дополнение множества $\mathbf{c}_{\boldsymbol{\alpha}}$, $\overline{\mathbf{c}}_{\boldsymbol{\alpha}} = \mathbf{X} \backslash \Gamma_t^{-1}(\lambda_{t\alpha})$, определено однозначно. Таким образом, выделение класса $\mathbf{c}_{\boldsymbol{\alpha}}$ порождает задачу классификации $\mathbf{c}_{\boldsymbol{\alpha}}/\overline{\mathbf{c}}_{\boldsymbol{\alpha}}$. Любая задача числового прогнозирования может быть сведена к последовательности корректно решаемых задач $\mathbf{c}_{\boldsymbol{\alpha}}/\overline{\mathbf{c}}_{\boldsymbol{\alpha}}$ [5].

Пусть задано подмножество признаков $p \subseteq [1 \dots n]$ и элемент решетки $c \in L(T(\mathbf{X}))$. Определим функцию

$$\boldsymbol{\rho}_{\mathbf{mc}}(x_i, c, p) = \{\rho_m(c, \Gamma_k^{-1}(\Gamma_k(x_i)),\ k \in p)\}.$$

При заданных ρ_m, p, $\mathbf{c}_{\boldsymbol{\alpha}}$ и $\overline{\mathbf{c}}_{\boldsymbol{\alpha}}$ для x_i вычислимы множества расстояний $\boldsymbol{\rho}_{\mathbf{mc}}(x_i, \mathbf{c}_{\boldsymbol{\alpha}}, p)$ и $\boldsymbol{\rho}_{\mathbf{mc}}(x_i, \overline{\mathbf{c}}_{\boldsymbol{\alpha}}, p)$. Обозначим

$$\boldsymbol{\rho}_{\mathbf{m}\boldsymbol{\alpha}}(x_i) = \boldsymbol{\rho}_{\mathbf{mc}}(x_i, \mathbf{c}_{\boldsymbol{\alpha}}, [1 \dots n]);$$
$$\boldsymbol{\rho}_{\mathbf{m}\overline{\boldsymbol{\alpha}}}(x_i) = \boldsymbol{\rho}_{\mathbf{mc}}(x_i, \overline{\mathbf{c}}_{\boldsymbol{\alpha}}, [1 \dots n]).$$

Для $x_i \in \mathbf{X}$ определено множество

$$\boldsymbol{\rho}_{\mathbf{m}}(x_i, p) = \{\rho_{mk_1k_2}(x_i, p) = \rho_m\left(\Gamma_{k_1}^{-1}\left(\Gamma_{k_1}(x_i), \Gamma_{k_2}^{-1}\left(\Gamma_{k_2}(x_i)\right)\right)\right),\ k_1, k_2 \in p,\ k_1 \neq k_2\},\ \boldsymbol{\rho}_{\mathbf{m}}(x_i) = \boldsymbol{\rho}_{\mathbf{m}}(x_i, [1 \dots n]).$$

На основе $\boldsymbol{\rho}_{\mathbf{m}\boldsymbol{\alpha}}(x_i)$ и $\boldsymbol{\rho}_{\mathbf{m}\overline{\boldsymbol{\alpha}}}(x_i)$ вводятся оценки релевантности ρ_m. По отношению к задаче $\mathbf{c}_{\boldsymbol{\alpha}}/\overline{\mathbf{c}}_{\boldsymbol{\alpha}}$ более релевантна или «информативна» такая метрика ρ_m, которая для всех $x \in \mathbf{c}_{\boldsymbol{\alpha}}$ минимизирует расстояния в списке $\boldsymbol{\rho}_{\mathbf{m}\boldsymbol{\alpha}}(x)$ и максимизирует расстояния в списке $\boldsymbol{\rho}_{\mathbf{m}\overline{\boldsymbol{\alpha}}}(x)$ (т. е. «приближает» объекты к их классам). Выделены два взаимосвязанных направления дальнейших исследований:

(1) нахождение подмножеств p признаков, «более информативных» для ρ_m;

(2) настройка/выбор ρ_m при фиксированном p.

Для $c' \in L(T(\mathbf{X}))$ определим $\vartheta_{\mathbf{mc}}$, операцию слияния списков $\boldsymbol{\rho}_{\mathbf{mc}}$:

$$\vartheta_{\mathbf{mc}}(c', c, p) = \bigcup_{y \in c'} \boldsymbol{\rho}_{\mathbf{mc}}(y, c, p).$$

Обозначим

$$\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\mathbf{c}, p) = \vartheta_{\mathbf{mc}}(\mathbf{c}, \mathbf{c}_{\boldsymbol{\alpha}}, p);\ \vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\mathbf{c}, p) = \vartheta_{\mathrm{mc}}(\mathbf{c}, \overline{\mathbf{c}}_{\boldsymbol{\alpha}}, p),$$

вычислим множества $\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\mathbf{c}_{\boldsymbol{\alpha}}, p)$ и $\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\overline{\mathbf{c}}_{\boldsymbol{\alpha}}, p)$ и сформируем эмпирические функции распределения (э.ф.р.) $\hat{\phi}(x)\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\mathbf{c}_{\boldsymbol{\alpha}}, p)$ и $\hat{\phi}(x)\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\overline{\mathbf{c}}_{\boldsymbol{\alpha}}, p)$. На пространстве однородных монотонно возрастающих функций

$$\mathbf{M}_{0\dots1}^{+} = \{f : [0 \dots 1] \to [0 \dots 1],\ x \geq y \Rightarrow f(x) \geq f(y)\}$$

введем функционал расстояния d_f: $\mathbf{M}_{0..1}^{+} \to [0 \dots 1]$ (максимальное уклонение Колмогорова $D(f(x), g(x)) = \sup_x |f(x) - g(x)|$, метрики фон Мизеса, Реньи и др.). Выбор d_f делает возможной постановку ряда задач топологического анализа данных:

(1) количественные оценки релевантности ρ_m как $d_f(\hat{\phi}(x)\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\mathbf{c}_{\boldsymbol{\alpha}}, p), \hat{\phi}(x)\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\overline{\mathbf{c}}_{\boldsymbol{\alpha}}, p))$ для разных $\mathbf{c}_{\boldsymbol{\alpha}}$, $\lambda_{t\alpha} \in \mathrm{I}_t$, $\alpha = \overline{1, |\mathrm{I}_t|}$;

(2) задачи оптимизации для увеличения разделения классов $\mathbf{c}_{\boldsymbol{\alpha}}/\overline{\mathbf{c}}_{\boldsymbol{\alpha}}$ ($\arg\max_{\rho_m, p} d_f(\hat{\phi}\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\overline{\mathbf{c}}_{\boldsymbol{\alpha}}, p), \hat{\phi}\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\mathbf{c}_{\boldsymbol{\alpha}}, p))$, $\arg\max_{\rho_m, p} d_f(\hat{\phi}\vartheta_{\mathbf{m}\overline{\boldsymbol{\alpha}}}, (\overline{\mathbf{c}}_{\boldsymbol{\alpha}}, p), \hat{\phi}\vartheta_{\mathbf{m}\overline{\boldsymbol{\alpha}}}(\mathbf{c}_{\boldsymbol{\alpha}}, p))$ и др.);

(3) определение ρ_q-метрик на пространстве объектов [2, с. 184–199] (например, в виде $d_f(\hat{\phi}\boldsymbol{\rho}_{\mathbf{m}\boldsymbol{\alpha}}(x, p), \hat{\phi}\boldsymbol{\rho}_{\mathbf{m}\boldsymbol{\alpha}}(y, p)), d_f(\hat{\phi}\boldsymbol{\rho}_{\mathbf{m}}(x, p), \hat{\phi}\boldsymbol{\rho}_{\mathbf{m}}(y, p))$);

(4) оценка близости метрик ρ_q к метрике разреза по классам $\mathbf{c}_{\boldsymbol{\alpha}}/\overline{\mathbf{c}}_{\boldsymbol{\alpha}}$;

(5) формулировка критериев разрешимости/регулярности задачи $\mathbf{c}_{\boldsymbol{\alpha}}/\overline{\mathbf{c}}_{\boldsymbol{\alpha}}$ [6];

(6) оценки компактности классов $\mathbf{c}_{\boldsymbol{\alpha}}$ и $\overline{\mathbf{c}}_{\boldsymbol{\alpha}}$ [3].

4 О способах порождения и отбора синтетических признаков на основании функций расстояния

Множества $\boldsymbol{\rho}_{\mathbf{m}\boldsymbol{\alpha}}(x_i, p)$, $\boldsymbol{\rho}_{\mathbf{m}\overline{\boldsymbol{\alpha}}}(x_i, p)$ и $\boldsymbol{\rho}_{\mathbf{m}}(x_i)$ и отдельные $\rho_m(\mathbf{c}_{\boldsymbol{\alpha}}, \Gamma_k^{-1}(\Gamma_k(x_i))$ используются для формирования синтетических числовых признаков $\Gamma_{k'}(x_i)$ объекта x_i, $k' = \overline{n+l+1, n+l+n_S}$. Значение синтетического признака $\Gamma_{k'}(x_i)$ зависит от выбора ρ_m, классов $\mathbf{c}_{\boldsymbol{\alpha}}$ и $\overline{\mathbf{c}}_{\boldsymbol{\alpha}}$ и от способа его вычисления:

(1) $\rho_m(\mathbf{c}_{\boldsymbol{\alpha}}, \Gamma_k^{-1}(\Gamma_k(x_i))$;

(2) $\rho_m(\overline{\mathbf{c}}_{\boldsymbol{\alpha}}, \Gamma_k^{-1}(\Gamma_k(x_i))$;

(3) $\rho_m(\mathbf{c}_{\boldsymbol{\alpha}}, \dots) - \rho_m(\overline{\mathbf{c}}_{\boldsymbol{\alpha}}, \dots)$;

(4) $1 - \rho_m(\mathbf{c}_{\boldsymbol{\alpha}}, \dots)$;

(5) значения э.ф.р. $\hat{\phi}(x)\boldsymbol{\rho}_{\mathbf{m}\boldsymbol{\alpha}}(x_i, p)$ при разных x (например, соответствующих процентилям $\hat{\phi}\boldsymbol{\rho}_{\mathbf{m}\boldsymbol{\alpha}}(x_i, p)$);

(6) значения $\hat{\phi}(x)\boldsymbol{\rho}_{\mathbf{m}\overline{\boldsymbol{\alpha}}}(x_i, p)$ при разных x;

(7) $\hat{\phi}(x + \Delta x)\boldsymbol{\rho}_{\mathbf{m}\boldsymbol{\alpha}}(x_i, p) - \hat{\phi}(x)\boldsymbol{\rho}_{\mathbf{m}\boldsymbol{\alpha}}(x_i, p)$ и $\hat{\phi}(x + \Delta x)\boldsymbol{\rho}_{\mathbf{m}\overline{\boldsymbol{\alpha}}}(x_i, p) - \hat{\phi}(x)\boldsymbol{\rho}_{\mathbf{m}\overline{\boldsymbol{\alpha}}}(x_i, p)$, где Δx — шаг.

Кроме того, $\mathbf{c}_{\boldsymbol{\alpha}}$ может определяться как $\Gamma_t^{-1}(\lambda_{t\alpha})$ или как $u(\lambda_{t\alpha})$; если $\mathbf{c}_{\boldsymbol{\alpha}} = \Gamma_t^{-1}(\lambda_{t\alpha})$, то $\overline{\mathbf{c}}_{\boldsymbol{\alpha}}$ может быть равно $\Gamma_t^{-1}(\lambda_{t\alpha+1})$; классы $\mathbf{c}_{\boldsymbol{\alpha}}/\overline{\mathbf{c}}_{\boldsymbol{\alpha}}$ t-й переменной могут определяться с использованием разбиений на различные процентили (которые определяются как подвыборка значений $\lambda_{t\alpha} \in \mathrm{I}_t$) и т. д.

Таким образом, предлагаемые схемы порождают значительное число синтетических признаков $\Gamma_{k'}(x_i)$ ($10n$ и более при n исходных признаках Γ_k), что делает необходимым введение процедур отбора признаков. Таргетная переменная $\Gamma_t(x_i)$ — числовая, и порождаемые признаки $\Gamma_{k'}(x_i)$ — также числовые. Для данного случая в прикладной математике имеется несколько различных подходов к оценке взаимосвязи $\Gamma_t(x_i)$ и $\Gamma_{k'}(x_i)$: корреляционные оценки (для линейных закономерностей), полиномная аппроксимация с оценкой качества (для нелинейных закономерностей) и методы теории вероятностей / математической статистики, не зависящие от вида закономерности (в том числе на основе «взаимной информации» [7]).

Наиболее фундаментальным представляется тестирование взаимосвязи двух переменных на основе «нулевой гипотезы» об их независимости. Пусть заданы пары тестируемых значений, (x_i, y_i), $i = \overline{1, \mathbf{n}_{(\mathrm{x,y})}}$, э.ф.р. $F_{xy}(x, y)$ характеризует совместное распределение x и y, а э.ф.р. $F_x(x)$ и $F_y(y)$ — индивидуальные распределения переменных. Эмпирическая функция распределения нулевой гипотезы (независимость x и y) определяется как $F_x(x)F_y(y)$.

Для оценки отличий между $F_{xy}(x, y)$ и $F_x(x)F_y(y)$ необходимо ввести расстояние между такими функциями (так называемую «статистику») и оценить достоверность различий посредством того или иного статистического теста. В качестве расстояния можно использовать функции d_f, адаптированные для 2-мерного случая (например, максимальное уклонение $D(\mathrm{F}_{xy}(x, y), \mathrm{F}_x(x)\mathrm{F}_y(y)) = \max(|\mathrm{F}_{xy}(x_i, y_i) - \mathrm{F}_x(x_i)\mathrm{F}_y(y_i)|)$) и статистический тест Колмогорова–Смирнова P_{KC} $(D(\mathrm{F}_{xy}(x, y), \mathrm{F}_x(x)\mathrm{F}_y(y)), n_{(x,y)})$. Тогда $1 - P_{\mathrm{KC}}$ характеризует «информативность» x относительно y.

Более универсальным подходом к оценке достоверности различий между $\mathrm{F}_{xy}(x, y)$ и $\mathrm{F}_x(x)\mathrm{F}_y(y)$ считается прямое вычисление выбранной статистики d_f на множествах пар значений (x_i, y_i), полученных датчиком случайных чисел.

Пусть *оператор* $\hat{\zeta}$*, семплирующий* множество $\mathbf{X}$, формирует набор семплов

$$\hat{\zeta}\mathbf{X} = \{a_1, a_2, \ldots, a_k, \ldots, a_{|\hat{\zeta}X|} | a_k \subset \mathbf{X}\},$$

а процедура random — датчик случайных чисел (в диапазоне $[0 \ldots 1]$). Для каждого семпла a_k принимается, что $n_{(x,y)} = |a_k|$, и вычисляется множество значений d_f для случайных выборок,

$$\mathrm{rnd}\,(\hat{\zeta}\mathbf{X}, d_f) = \left\{ d_f \left(\mathrm{F}_{xy}(x_{ij}, y_{ij}), \mathrm{F}_x(x_{ij})\mathrm{F}_y(y_{ij}), \right.\right.$$
$$\left.\left. x_{ij}, y_{ij} = \mathrm{random},\ j = \overline{1, |a_i|} \right),\ i = \overline{1, \left|\hat{\zeta}X\right|} \right\}.$$

Для $a \in \hat{\zeta}\mathbf{X}$ значение $P(d_f, \hat{\zeta}\mathbf{X}, a, k', t) = 1 - \hat{\phi}(d_f(\mathrm{F}_{k't}(\Gamma_{k'}(z), \Gamma_t(z)), F_{k'}(\Gamma_{k'}(z))\mathrm{F}_t(\Gamma_t(z)))|z \in a)\mathrm{rnd}\,(\hat{\zeta}\mathbf{X}, d_f)$ — статистическая достоверность «зависимости» $\Gamma_t(z)$ и $\Gamma_{k'}(z)$ по статистике d_f на семпле a, а $1 - P(d_f, \hat{\zeta}\mathbf{X}, a, k', t)$ количественно оценивает зависимость.

При заданном способе оценки зависимости $1 - P(d_f, \hat{\zeta}\mathbf{X}, a, k', t)$ задача отбора информативных признаков решается посредством так называемого B-алгоритма, исходно разработанного для построения оптимальных словарей финальных информаций (чему и соответствует литера «B») [8]. Данный алгоритм, основанный на критерии разрешимости по Журавлёву, позволяет выбирать множества финальных информаций на основе максимального частичного покрытия при минимуме элементов покрытия. Замена мощности пересечения множеств на $1 - P(d_f, \hat{\zeta}\mathbf{X}, a, k', t)$ приведет к тому, что B-алгоритм будет выбирать минимум признаков с максимальной «информативностью» (наиболее информативные признаки, см. теоремы 1, 7 и 8 работы [8]).

Таким образом, в рамках развиваемого формализма синтез более информативных синтетических $\Gamma_{k'}(x_i)$ осуществляется в 5 стадий:

(1) определяется набор исходных (как правило, «низкоинформативных») признаков $\Gamma_k(x_i)$ и таргетная переменная $\Gamma_t(x_i)$;

(2) вводится набор метрик ρ_m, оценивается их релевантность $d_f(\hat{\phi}(x)\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\mathbf{c}_{\boldsymbol{\alpha}}, p), \hat{\phi}(x)\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\overline{\mathbf{c}}_{\boldsymbol{\alpha}}, p))$ для каждого класса $\mathbf{c}_{\boldsymbol{\alpha}}$ значений t-й переменной и отбираются наиболее релевантные ρ_m;

(3) посредством каждой из отобранных ρ_m порождаются синтетические признаки $\Gamma_{k'}(x_i)$;

(4) посредством вычислений $1 - P(d_f, \hat{\zeta}\mathbf{X}, a, k', t)$ и B-алгоритма отбирается минимальное число признаков максимальной «информативности»;

(5) применяется алгоритм прогнозирования таргетной переменной (корректор по Журавлёву–Рудакову).

5 Экспериментальная апробация

Формализм апробирован на комплексе задач фармакоинформатики: получение количествен-

Ранговые корреляции между экспериментальными и расчетными значениями EC_{50} и других величин хемокиномного анализа: r — коэффициент ранговой корреляции на обучении; r_c — на контроле. Усреднение r и r_c проводилось по 2400 выборкам хемокиномных данных

Эксперимент	r	r_c
f_{θ_k}-алгоритмы, корректор — нейросеть	$\mathbf{0{,}88 \pm 0{,}15}$	$\mathbf{0{,}86 \pm 0{,}20}$
Синтетические $\Gamma_{k'}(x_i)$, корректор — нейросеть (2 слоя)	$0{,}45 \pm 0{,}22$	$0{,}22 \pm 0{,}21$
Синтетические $\Gamma_{k'}(x_i)$, корректор — нейросеть (10 слоев)	$0{,}52 \pm 0{,}25$	$0{,}21 \pm 0{,}20$
Синтетические $\Gamma_{k'}(x_i)$, корректор — «случайный лес», вариант 1	$0{,}98 \pm 0{,}15$	$0{,}67 \pm 0{,}31$
Синтетические $\Gamma_{k'}(x_i)$, корректор — «случайный лес», вариант 2	$0{,}99 \pm 0{,}14$	$0{,}71 \pm 0{,}35$
Синтетические $\Gamma_{k'}(x_i)$, полиномные корректоры, вариант 1	$\mathbf{0{,}93 \pm 0{,}11}$	$\mathbf{0{,}90 \pm 0{,}23}$
Синтетические $\Gamma_{k'}(x_i)$, полиномные корректоры, вариант 2	$\mathbf{0{,}95 \pm 0{,}08}$	$\mathbf{0{,}86 \pm 0{,}27}$

ных оценок ингибирования киназ протеома перспективными лекарствами (хемокиномный анализ) [9]. Использованы 2400 выборок данных «молекула–свойство» из ProteomicsDB; свойства молекул включили константы EC_{50} и активности для концентраций $(E_j(C_i))$.

Исходные признаки $\Gamma_k(x_i)$ определялись как булевы инварианты над множествами χ-цепей и χ-узлов хемографов x_i, как и в [9]. Таргетная $\Gamma_t(x_i)$ определялась как числовое значение прогнозируемого свойства. В качестве ρ_m использовались функции расстояния на множествах, векторах и э.ф.р. (всего 65 функций из справочника [2]). Классы $\mathbf{c}_{\boldsymbol{\alpha}}$ определялись как квартили значений Γ_t. Векторы элементов $L(T(\mathbf{X}))$ формировались из оценок v_α^+, v_α^- и d_α [4] для каждого $\mathbf{c}_{\boldsymbol{\alpha}}$. Релевантность ρ_m по $d_f(\hat{\phi}(x), \vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\mathbf{c}_{\boldsymbol{\alpha}}, p), \hat{\phi}(x)\vartheta_{\mathbf{m}\boldsymbol{\alpha}}(\overline{\mathbf{c}}_{\boldsymbol{\alpha}}, p))$ оценивалась для каждого $\mathbf{c}_{\boldsymbol{\alpha}}$, d_f — максимальное уклонение. Синтетические признаки $\Gamma_{k'}(x_i)$ порождались всеми перечисленными выше способами; их отбор проводился B-алгоритмом с использованием $1 - P(d_f, \hat{\zeta}\mathbf{X}, a, k', t)$.

В качестве корректоров использовались нейронные сети с несколькими слоями (от 2 до 10) с функцией активации softmax, полиномы различных конструкций (более 20 формул, в том числе квазиполиномные модели с элементарными функциями) и «случайные леса» решающих деревьев. Оператор семплирования $\hat{\zeta}$ был реализован как десятикратная кросс-валидация с делением каждой выборки объектов на 80% (обучение) и 20% (контроль). Результаты экспериментов суммированы в таблице.

Наилучший результат применения нового «топологического» формализма с полиномным корректором ($r_c = 0{,}90 \pm 0{,}23$) немного превзошел наилучший результат применения метода опорных функций (композиций вида $f_{\theta_k} = g(f_1(\sum \omega_k^j x_k), \ldots$ $\ldots, f_l(\sum \omega_k^j x_k))$, см. [9]), для которого $r_c = 0{,}86 \pm$ $\pm\, 0{,}20$. Полиномными формулами, наиболее часто показывавшими наилучший результат, оказались полиномы 1-й или 2-й степеней с произведениями переменных первой степени, полиномы 5-й степени, квазиполиномы 5-й степени с сигмоидами и Фурье-полиномы 3-й степени.

Нейросетевые корректоры всех использованных конфигураций отличались крайне низкими показателями ($r = 0{,}45 \pm 0{,}22$, $r_c = 0{,}22 \pm 0{,}21$), а «случайный лес» приводил к существенному переобучению (см. таблицу). При этом в 290 из 2400 выборок данных (12%) «случайный лес» приводил к улучшению результатов по сравнению с наилучшими полиномными корректорами, а в 1670 из 2400 выборок данных (70%) — к ухудшению.

Анализ синтетических признаков $\Gamma_{k'}(x_i)$, вошедших в наилучшие полиномные модели, показал, что среди более информативных (по оценке $1 - P(d_f, \hat{\zeta}\mathbf{X}, a, k', t)$) признаков чаще всего встречались признаки, порождаемые с использованием э.ф.р. на основе опорных цепей (теорема 1 в 1-й части работы [1]), среди наименее информативных — исходные признаки $\Gamma_k(x_i)$ и признаки на основе отдельных расстояний $\rho_m(\mathbf{c}_{\boldsymbol{\alpha}}, \Gamma_k^{-1}(\Gamma_k(x_i))$. Функциями ρ_m, наиболее часто порождающими информативные $\Gamma_{k'}(x_i)$ на пространстве э.ф.р., оказались максимальное уклонение Колмогорова, «косое» расстояние, метрики Lp, Реньи, $\chi 2$, фон Мизеса, инженерная [2]. В среднем по всем выборкам данных эти 7 разновидностей ρ_m порождали более 50% самых информативных признаков $\Gamma_{k'}(x_i)$, отобранных B-алгоритмом.

6 Заключение

Предлагаемый подход к порождению информативных синтетических признаков подразумевает последовательные трансформации описаний объекта:

(1) исходное множество значений признаков;
(2) множество соответствующих элементов решетки;
(3) множество расстояний (измеряемых посредством ρ_m) между элементами решетки, соответствующими классам и признакам;

(4) множество э.ф.р. расстояний, измеренных заданными ρ_m;

(5) множество синтетических признаков объекта.

Использование многочисленных метрик на стадии порождения признаков позволяет рассматривать развиваемый формализм как вариант развития идеологии АВО (алгоритмы вычисления оценок) научной школы Ю. И. Журавлёва. Экспериментальная апробация предлагаемого подхода на 2400 однородных задачах фармакоинформатики позволила повысить аккуратность и обобщающую способность алгоритмов.

Литература

1. *Торшин И. Ю.* О порождении синтетических признаков на основе опорных цепей и произвольных метрик в рамках топологического подхода к анализу данных. Часть 1. Включение в формализм эмпирических функций расстояния // Информатика и её применения, 2024. Т. 18. Вып. 1. С. 71–77. doi: 10.14357/19922264240110. EDN: RIVOXR.
2. *Деза Е. И., Деза М. М.* Энциклопедический словарь расстояний / Пер. с англ. — М.: Наука, 2008. 444 с. (*Deza E. I., Deza M. M.* Dictionary of distances. — North-Holland: Elsevier, 2006. 412 p. doi: 10.1016/B978-0-444-52087-6.X5000-8.)
3. *Torshin I. Y., Rudakov K. V.* Combinatorial analysis of the solvability properties of the problems of recognition and completeness of algorithmic models. Part 2: Metric approach within the framework of the theory of classification of feature values // Pattern Recognition Image Analysis, 2017. Vol. 27. No. 2. P. 184–199. doi: 10.1134/S1054661817020110.
4. *Торшин И. Ю.* О формировании множеств прецедентов на основе таблиц разнородных признаковых описаний методами топологической теории анализа данных // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 2–7. doi: 10.14357/19922264230301. EDN: AQEUYO.
5. *Torshin I. Yu., Rudakov K. V.* On the procedures of generation of numerical features over partitions of sets of objects in the problem of predicting numerical target variables // Pattern Recognition Image Analysis, 2019. Vol. 29. No. 4. P. 654–667. doi: 10.1134/S1054661819040175.
6. *Torshin I. Y., Rudakov K. V.* Combinatorial analysis of the solvability properties of the problems of recognition and completeness of algorithmic models. Part 1: Factorization approach // Pattern Recognition Image Analysis, 2017. Vol. 27. No. 1. P. 16–28. doi: 10.1134/S1054661817010151.
7. *Sosa-Cabrera G., Gómez-Guerrero S., García-Torres M., Schaerer C. E.* Feature selection: A perspective on inter-attribute cooperation // Int. J. Data Science Analytics, 2024. Vol. 17. P. 139–151. doi: 10.1007/s41060-023-00439-z.
8. *Torshin I. Y.* Optimal dictionaries of the final information on the basis of the solvability criterion and their applications in bioinformatics // Pattern Recognition Image Analysis, 2013. Vol. 23. No. 2. P. 319–327. doi: 10.1134/S1054661813020156.
9. *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: IGSPEW.

Поступила в редакцию 09.04.24

ON THE GENERATION OF SYNTHETIC FEATURES BASED ON SUPPORT CHAINS AND ARBITRARY METRICS WITHIN THE FRAMEWORK OF A TOPOLOGICAL APPROACH TO DATA ANALYSIS. PART 2. EXPERIMENTAL TESTING ON PHARMACOINFORMATICS PROBLEMS

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: Consideration of precedent relationships between features and a target variable in the form of sets of Boolean lattice elements indicates the possibility of generating synthetic features using metric distance functions. Approaches to (i) assessing the relevance ("informativeness") of metrics in relation to the problems being solved; (ii) generating; and (iii) selecting synthetic features that are more informative than the original feature descriptions are formulated. The results of topological analysis of 2400 samples of "molecule–property" data from ProteomicsDB made it possible to obtain fairly effective algorithms for predicting the properties of molecules (rank correlation in cross-validation is 0.90 ± 0.23). Using this sample of problems, metrics have been established

that most often generate informative synthetic features: maximum Kolmogorov deviation, “oblique” distance, and Lp, Renyi, and von Mises metrics. To solve the studied set of problems, the advantage of polynomial correctors compared to neural network and random forest correctors is shown.

Keywords: topological data analysis; lattice theory; algebraic approach of Yu. I. Zhuravlev; pharmacoinformatics

DOI: 10.14357/19922264240207 **EDN:** OTXCUD

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. Torshin, I. Yu. 2024. O porozhdenii sinteticheskikh priznakov na osnove opornykh tsepey i proizvol’nykh metrik v ramkakh topologicheskogo podkhoda k analizu dannykh. Chast’ 1. Vklyuchenie v formalizm empiricheskikh funktsiy rasstoyaniya [On the generation of synthetic features based on support chains and arbitrary metrics within a topological approach to data analysis. Part 1. Inclusion of empirical distance functions into the formalism]. *Informatika i ee Primeneniya — Inform Appl.* 18(1):71–77. doi: 10.14357/19922264240110. EDN: RIVOXR.
2. Deza, E. I., and M. M. Deza. 2006. *Dictionary of distances*. North-Holland: Elsevier. 412 p. doi: 10.1016/B978-0-444-52087-6.X5000-8.
3. Torshin, I. Yu., and K. V. Rudakov. 2017. Combinatorial analysis of the solvability properties of the problems of recognition and completeness of algorithmic models. Part 2: Metric approach within the framework of the theory of classification of feature values. *Pattern Recognition Image Analysis* 27(2):184–199. doi: 10.1134/S1054661817020110.
4. Torshin, I. Yu. 2023. O formirovanii mnozhestv pretsedentov na osnove tablits raznorodnykh priznakovykh opisaniy metodami topologicheskoy teorii analiza dannykh [On the formation of sets of precedents based on tables of heterogeneous feature descriptions by methods of topological theory of data analysis]. *Informatika i ee Primeneniya — Inform Appl.* 17(3):2–7. doi: 10.14357/19922264230301. EDN: AQEUYO.
5. Torshin, I. Yu., and K. V. Rudakov. 2019. On the procedures of generation of numerical features over partitions of sets of objects in the problem of predicting numerical target variables. *Pattern Recognition Image Analysis* 29(4):654–667. doi: 10.1134/S1054661819040175.
6. Torshin, I. Y., and K. V. Rudakov. 2017. Combinatorial analysis of the solvability of the problems of recognition, completeness of algorithmic models. Part 1: Factorization approach. *Pattern Recognition Image Analysis* 27(1):16–28. doi: 10.1134/S1054661817010151.
7. Sosa-Cabrera, G., S. Gymez-Guerrero, M. García-Torres, and C. E. Schaerer. 2024. Feature selection: A perspective on inter-attribute cooperation. *Int. J. Data Science Analytics* 17:139–151. doi: 10.1007/s41060-023-00439-z.
8. Torshin, I. Y. 2013. Optimal dictionaries of the final information on the basis of the solvability criterion and their applications in bioinformatics. *Pattern Recognition Image Analysis* 23(2):319–327. doi: 10.1134/S1054661813020156.
9. Torshin, I. Yu. 2023. O zadachakh optimizatsii, voznikayushchikh pri primenenii topologicheskogo analiza dannykh k poisku algoritmov prognozirovaniya s fiksirovannymi 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.

Received April 9, 2024

Contributor

Torshin Ivan Y. (b. 1972) — Candidate of Science (PhD) in physics and mathematics, Candidate of Science (PhD) in chemistry, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str, Moscow 119333, Russian Federation; tiy135@yahoo.com

ВЫЯВЛЕНИЕ ПРИЧИННО-СЛЕДСТВЕННЫХ СВЯЗЕЙ ПРИ ПОКРЫТИИ ПРИЧИН

А. А. Грушо[1], Н. А. Грушо[2], М. И. Забежайло[3], В. В. Кульченков[4], Е. Е. Тимонина[5]

Аннотация: Задачи поиска причинно-следственных связей имеют большое значение в медицинской диагностике, поиске первопричин сбоев в программно-технических системах, информационной безопасности. Объяснимость формируемых заключений, получаемых в результате сложных вычислений с помощью методов искусственного интеллекта (ИИ), чаще всего реализуется с помощью причинно-следственных связей. В работе исследована возможность выявления причинно-следственных связей в случаях, когда причина находится в неразделимом объекте, доступном наблюдению. В таких случаях говорят, что свойство «причина» покрыто объектом, в котором присутствуют и другие свойства данных. Следствия причин проявляются в других информационных пространствах. Задача выявления причинно-следственных связей исследуется в присутствии других случайных данных, не относящихся к зависимости, порождаемой причинно-следственной связью. Рассматривается модель детерминированной причинно-следственной связи при наличии значительного числа случайно возникающих свойств, которые не связаны с причинно-следственным влиянием одних свойств на другие.

Ключевые слова: искусственный интеллект; компьютерный анализ данных; причинно-следственные связи; покрытие причин

DOI: 10.14357/19922264240208 **EDN:** MKXMZY

1 Введение

Центральным компонентом прикладных систем ИИ и решений выступает их процедурный «инструментарий» — проблемно-ориентированные математические модели, методы и алгоритмы, способные отразить как общие принципы интеллектуального анализа данных, так и специфику конкретных областей приложения. В свою очередь, среди такого рода «инструментов» анализа данных и поддержки принятия решений особое место отводится средствам восстановления причинно-следственных связей, неявным образом представленных в обрабатываемых эмпирических данных.

Например, в системах обеспечения кибербезопасности именно причинно-следственные зависимости, изначально скрытые в анализируемых данных, позволяют строить результативные математические модели и методы фильтрации Big Data на предмет идентификации в них аномальных (например, вредоносных) фрагментов «малого» размера [1].

Дополнительные возможности для принятия выводов и рекомендаций, формируемых системами ИИ, дает формирование содержательных, использующих термины и понятия анализируемой предметной области интерпретаций и объяснений. Обычно объяснение — это ответ на вопрос ПОЧЕМУ? Именно по этой причине ключевая роль в работе прикладных систем ИИ отводится использованию причинно-следственных связей, изначально скрытых в анализируемых данных, однако восстанавливаемых из них соответствующей системой ИИ.

Итак, задачи поиска причинно-следственных связей имеют большое значение в целом ряде значимых приложений — в медицинской диагностике [2, 3], поиске первопричин сбоев в программно-технических системах [4], информационной безопасности [1, 5] и др. При поиске причинно-следственных связей часто использовались методы статистики [6–8], однако в основном речь шла об анализе причинно-следственных связей в условиях статистически обнаруживаемой независимости или, наоборот, зависимости случайных величин.

В данной работе исследуется задача выявления причинно-следственных связей на фоне других случайных данных, не относящихся к исследуемой зависимости. При этом рассмотрены условия, когда

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, grusho@yandex.ru
[2]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, info@itake.ru
[3]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, m.zabezhailo@yandex.ru
[4]Банк ВТБ (ПАО), vlad.kulchenkov@gmail.com
[5]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, eltimon@yandex.ru

причина спрятана (покрыта) в объекте наряду с другими свойствами данных [9].

2 Математическая модель условий поиска причинно-следственных связей

Пусть $A = \{x_1, \ldots, x_n\}$ — это множество наблюдаемых свойств данных. Объектом называется набор свойств из A. Рассмотрим простейший случай, когда каждый объект содержит s свойств и порождается случайно и независимо с вероятностью $1/\binom{n}{s}$.

Исходные данные представляют собой последовательность объектов длины M. Случайность последовательности объектов вводится с помощью равновероятной полиномиальной схемы из M испытаний с известными вероятностями $1/\binom{n}{s}$. Построенную модель будем называть информационным пространством IS_1. Далее информационные пространства будем отождествлять с последовательностями, которые построены по соответствующим этим пространствам вероятностным схемам.

Вместе с IS_1 построим другое информационное пространство IS_2, в котором множество свойств $B = \{y_1, \ldots, y_N\}$, на которых построена полиномиальная схема длины M с вероятностями исходов $\{p(y_j)\}$. Пусть свойство x из A служит причиной свойства y из B. Это значит, что при наличии связи IS_1 и IS_2 (далее предполагаем ее наличие) при появлении в последовательности объектов IS_1 объекта O со свойством x в последовательности IS_2 в тот же момент возникает свойство y, и от данного свойства оставшаяся часть последовательности IS_2 сдвигается на 1 вправо. Далее появление свойств в IS_2 соответствует независимой полиномиальной схеме, определенной выше для IS_2, до следующего появления в IS_1 объекта со свойством x.

3 Случай выявления одной причинно-следственной связи

В условиях построенной модели можно статистически определять причинно-следственные связи. Предположим, что M достаточно велико, чтобы относительные частоты свойств в полиномиальной схеме IS_2 с достаточной степенью уверенности однозначно идентифицировали различные вероятности этой схемы.

Пусть свойство x_i служит причиной появления свойства y_j. Вероятность того, что в данном объекте есть x_i, равна

$$\frac{\binom{n-1}{s-1}}{\binom{n}{s}} = \frac{s}{n}.$$

Обозначим частоты свойств, встретившихся в IS_1, через v (что равносильно числу появлений объектов, содержащих соответствующее свойство), а частоты свойств в последовательности IS_2 через μ. Если x_i служит причиной появления y_j, то в IS_2 реальная длина последовательности равна

$$M^* = M + v(x_i),$$

где $v(x_i)$ — число появлений объектов со свойством x_i, что также соответствует числу появлений x_i в IS_1. Тогда частота встречаемости y_i будет равна

$$\mu^*(y_j) = v(x_i) + \mu(y_j).$$

При фиксированных значениях всех параметров в IS_1 и IS_2 (кроме условия $M \to \infty$) получим следующие соотношения для относительных частот:

$$\frac{v(x_i)}{M} \to \frac{s}{n}, \quad \frac{\mu(y_j)}{M} \to p(y_j),$$

и

$$\frac{\mu^*(y_j)}{M} = \frac{v(x_i)}{M} + \frac{\mu(y_j)}{M} > p(y_j).$$

Отсюда делаем вывод, что y_j — это следствие какой-то причины из IS_1.

Для определения причины x_i сделаем преобразование данных IS_1, которое назовем операцией *синхронизации*. Пусть символ α не используется в обозначениях объектов и x — произвольное фиксированное свойство из A. Тогда в последовательности IS_1 после каждого случая появления объекта со свойством x вставим символ α. Если в последовательности IS_1 встретились подряд несколько объектов со свойством x, то после них в последовательность вставляется такое же число символов α. Построенную последовательность объектов будем обозначать $\mathrm{IS}_1^*(x)$.

Лемма 1. *Если x_i служит причиной y_j, то последовательность $\mathrm{IS}_1^*(x_i)$ синхронизируется с последовательностью IS_2 в том смысле, что каждое новое появление объекта с x_i в $\mathrm{IS}_1^*(x_i)$ порождает в тот же момент появление нового свойства y_j с тем же порядковым номером в последовательности IS_1^*. При этом длины последовательностей IS_2 и $\mathrm{IS}_1^*(x_i)$ совпадают так, что не происходит смещения x_i и порожденных ими y_j, т. е. они находятся на местах с одинаковыми номерами во вновь образованных последовательностях.*

Доказательство. В IS_2 синхронно с вкраплением объекта y_j происходит смещение остальной части последовательности вправо на один шаг. Аналогичное вкрапление на один шаг вправо в IS_1 делается каждый раз с помощью вставки элемента α сразу после появления объекта со свойством x_i. Из того, что появление объектов и свойств в IS_2 происходит одновременно, получаем, что очередное появление объекта со свойством x_i в $\mathrm{IS}_1^*(x_i)$ совпадает с появлением нового встроенного свойства y_j в IS_2 на том же месте (с тем же номером места). Таким образом, число вставок элемента α совпадает с числом вставок свойства y_j, а места появления объектов со свойством x_i синхронно появляются с вкраплениями y_j. Лемма доказана.

Лемма 2. *В последовательности* $\mathrm{IS}_1^*(x_i)$ *свойство* x_i *встречается только в сочетании с* α*, а во всех остальных объектах свойства* x_i *нет.*

Доказательство. Каждое появление объекта со свойством x_i порождает вкрапление y_j в последовательность IS_2. Отсюда следует утверждение леммы.

Лемма 3. *Необходимым условием того, что свойство* x *служит причиной* y_j*, является то, что в* $\mathrm{IS}_1^*(x)$ *всем объектам с* x *синхронно поставлено в соответствие* y_j*.*

Доказательство. Если x служит причиной y_j, то каждая встреча x порождает в IS_2 свойство y_j, а вне детерминированного порождения y_j свойство x в $\mathrm{IS}_1^*(x)$ появляться не может. $\mathrm{IS}_1^*(x)$ в случае, когда x служит причиной y_j, синхронизирует появления x и y_j. Поэтому появление x в этой последовательности вне сочетания с y_j невозможно. Лемма доказана.

Используем алгоритм синхронизации для каждого свойства x. Решение о принятии x в качестве возможной причины для y_j принимается (при выполнении необходимого условия леммы 3) по максимальному соответствию относительных частот x и y_j соответствующим вероятностям.

Рассмотрим задачу поиска в объектах истинной причины следствия y_j. Для каждого x из встретившихся в объектах IS_1 построим последовательность объектов из $\mathrm{IS}_1^*(x)$. Если предположить, что x служит причиной y_j, то в этой последовательности все объекты с x появляются тогда, когда свойство x соответствует y_j. Если x не служит причиной y_j, то в этой последовательности $\mathrm{IS}_1^*(x)$ будут объекты, содержащие неизвестную причину, и объекты x, соответствующие случайным появлениям y_j в последовательности IS_2 (см. необходимое условие).

Для любого x среднее число появлений x равно Ms/n, а в последовательности IS_2 по-прежнему будет $v(x_i)$ свойств y_j, порожденных реальной причиной, и $\mu(y_j)$ свойств y_j без участия свойства x_i. По теореме Муавра–Лапласа

$$v(x_i) = \frac{Ms}{n} + o\left(\ln M\sqrt{M}\right);$$

$$\mu(y_j) = Mp(y_j) + o\left(\ln M\sqrt{M}\right).$$

В то же время

$$v(x) = \frac{Ms}{n} + o\left(\ln M\sqrt{M}\right),$$

и разница с $v(x_i)$ априори может быть компенсирована колебаниями $\mu(y_j)$.

Отсюда следует, что для любого объекта со свойством $x \neq x_i$ необходимо изучить число случайных появлений объектов, которые случайно получаются в $\mathrm{IS}_1^*(x)$ под возникшими свойствами y_j. Для понимания проблемы приведем два примера.

Пример 1. Предположим, что $p(y_j) = 0$. Тогда в последовательности IS_2 следствие y_j появляется только в результате появления в IS_1 причины x_i. Тогда для произвольного $x \neq x_i$ математическое ожидание числа объектов на местах появления y_j равно

$$\frac{Ms(s-1)}{n(n-1)} < \frac{Ms}{n}.$$

Доказательство будет приведено далее. Тогда остальные объекты со свойством x не соответствуют y_j, что по лемме 3 означает, что x не может быть причиной y_j.

Таким образом, приведенный выше алгоритм позволяет однозначно выявить причину.

Пример 2. Предположим, что $p(y_j) = 1$. Тогда в последовательности IS_2 свойство y_j появляется на каждом месте последовательности IS_2. Отсюда для произвольного $x \neq x_i$ в $\mathrm{IS}_1^*(x)$ может случайно иметь больше объектов с x, чем последовательность $\mathrm{IS}_1^*(x_i)$ имеет объектов с x_i. В этом случае алгоритм, основанный на максимизации числа выявленных вкраплений, порождающих y_j, заведомо выберет ложное решение. Поэтому принятие правильного решения потребует выявления дополнительных условий.

С этой целью будем вычислять все частоты сочетаний свойств $x \neq x_i$ и x_i в появившихся объектах. Вероятность появления объектов со свойствами x и x_i равна

$$\frac{\binom{n-2}{s-2}}{\binom{n}{s}} = \frac{s(s-1)}{n(n-1)}.$$

Математическое ожидание числа таких объектов — $Ms(s-1)/(n(n-1))$. Вероятность появления объекта со свойством x_i, но без свойства x равна $\binom{n-2}{s-1}/\binom{n}{s}$. Тогда число y_j, возникших детерминированно, но не соответствующих объектам со свойством x, равно в среднем $M\binom{n-2}{s-1}/\binom{m}{s}$.

Отметим, что

$$\binom{n-2}{s-1}+\binom{n-2}{s-2}=\binom{n-1}{s-1}.$$

Отсюда вероятность появления x_i равна s/n, что соответствует полученному раньше выражению.

Вероятность появления объекта со свойством x, но без свойства x_i равна $\binom{n-2}{s-1}/\binom{n}{s}$. Отсюда следует, что для подтверждения x в качестве причины y_j необходимо, чтобы область случайного появления y_j, а это в среднем $Mp(y_j)$, смогла покрыть в среднем $M\binom{n-2}{s-1}/\binom{n}{s}$ объектов со свойством x. Исходя из того что недетерминированные появления свойств y_j не зависят от появления объектов со свойством x в $\mathrm{IS}_1^*(x)$, математическое ожидание числа «покрытых» y_j объектами со свойством x удовлетворяет неравенству

$$\frac{Mp(y_j)\binom{n-2}{s-1}}{\binom{n}{s}}<\frac{M\binom{n-2}{s-1}}{\binom{n}{s}}.$$

При $p(y_j)<1$ разность математических ожиданий имеет порядок M, поэтому из теоремы Муавра–Лапласа следует, что необходимое число «покрытий» невозможно с вероятностью, стремящейся к 1. Отсюда по лемме 3 получаем следующую теорему.

Теорема 1. *При $M\to\infty$, $p(y_j)<1$ с вероятностью, стремящейся к единице, причина появления y_j определяется однозначно.*

Если в IS_2 в схеме серий допустить $n\to\propto$, то $\mu(y_j)/M\to 0$ по вероятности. Тогда из соотношения

$$\frac{\mu^*(y_j)}{M}=\frac{v(x_i)}{M}+\frac{\mu(y_j)}{M}$$

сразу следует, что y_j — это следствие некоторой причины из IS_1. Приведенный выше пример 1 также позволяет в этом случае выделить x_i как причину y_j.

4 Случай выявления нескольких причинно-следственных связей в разных информационных пространствах

Предположим, что условия построения случайных последовательностей $\mathrm{IS}_1,\mathrm{IS}_2,\ldots,\mathrm{IS}_{k+1}$, $k\le n$, таковы, что существуют k свойств, которые служат причинами в IS_1, и порождающих одиночные следствия в соответствующих пространствах $\mathrm{IS}_2,\ldots,\mathrm{IS}_{k+1}$.

Для простоты рассмотрим случай двух причин и их следствий в IS_2 и в IS_3, где IS_3 основано на свойствах $C=\{z_1,\ldots,z_S\}$ и определяется полиномиальной схемой длины M и вероятностями $\{p(z_i)\}$. Пусть для простоты объект со свойством x_1 порождает следствие y_1 в IS_2, а объект со свойством x_2 порождает следствие z_1 в IS_3. В условиях построения причинно-следственных связей в исходной модели необходимо решить задачу нахождения всех причин и следствий.

Предположим, что M достаточно велико, чтобы относительные частоты свойств в полиномиальных схемах IS_2 и IS_3 с достаточной степенью уверенности однозначно идентифицировали различные неравные вероятности каждой схемы.

Если в IS_1 присутствуют только две причины, то реальная длина последовательности IS_2 увеличилась и стала равной

$$M_1^*=M+v(x_1).$$

В IS_3 реальная длина последовательности также увеличилась и стала равной

$$M_2^*=M+v(x_2).$$

При этом

$$\mu^*(y_1)=v(x_1)+\mu(y_1);\ \eta^*(z_1)=v(x_2)+\eta(z_1).$$

Отсюда следует, что y_1 и z_1 стали следствиями каких-то причин из IS_1. Осталось определить причины этих следствий.

Лемма 4. *Алгоритмы теоремы* 1 *поиска причин y_1 и z_1 в пространстве* IS_1 *функционируют независимо и не влияют на последовательность их выполнения.*

Доказательство. Пространства $\mathrm{IS}_1^*(x)$ для IS_2 и для IS_3 могут быть использованы независимо, так как они будут соотноситься с разными независимыми последовательностями в IS_2 и в IS_3. Аналогично отсев по необходимому условию леммы 3 также основан на независимо построенных последовательностях IS_2 и IS_3. Лемма доказана.

Суммируем результаты в следующей теореме.

Теорема 2. *Пусть для произвольного $k \leq n$ определены информационные пространства* $\mathrm{IS}_1, \mathrm{IS}_2, \ldots, \mathrm{IS}_{k+1}$, *где* IS_1 *содержит последовательность длины* M *равновероятных объектов, а* $\mathrm{IS}_2, \ldots, \mathrm{IS}_{k+1}$ *— последовательности полиномиальных испытаний длины* M *с разными вероятностными схемами, в которых все вероятности больше* 0. *Если все пространства связаны с* IS_1 *так, что находящиеся в* IS_1 k *выделенных свойств служат причинами появления в соответствующих пространствах уникальных следствий (каждое следствие может появляться случайно или под воздействием причины из* IS_1*), то все причинно-следственные связи однозначно определяются с вероятностью, стремящейся к единице*.

Доказательство. По лемме 4 поиск каждой пары причина–следствие может осуществляться независимо друг от друга в любом порядке. При этом сначала на каждом $\mathrm{IS}_2, \ldots, \mathrm{IS}_{k+1}$ определяются свойства, ставшие следствиями неизвестных причин свойств из IS_1, а затем с помощью построенного алгоритма выявляются свойства причины найденных следствий. Теорема доказана.

5 Заключение

В работе исследована возможность выявления причинно-следственных связей в случаях, когда причина находится в неразделимом объекте, доступном наблюдению. Следствия причин проявляются в других информационных пространствах. Шум, препятствующий простому решению задачи, состоит из свойств данных, окружающих причины и следствия, не относящихся к причинно-следственным связям. Такие условия моделируют ситуации, когда в рассматриваемых данных возможна декомпозиция и выделение фрагментов причинно-следственных связей, представляющих интерес для исследователя.

Методика частично апробирована в решении задач поиска «следов» определенных типов инсайдеров в больших данных мониторинга безопасности одного из российских банков.

Литература

1. *Смирнов Д. В.* Методика проблемно-ориентированного анализа Big Data в режиме ограниченного времени // Int. J. Open Information Technologies, 2021. Vol. 9. Iss. 9. P. 88–94. EDN: NKHHGS.
2. *Höfler M.* Causal inference based on counterfactuals // BMC Med. Res. Methodol., 2005. Vol. 5. Art. 28. 12 p. doi: 10.1186/1471-2288-5-28.
3. *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.
4. *Reimer J., Wang Y., Laridi S., Urdich J., Wilmsmeier S., Palmer G.* Identifying cause-and-effect relationships of manufacturing errors using sequence-to-sequence learning // Sci. Rep. — U.K., 2022. Vol. 12. Art. 22332. 11 p. doi: 10.1038/s41598-022-26534-y.
5. *Grusho A., Grusho N., Zabezhailo M., Timonina E.* Evaluation of trust in computer-computed results // Comm. Com. Inf. Sc., 2022. Vol. 1552. P. 420–432. doi: 10.1007/978-3-030-97110-6_33.
6. *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.
7. *Pearl J.* The mathematics of causal inference // Joint Statistical Meetings Proceedings. — ASA, 2013. P. 2515–2529.
8. *Zhang X., Hu W., Yang F.* Detection of cause-effect relations based on information granulation and transfer entropy // Entropy, 2022 Jan 28. Vol. 24. Art. 212. 18 p. doi: 10.3390/e24020212.
9. *Грушо А. А., Грушо Н. А., Забежайло М. И., Тимонина Е. Е., Шоргин С. Я.* Сложные причинно-следственные связи // Информатика и её применения, 2023. Т. 17. Вып. 2. С. 84–89. doi: 10.14357/19922264230212. EDN: TGXQIW.

Поступила в редакцию 09.04.24

IDENTIFICATION OF CAUSE-AND-EFFECT RELATIONSHIPS WHEN COVERING CAUSES

A. A. Grusho[1], N. A. Grusho[1], M. I. Zabezhailo[1], V. V. Kulchenkov[2], and E. E. Timonina[1]

[1]Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

[2]VTB Bank, 43-1 Vorontsovskaya Str., Moscow 109147, Russian Federation

Abstract: The tasks of identification of cause-and-effect relationships are of great importance in medical diagnostics, finding the root causes of failures in software and hardware systems, and information security. The explainability of the formed conclusions obtained as a result of complex calculations using artificial intelligence methods is most

often realized using causal relationships. The paper investigated the possibility of identification of cause-and-effect relationships in cases where the cause is in an inseparable object available for observation. In such cases, it is said that the ”cause” property is covered by an object in which other data properties are present. Effects of causes appear in other information spaces. The cause-and-effect identification problem is investigated in the presence of other random data not related to the relationship generated by the cause-and-effect relationship. The model of deterministic cause-and-effect relationship is considered in the presence of a significant number of randomly occurring properties that are not related to the causal effect of some properties on others.

Keywords: artificial intelligence; computer data analysis; cause and effect; covering causes

DOI: 10.14357/19922264240208 **EDN:** MKXMZY

References

1. Smirnov, D. V. 2021. Metodika problemno-orientirovannogo analiza Big Data v rezhime ogranichennogo vremeni [Methodology of problem-oriented Big Data analysis in limited time mode]. *Int. J. Open Information Technologies* 9(9):88–94. EDN: NKHHGS.
2. Höfler, M. 2005. Causal inference based on counterfactuals. *BMC Med. Res. Methodol.* 5:28. 12 p. doi: 10.1186/1471-2288-5-28.
3. Richens, J. G., C. M. Lee, and S. Johri. 2020. Improving the accuracy of medical diagnosis with causal machine learning. *Nat. Commun.* 11(1):3923. 9 p. doi: 10.1038/s41467-020-17419-7.
4. Reimer, J., Y. Wang, S. Laridi, J. Urdich, S. Wilmsmeier, and G. Palmer. 2022. Identifying cause-and-effect relationships of manufacturing errors using sequence-to-sequence learning. *Sci. Rep. — U.K.* 12:22332. 11 p. doi: 10.1038/s41598-022-26534-y.
5. Grusho, A., N. Grusho, M. Zabezhailo, and E. Timonina. 2022. Evaluation of trust in computer-computed results. *Comm. Com. Inf. Sc.* 1552:420–432. doi: 10.1007/978-3-030-97110-6_33.
6. 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.
7. Pearl, J. 2013. The mathematics of causal inference. *Joint Statistical Meetings Proceedings*. ASA. 2515–2529.
8. Zhang, X., W. Hu, and F. Yang. 2022. Detection of cause-effect relations based on information granulation and transfer entropy. *Entropy* 24(2):212. 18 p. doi: 10.3390/e24020212.
9. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, E. E. Timonina, and S. Ya. Shorgin. 2023. Slozhnye prichinno-sledstvennye svyazi [Complex cause-and-effect relationships]. *Informatika i ee Primeneniya — Inform. Appl.* 17(2):84–89. doi: 10.14357/19922264230212. EDN: TGXQIW.

Received April 9, 2024

Contributors

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; grusho@yandex.ru

Grusho Nikolai A. (b. 1982) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; info@itake.ru

Zabezhailo Michael I. (b. 1956) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; m.zabezhailo@yandex.ru

Kulchenkov Vladislav V. (b. 1989) — deputy head, Portfolio Analysis Department, VTB Bank, 43-1 Vorontsovskaya Str., Moscow 109147, Russian Federation; vlad.kulchenkov@gmail.com

Timonina Elena E. (b. 1952) — Doctor of Science in technology, professor, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; eltimon@yandex.ru

ПРИМЕНЕНИЕ РАЗЛОЖЕНИЯ ИЗОБРАЖЕНИЯ С ПОМОЩЬЮ ДИСКРЕТНОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ ДЛЯ ПОСТРОЕНИЯ АРХИТЕКТУРЫ ШУМОПОДАВЛЯЮЩЕЙ НЕЙРОННОЙ СЕТИ

А. С. Коваленко[1]

Аннотация: Подавление шума на цифровых изображениях — одна из самых распространенных задач в области обработки изображений. На данный момент широкое применение имеют подходы подавления шума, основанные на применении сверточных нейронных сетей (CNN, convolutional neural network). При этом, как правило, обучение модели строится на минимизации функции ошибки между результатом работы сети и ожидаемым эталонным изображением и дополнительно не используются различные представления двумерного сигнала изображения и их свойства для оптимизации обучения архитектур шумоподавляющих сетей. Предложен подход к обучению нейронных сетей подавлять шум. Описанный подход основан на применении N-кратного быстрого вейвлет-преобразования Хаара (БВПХ). Такое представление дискретного сигнала изображения позволяет отказаться от классической архитектуры автоэнкодера и использовать только его часть, кодирующую сигнал, что приводит к значительному сокращению параметров модели и ускоряет работу сети.

Ключевые слова: нейронные сети; глубокое обучение; шумоподавление изображений; методы обработки изображений

DOI: 10.14357/19922264240209 **EDN:** UEQSXP

1 Введение

Наилучшие результаты решения задачи подавления шума на цифровых изображениях демонстрируют подходы, основанные на применении глубоких сверточных нейронных сетей. Как правило, данные сети имеют архитектуры, схожие с моделью U-Net [1], где сеть представлена в виде автокодировщика со сквозной передачей сигнала между слоями кодировщика входного сигнала и его декодера. Общая схема модели U-Net изображена на рис. 1. Авторы работы [2] рассматривают различные модификации архитектуры U-Net для подавления шума на входном изображении и подходы к обучению таких моделей.

Изображение, передаваемое в нейронную сеть, можно рассматривать как сумму значений элементов матрицы чистого изображения I с матрицей, содержащей шум, получаемый из некоторого распределения P, и может быть записано выражением:

$$\tilde{I} = I + \alpha, \;\; \alpha \sim P\,.$$

Поскольку погрешность приема оптического сигнала зависит от физических свойств КМОП-сенсора (КМОП — комплементарная структура металл–оксид–полупроводник), то для каждой модели существует некоторое уникальное распределение P, генерирующее шумовую составляющую сигнала. Также на уровень шума будут иметь значительное влияние настройки камеры и условия съемки [3], при увеличении уровня светочувствительности сенсора будет возрастать и отношение шума к чистому сигналу.

Задача обучения нейронной сети f заключается в поиске оптимального набора весов слоев сети w, при котором будет достигнут минимум аппроксимированного эмпирического риска:

$$\tilde{Q}(w, X^l) = \sum_{i=1}^{l} \mathcal{L}(I_i, f(\tilde{I}_i, w)) \to \min_w\,.$$

В качестве функции ошибки для обучения модели могут быть выбраны меры схожести изображений L_1 и L_2, функция показателя индекса структурного сходства (SSIM, structure simularity) [4] или комбинация нескольких мер сходства изображений, как предлагаемая авторами работы [5] функция ошибки MIX, которая представляет собой взвешенную сумму нормы L_1 и многомасштабной SSIM (MS-SSIM, multiscale SSIM).

Для повышения качества работы U-Net-подобных моделей в их архитектуру интегрируют слои межканального и пространственного внимания [6]. Эти дополнительные слои позволяют модели извлекать не только локальные признаки изображе-

[1]Южный федеральный университет, Институт математики, механики и компьютерных наук им. И. И. Воровича, akov@sfedu.ru

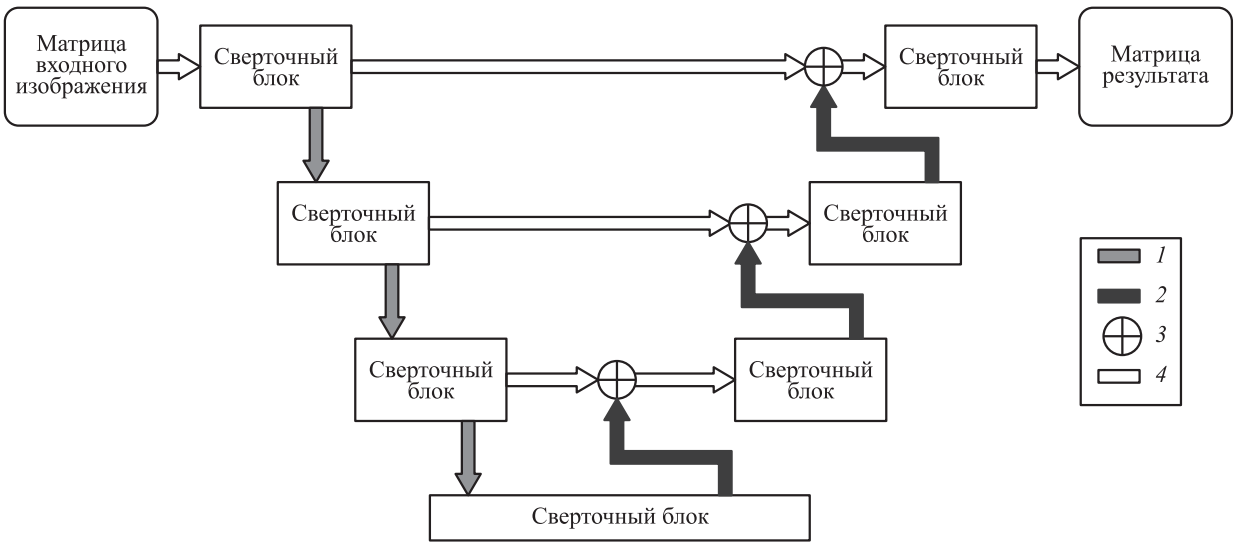


Рис. 1 Общая схема архитектуры U-Net: *1* — понижение разрешения в 2 раза; *2* — повышение разрешения в 2 раза; *3* — конкатенация матриц по каналам; *4* — передача выходных значений блока

ния, но и работать с глобальными особенностями изображения [7]. Также для улучшения шумоподавляющих свойств обучаемой модели могут использоваться различные преобразования значений скрытого пространства слоев сети. Авторы архитектуры Multilevel Wavelet-CNN (MWCNN) [8] используют дискретное вейвлет-преобразование Хаара для разложения сигнала, передаваемое между слоями модели. Это позволяет улучшить качество восстановления высокочастотной компоненты входного изображения и минимизировать эффект размытия обработанного сетью изображения.

Авторы работы [9] используют вейвлет-преобразования в комбинации с методом объединения признаков, извлеченных разными блоками модели для сохранения информации в восстанавливаемом изображении о текстурах и границах объектов. Повышение разрешения карт признаков путем применения вейвлет-преобразований также используют авторы работы [10]. Подход, схожий с применением вейвлет-преобразования для обработки признаков входного сигнала, применяется в работе [11].

Перечисленные выше подходы используют вейвлет-преобразования для улучшения извлечения скрытыми слоями сети более устойчивых признаков из изображения, но не используются в построении функции ошибки. Если рассматривать задачу повышения разрешения изображения, то существуют работы, где функция ошибки при обучении модели основывается на применении преобразовании Хаара. Так, в работе [12] сеть изучает матрицы коэффициентов, полученные преобразованием Хаара, которые необходимо добавить к вейвлет-матрицам входного изображения для уточнения его границ. Данный подход уже использует обратное вейвлет-преобразование для получения итогового изображения.

Также существуют подходы, использующие специальные функции ошибок, которые учитывают значения скрытых пространств слоев нейронной сети. Для решения задачи удаления фона на изображении авторы работы [13] разработали собственную U-Net-подобную архитектуру — IS-Net, а также функцию ошибки для ее обучения, которая требует от глубоких слоев модели строить результирующую матрицу маски главного объекта в разных масштабах, далее результаты работы всех слоев объединяются для построения финальной маски объекта.

На основе идей повышения разрешения изображения с помощью предсказываемых матриц высокочастотных коэффициентов [12] и построения функции ошибки, учитывающей скрытое состояние слоев модели [13], предлагается подход к обучению архитектур, подобных ResNet [14], без необходимости добавления обучаемого декодера для преобразования скрытого пространства модели в изображение.

2 Предлагаемый подход

2.1 Модификация архитектуры классификации

Классические архитектуры для решения задачи классификации изображений, как правило, состоят из блоков, содержащих сверточные слои, слои

нормализации и операций объединения [15]. Операции объединения максимумов (Max Pooling) вычисляют максимальное значение для предсказанных матриц признаков предыдущими сверточными слоями и используют их для создания матриц признаков с пониженной дискретизацией. Эти операции используются в архитектуре ResNet [14] и позволяют извлекать наиболее устойчивые признаки из предыдущих предсказаний при уменьшении их размера в 2 раза. Последними слоями в архитектурах для решения задач классификации служат полносвязные слои, которые строят распределение вероятностей классов, содержащихся на входном изображении. Для построения шумоподавляющей архитектуры сети, рассматриваемой в данной работе, слои классификации не используются и удаляются из применяемых моделей классификации. Удаление последних слоев из модели ResNet делает ее полносверточной архитектурой [16], и она становится инвариантной к размеру входного изображения.

Из выбранной архитектуры для задачи классификации рассматриваются промежуточные матрицы признаков, полученные после применения каждого из сверточных блоков. Обозначим набор данных значений $P = \{p_i\}_{i=1}^{5}$. К каждой матрице признаков p_i применим сверточный слой Conv_i с размером ядра $\dim(K_i) = 9 \times C_i \times 3 \times 3$ при $i \leq \leq 2$ или $\dim(K_i) = 9 \times C_i \times 1 \times 1$ для остальных уровней блоков, где C_i — число каналов у выходной матрицы соответствующего блока модели. Выбор разного размера ядер сверточных слоев обусловлен различной размерностью матриц признаков. У матриц с более глубоких уровней размерность ниже. После применения сверточных слоев к набору значений P получается новый набор $F = \{f_i\}_{i=1}^{5}$, где $f_i = \mathrm{Conv}_i(p_i)$. Дополнительно к сверточным слоям Conv_i был добавлен механизм канального и пространственного внимания CBAM (Convolutional Block Attention Module) [6]. Модифицированную архитектуру с параметрами внутренних слоев w обозначим $\Phi(\tilde{I}, w)$.

Схема предлагаемой архитектуры нейронной сети Φ для предсказания чистого изображения I по входному изображению с шумом $\tilde{I}$ изображена на рис. 2.

2.2 Вейвлет-преобразование Хаара

Вейвлет-преобразование Хаара (ВПХ) позволяет декомпозировать сигнал на две компоненты: аппроксимацию сигнала и детализацию сигнала [17]. Для получения следующего уровня разложения ВПХ применяется к полученному аппроксимационному сигналу. При условии, что первоначальный дискретный сигнал представлен массивом их 2^n чисел, можно рекурсивно применить дискретное ВПХ n раз к данному сигналу, получая n-кратное применение дискретного вейвлет-преобразование Хаара.

При рассмотрении одномерного сигнала в виде набора значений $F = \{f_i\}_{i=1}^{N}$ коэффициенты Хаара для аппроксимации a_i и детализации d_i вычисляются по формулам:

$$a_i = \frac{f_{2i} + f_{2i+1}}{\sqrt{2}};\ \ d_i = \frac{f_{2i} - f_{2i+1}}{\sqrt{2}},\ \ i \in \left[1, \frac{N}{2}\right].$$

В случае использования БВПХ формулы для вычисления коэффициентов будут иметь следующий вид:

$$a_i = \frac{f_{2i} + f_{2i+1}}{2};\ d_i = \frac{f_{2i} - f_{2i+1}}{2},\ i \in \left[1, \frac{N}{2}\right].$$

Данная модификация преобразования является вычислительно более быстрым по сравнению с оригинальным.

Обратное БВПХ для получения первоначальных значений сигнала из коэффициентов БВПХ можно получить по формуле

$$f_i = a_{\lfloor i/2 \rfloor} + (-1)^{i-1 \bmod 2} d_{\lfloor i/2 \rfloor}.$$

Так как в данной работе происходит обработка цифровых изображений, то необходимо использовать БВПХ для дискретного двумерного сигнала. В случае двумерного сигнала БВПХ сначала применяется к строкам матрицы изображения I. Обозначим матрицы получаемых коэффициентов $a_j^{\mathrm{row}_i}$ и $d_j^{\mathrm{row}_i}$, а далее повторно применим БВПХ к столбцам матриц, полученных после первого разложения. В результате данных преобразований будут получены матрицы коэффициентов $A = \{a_{i,j}\}_{i=1,j=1}^{H,W}$, $H = \{h_{i,j}\}_{i=1,j=1}^{H,W}$, $V = \{v_{i,j}\}_{i=1,j=1}^{H,W}$, $D = \{d_{i,j}\}_{i=1,j=1}^{H,W}$. Матрица A содержит информацию об аппроксимации двумерного сигнала, матрицы H, V и D — о горизонтальных, вертикальных и диагональных различиях значений соответственно.

2.3 Применение вейвлет-преобразования Хаара для обучения сети

При применении n-кратного ВПХ к квадратному изображению размером $N \times N$ матрицы коэффициентов будут иметь размер $N/2^n \times N/2^n$, поскольку с каждым применением ВПХ к изображению или последующей аппроксимационной матрице размер будет уменьшаться ровно в 2 раза.

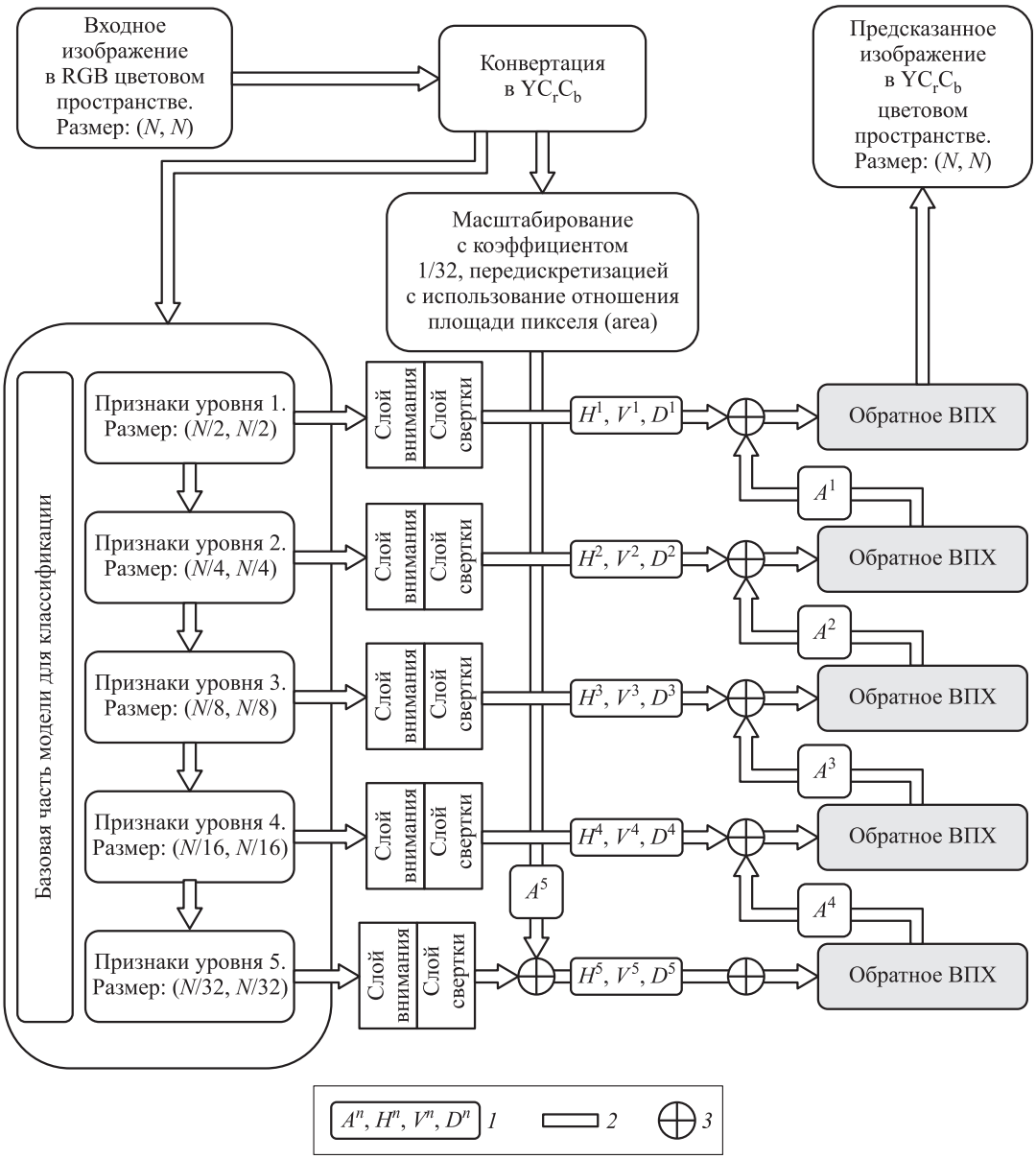


Рис. 2 Схема предлагаемой архитектуры модели подавления шума на изображении: *1* — матрицы коэффициентов n-кратного ВПХ; *2* — передача выходных значений блока; *3* — конкатенация матриц по каналам

Применяя ВПХ к многоканальному изображению, коэффициенты надо рассчитывать для каждого канала отдельно. Если объединить разложения для всех каналов изображения, получится матрица размера $12 \times N/2 \times N/2$, содержащая в себе конкатенацию матриц A, H, V и D, при условии, что используемое изображение I имеет три канала и размер $N \times N$, где N представляет собой некоторую степень 2, причем эта степень $q \geq 5$.

Во время обучения модель учится предсказывать изображение без шума I по входной матрице изображения $\tilde{I}$. Для сокращения параметров сети предлагается учить слои сети предсказывать матрицы f_i, равные объединениям матриц коэффициентов деталей (H^i, V^i, D^i) i-кратного БВПХ, примененного к I. Данная оптимизация позволяет отказаться от построения слоев декодировки результирующего изображения из скрытого состояния сети.

Для построения итогового изображения по высокочастотным коэффициентам необходимо применять обратное БВПХ к соответствующим матрицам коэффициентов. Обозначив прямое БВПХ как DWT, а обратное — IWT, формулы для расчета матриц коэффициентов можем записать в виде

$$A^i, H^i, V^i, D^i = \mathrm{DWT}\left(A^{i-1}\right), \;\; A^0 = I\,,$$
$$A^i = \mathrm{IWT}\left(A^{i-1}, H^{i-1}, V^{i-1}, D^{i-1}\right).$$

Таким образом, нейронная сеть Ф будет учиться предсказывать наборы коэффициентов БВПХ $\{H^i, V^i, D^i\}_{i=1}^5$, но для восстановления изображения I необходима аппроксимация изображения 5-кратного применения ВПХ — A^5, поскольку она необходима для вычисления A^4. Для получения A^5 предлагается применить 5-кратное БВПХ ко входному изображению с шумом $\tilde{I}$ и использовать его коэффициенты приближения изображения $\tilde{A}^5$. Матрица $\tilde{A}^5$ будет близка к матрице приближения чистого изображения A^5, поскольку с каждым применением БВПХ уровень шума в каждом последующем приближении $\tilde{A}^i$ снижается [18], при этом информация о шуме α будет содержаться в матрицах $\tilde{H}^i$, $\tilde{V}^i$ и $\tilde{D}^i$ [19].

Для сокращения вычислений при расчете матрицы $\tilde{A}^5$ необходимо применять БВПХ только к коэффициентам приближения $\{\tilde{A}^q\}_{q=1}^4$. Так, коэффициенты матрицы $\tilde{A}^q$ будут вычисляться по формуле:

$$\tilde{A}^q_{i,j} = \frac{\tilde{A}^{q-1}_{2i,2j} + \tilde{A}^{q-1}_{2i,2j+1} + \tilde{A}^{q-1}_{2i+1,2j} + \tilde{A}^{q-1}_{2i+1,2j+1}}{4},$$
$$i, j \in \left[1, \frac{N}{2^q}\right].$$

Если выразить $\tilde{A}^q$ через значения пикселей изображения $\tilde{I}$, формула примет вид:

$$\tilde{A}^q_{i,j} = \sum_{m=1}^{2q}\sum_{s=1}^{2q} \frac{\tilde{I}_{2qi+m,2qj+s}}{4q^2}, \;\; i, j \in \left[1, \frac{N}{2^q}\right]. \quad (1)$$

Выражение (1) совпадает с формулой вычисления пикселя при масштабировании изображения с коэффициентом $1/2^q$ с помощью метода передискретизации с использованием отношения площади пикселя. Данный метод интерполяции изображения называется Area в терминологии библиотеки обработки изображений OpenCV [20]. Матрицу $\tilde{A}^5$ можно напрямую вычислить, применяя данный метод интерполяции к изображению $\tilde{I}$.

2.4 Сравнение с существующими подходами к внедрению вейвлет-преобразований в архитектуры шумоподавляющих моделей

Архитектура MWCNN, разработанная авторами работы [8], основана на применении вейвлет-преобразований для понижения размерности карт признаков при увеличении числа каналов. Это позволило отказаться от операций пулинга (Pooling) при проектировании модели. Модель построена на последовательном применении ВПХ и сверточных слоев к коэффициентам ВПХ. В качестве операций обратного пулинга или повышения размерности карт признаков авторами применяется обратное ВПХ. Также в модели применяется сквозная передача признаков из кодирующей части модели в декодирующую, что делает ее схожей в общем виде с классической архитектурой U-Net.

Авторы работы [10] применяют вейвлет-преобразования для выделения высокочастотной пространственной информации сигнала. К выделенной высокочастотной компоненте применяется блок сверточных слоев для улучшения высокочастотных признаков сигнала.

На основе идеи разложения сигнала на компоненты с помощью вейвлет-преобразований основан подход, описанный в работе [11]. Авторы применяют преобразование с обучаемым ядром, которое позволяет в процессе обучения модели эффективно разделять сигнал на две компоненты. Данные преобразования по аналогии с работой [10] применяются вместе со сверточными слоями для извлечения признаков из сигнала.

В отличие от подходов [8, 10, 11] предлагаемая архитектура не использует ВПХ для разложения сигнала изображения на компоненты с целью последующей обработки сверточными слоями. Обратное ВПХ используется для повышения разрешения изображения по предсказанным коэффициентам ВПХ аналогично подходу [12]. Но используется не однократное ВПХ, а 5-кратное, что позволяет преобразовать изображение размера 32×32 в 256×256. Таким образом, по входному зашумленному изображению модель предсказывает коэффициенты для каждого уровня ВПХ, позволяя по ним восстанавливать изображение исходного размера без шума. Использование описанного способа получения результирующего изображения заменяет применение декодирующей части U-Net подобных архитектур [8, 10].

3 Эксперименты по обучению моделей

3.1 Обучающий набор данных

Обучающая выборка состояла из объединения нескольких наборов данных. Для обучения на изображениях с шумом, полученных с реальных сенсоров камер, использовался открытый набор данных Smartphone Image Denoising Dataset (SIDD) [21].

Набор SIDD предоставляет реальные зашумленные изображения и соответствующие им чистые изображения. Обучающая часть набора содержит 320 изображений высокого разрешения, а проверочная часть содержит 1280 пар изображений, имеющих размер 256×256 точек. Съемка проводилась авторами на 5 мобильных устройств с КМОП-сенсорами.

Также для обучения использовались изображения из наборов Set5, Set14, Sun-Hays 80 [22] и DIV2K [23]. Данные наборы содержат изображения в двух масштабах и, как правило, используются для обучения моделей, повышающих разрешение изображения. Так как изображения в наборах не содержат шума [22], они могут использоваться в задаче обучения шумоподавляющих моделей. Авторы работ [8, 10] используют перечисленные наборы для обучения моделей подавлять шум на изображениях. Для получения изображений с шумом авторы добавляют к матрицам дополнительный гауссовский шум. Таким образом получаются пары изображений для обучения шумоподавляющих нейронных сетей.

В проводимых экспериментах к чистым изображениям из наборов [22, 23] попиксельно добавлялся дополнительный гауссовский шум с фиксированным параметром математического ожидания, равным нулю, и изменяемым значением среднеквадратичного отклонения σ. При каждой загрузке изображения параметр σ выбирался случайным образом из равномерного распределения $R(\sigma|a,b)$ с диапазоном $a = 0$, $b = 90$. Шум для каждого пикселя изображения семплировался независимо от остальных пикселей. Получение изображения с шумом приводится в формуле:

$$\tilde{I}_{i,j} = I_{i,j} + \alpha_{i,j}, \;\; \alpha_{i,j} \sim N(0, \sigma|R(0,90)), \\ i \in [1,H], \; j \in [1,W].$$

Для тестирования работы обученных моделей использовалась валидационная часть набора данных SIDD, а также набор BSD68 [24] с заранее наложенным шумом для корректного сравнения с результатами работ [8, 10, 11]. Дополнительно модель тестировалась на изображениях из набора The Darmstadt Noise Dataset (DND) [25].

3.2 Исследуемые архитектуры

Предлагаемая архитектура, использующая предсказания коэффициентов матриц ВПХ, в приведенных исследованиях обозначается как Wavelets Prediction Network (WPNet), а указанная в скобках архитектура используется в качестве базовой модели.

Для сравнения результатов дополнительно строился декодер, использующийся в классической архитектуре U-Net [1]. В оригинальной реализации декодера использовались два варианта слоев для повышения разрешения: слои деконволюции и операции билинейной интерполяции с последующим применением сверточных слоев. В экспериментах обучались оба варианта декодеров. В качестве кодировщика для модели U-Net была выбрана архитектура ResNet10.

Для обучения моделей U-Net использовалась функция ошибки $\mathcal{L}_{\text{MIX}}$ [5].

3.3 Параметры обучения

Архитектура модели и код обучения реализованы на фреймворке глубокого обучения PyTorch [26]. Реализации архитектур моделей классификации для интеграции в предлагаемый подход использовались из библиотеки timm [27].

Модели обучались на случайных срезах из изображений размером 256×256 пикселей. Вырезанные части изображений конвертировались в цветовое пространство $\text{YC}_\text{r}\text{C}_\text{b}$, где наибольший вклад в детализацию изображения вкладывает компонента изображения Y, по которой происходит оценка качества работы моделей в работе [28]. Для обучения параметров модели w применялся метод стохастической оптимизации с адаптивным параметром скорости обучения AdaSmooth [29]. Начальное значение параметра скорости обучения задавалось равным 0,001. В качестве функции ошибки использовалась $\mathcal{L}_{\text{MIX}}$ [5].

Модель запускалась на входном изображении и на преобразованных изображениях. В качестве преобразований использовались повороты на 90°, 180° и 270°, а также отражения изображения по вертикали и горизонтали. После применения модели к этим изображениям использовались соответствующие обратные преобразования. Итоговое изображение получалось после усреднения результатов работы модели на преобразованных матрицах.

Для валидации использовалась метрика пикового отношения сигнала к шуму PSNR (peak signal-to-noise ratio). Обучение модели останавливалось при выходе графика результата валидационной метрики на плато.

Эксперименты проводились на вычислительной машине с графическим ускорителем Nvidia RTX 4090, процессором Intel i9-10920X и объемом оперативной памяти 64 ГБ. При размере входных изображений 256×256 в процессе обучения использовались пакеты размером 128 (batch size).

4 Результаты

Предлагаемая архитектура, использующая предсказания коэффициентов матриц ВПХ, в приведенных результатах обозначается как WPNet, а указанная в скобках архитектура используется в качестве базовой модели.

Для сравнения результатов использовались метрики пикового отношения сигнала к шуму (PSNR) и структурного сходства изображений (SSIM).

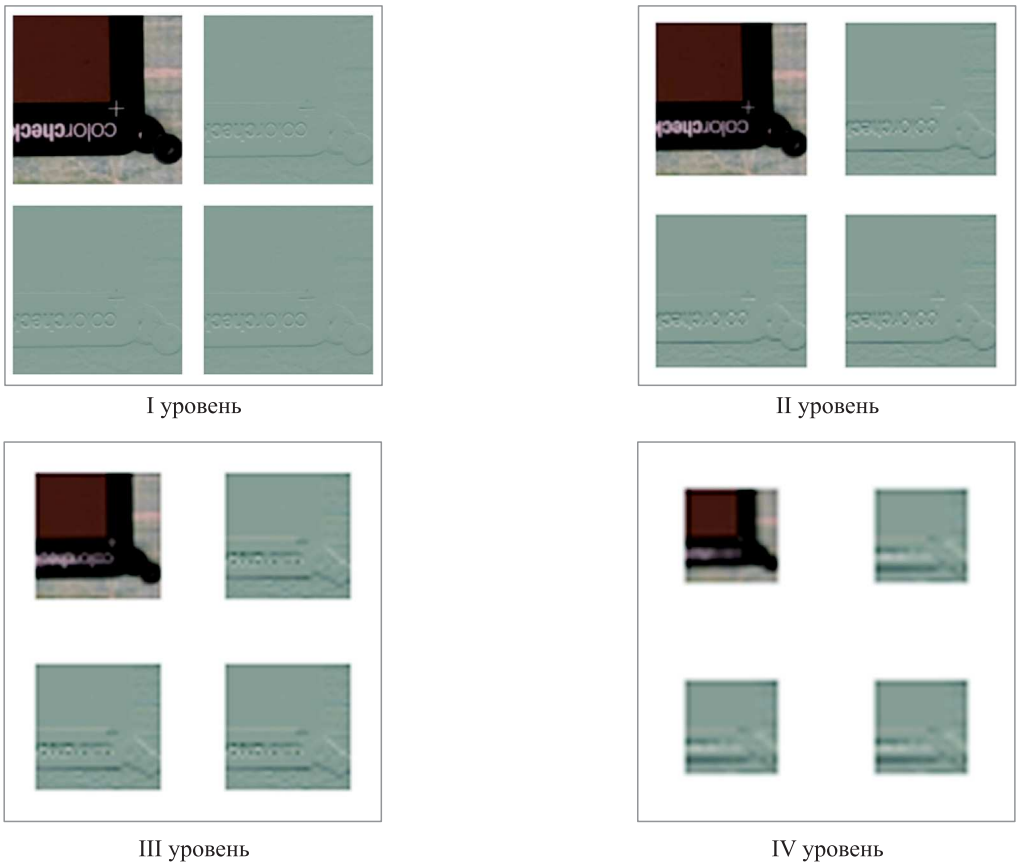
Оценка качества подавления шума оценивалась как на реальных изображениях из валидационной части набора SIDD [21], содержащих шум, так и на изображениях из набора BSD68 [24] с добавочным гауссовским шумом.

При оценке отклонения предсказываемых моделью коэффициентов ВПХ от эталонных значений использовалась метрика Smooth L_1 [30]. Для оценки строились средние значения и среднеквадратичные отклонения по множеству всех рассматриваемых изображений из набора. Данные, построенные на валидационной части набора SIDD приведены в табл. 1, все значения умножены на 10^4. В табл. 1 не приведены коэффициенты аппроксимации A, поскольку они строятся по входному изображению в соответствии с формулой (1). Коэффициенты на первых уровнях модели имеют большее отклонение, чем выходы более глубоких слоев, поскольку результат более глубоких уровней модели получен применением большего числа слоев. Но в предлагаемой архитектуре результат напрямую строится из всех уровней коэффициентов ВПХ. И коэффициенты первых применений ВПХ меньше влияют на предсказываемое изображение, чем коэффициенты дальнейших применений ВПХ.

Таблица 1 Сравнение отклонения коэффициентов БВПХ модели WPNet

Коэффициенты БВПХ		Уровень модели				
		1	2	3	4	5
H	$\mathbb{E}$	2,105	1,132	0,589	0,176	0,0424
	STD	2,175	1,395	0,583	0,175	0,0426
V	$\mathbb{E}$	2,108	1,393	0,586	0,174	0,0417
	STD	2,177	1,394	0,580	0,172	0,0421
D	$\mathbb{E}$	1,131	0,939	0,474	0,165	0,0446
	STD	1,183	0,955	0,471	0,164	0,0446

Пример предсказанных матриц с коэффициентами n-кратного БВПХ приведен на рис. 3.



I уровень II уровень

III уровень IV уровень

Рис. 3 Пример предсказанных моделью коэффициентов 4-кратного БВПХ

Таблица 2 Сравнение качества и размера предлагаемой архитектуры с другими моделями на наборе данных BSD68

Название модели	$\sigma = 15$		$\sigma = 25$		$\sigma = 50$		Число параметров модели $\times 10^6$
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
BM3D [31]	31,08	0,872	28,57	0,802	25,62	0,687	—
DnCNN [32]	31,73	0,891	29,23	0,828	26,23	0,719	0,56
MWCNN [8]	31,86	0,895	29,41	0,836	26,53	0,737	24,92
MWDCNN [10]	31,77	—	29,28	—	26,29	—	5,24
WINNet [11]	31,7	—	29,24	—	26,31	—	0,17
U-Net(ResNet10)	31,26	0,898	30,29	0,872	29,06	0,83	6,47
U-Net(ResNet10) + Bilinear	32,16	0,906	30,79	0,874	29,3	0,827	7,28
WPNet(MobileNetV2)	31,65	0,894	30,2	0,861	28,6	0,81	1,83
WPNet(ResNet10)	31,5	0,901	30,33	0,87	28,87	0,824	4,99

Таблица 3 Сравнение качества и размера предлагаемой архитектуры с другими моделями на валидационном наборе SIDD

Название модели	SIDD		DND		Число параметров модели $\times 10^6$
	PSNR	SSIM	PSNR	SSIM	
BM3D [31]	25,65	0,685	34,51	0,851	—
DnCNN [32]	35,13	0,896	37,03	0,932	0,56
U-Former [28]	39,89	0,96	39,98	0,955	26,87
U-Net(ResNet10)	37,65	0,897	37,62	0,947	6,47
U-Net(ResNet10) + Bilinear	37,79	0,891	38,03	0,946	7,28
WPNet(MobileNetV2)	36,77	0,913	36,14	0,937	1,83
WPNet(ResNet10)	37,28	0,927	36,6	0,944	4,99

На наборе BSD68 с добавочным шумом предлагаемая модель сравнивалась с результатами широко распространенных работ BM3D [31], DnCNN [32], а также с подходами, использующими вейвлет-преобразования [8, 10, 11]. Результаты сравнения на изображениях с разным параметром σ добавочного шума приведены в табл. 2, где шум генерировался из нормального распределения $N(0, \sigma)$. Для подхода BM3D не указывается число параметров, поскольку он не основан на использовании обучаемых нейронных сетей.

Для сравнения на наборе SIDD использовались подходы [31, 32] и современная архитектура, основанная на механизме самовнимания U- Former [28]. Авторы работ [8, 10, 11] не проводили исследований по обучению и тестированию моделей на наборе SIDD. Результаты сравнения качества подавления шума на изображениях из набора SIDD приведены в табл. 3.

Также оценивалась производительность предлагаемой архитектуры в сравнении с другими архитектурами, использующими ВПХ, и с U-Net вариантами модели. Замеры времени работы моделей производились на процессоре Intel i9-10920X. Каждая модель запускалась 100 раз на изображении размером 256×256, и время запусков усреднялось. Для запуска тестов использовалась версия 2.2.1 фреймворка PyTorch и операционная система Ubuntu 22.04. Результаты сравнения приведены в табл. 4. По результатам сравнения предлагаемая архитектура показала меньшее время работы, чем аналогичная U-Net-подобная архитектура, и превзошла по скорости работы модели [8, 10, 11].

Таблица 4 Сравнение скорости работы предлагаемой архитектуры с другими моделями

Название модели	Время, мс
MWCNN [8]	109
MWDCNN [10]	292
WINNet [11]	2297
U-Net (ResNet10)	72
WPNet (ResNet10)	56

Полученные модели достигают схожего качества в сравнении с современными архитектурами при меньшем числе параметров. При сравнении с архитектурами, использующими вейвлет-преобразования [8, 10, 11], удалось добиться лучшего качества подавления добавочного шума. Если рассматривать предлагаемый подход к декодиро-

ванию предсказанных признаков в изображение, достигается качество, схожее с использованием декодирующей части архитектуры U-Net.

5 Заключение

Реализованный подход позволяет строить архитектуру шумоподавляющей сети из модифицированной модели для задачи классификации. При таком построении нет необходимости использовать слои для декодирования скрытого пространства, построенного кодировщиком входного изображения, что позволяет сократить число параметров модели. Обученные модели демонстрируют высокую скорость обработки изображений при достижении качества подавления шума, сопоставимого с современными подходами.

Программный код с реализацией подхода и запуска обучения предлагаемой модели содержится в GitHub-репозитории по следующей URL-ссылке: https://github.com/AlexeySrus/WPNet.

Литература

1. *Ronneberger O., Fischer P., Brox T.* U-Net: Convolutional networks for biomedical image segmentation // Medical Image Computing and Computer-Assisted Intervention Proceedings. — Cham: Springer International Publishing, 2015. P. 234–241.
2. *Komatsu R., Gonsalves T.* Comparing U-Net based models for denoising color images // AI, 2020. Vol. 1. Iss. 4. P. 465–486. doi: 10.3390/ai1040029.
3. *Hasinoff S. W.* Saturation (imaging) // Computer vision: A reference guide. — Boston, MA, USA: Springer, 2014. P. 699–701. doi: 10.1007/978-0-387-31439-6_483.
4. *Wang Z., Bovic A., Sheikh H., Simoncelli E.* Image quality assessment: From error visibility to structural similarity // IEEE T. Image Process., 2004. Vol. 13. Iss. 4. P. 600–612. doi: 10.1109/TIP.2003.819861.
5. *Zhao H., Gallo O., Frosio I., Kautz J.* Loss functions for image restoration with neural networks // IEEE Transactions Computational Imaging, 2017. Vol. 3. Iss. 1. P. 47–57. doi: 10.1109/TCI.2016.2644865.
6. *Woo S., Jongchan P., Joon-Young L., In-So K.* CBAM: Convolutional block attention module. — Cornell University, 2018. 17 p. arXiv:1807.06521.
7. *Jiang J., Xiangming H., Zhao Y., Xu X., Cui Y.* SDAUNet: A simple dual attention mechanism UNet for mixed noise removal // IET Image Process., 2023. Vol. 17. Iss. 13. P. 3884–3896. doi: 10.1049/IPR2.12905.
8. *Liu P., Zhang H., Zhang K., Lin L., Zuo W.* Multi-level wavelet-CNN for image restoration // IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops Proceedings. — Los Alamitos, CA, USA: IEEE Computer Society, 2018. P. 886–895. doi: 10.1109/CVPRW.2018.00121.
9. *Batziou E., Ioannidis K., Patras I., Vrochidis S., Kompatsiaris I.* Low-light image enhancement based on U-Net and Haar wavelet pooling // MultiMedia modeling. — Cham: Springer Nature Switzerland, 2023. P. 510–522.
10. *Tian C., Zheng M., Zuo W., Zhang B., Zhang Y., Zhang D.* Multi-stage image denoising with the wavelet transform // Pattern Recogn., 2023. Vol. 134. Art. 109050. doi: 10.1016/j.patcog.2022.109050.
11. *Huang J.-J., Dragotti P. L.* WINNet: Wavelet-inspired invertible network for image denoising // IEEE T. Image Process., 2021. Vol. 31. P. 4377–4392.
12. *Guo T., Mousavi H., Vu T., Monga V.* Deep wavelet prediction for image super-resolution // Conference on Computer Vision and Pattern Recognition Workshops Proceedings. — Los Alamitos, CA, USA: IEEE Computer Society, 2017. P. 1100–1109. doi: 10.1109/CVPRW.2017.148.
13. *Qin X., Dai H., Hu X., Fan D.-P., Shao L., Van G.* Highly accurate dichotomous image segmentation // Computer vision. — Cham: Springer Nature Switzerland, 2022. P. 38–56.
14. *He K., Zhang H., Ren S., Sun J.* Deep residual learning for image recognition // Conference on Computer Vision and Pattern Recognition Proceedings. — Los Alamitos, CA, USA: IEEE Computer Society, 2016. P. 770–778. doi: 10.1109/CVPR.2016.90.
15. *Scherer D., Muller A., Behnke S.* Evaluation of pooling operations in convolutional architectures for object recognition // Artificial neural networks. — Berlin, Heidelberg: Springer, 2010. P. 92–101. doi: 10.1007/978-3-642-15825-4_10.
16. *Макаренко А. В.* Глубокие нейронные сети: зарождение, становление, современное состояние // Проблемы управления, 2020. Т. 2. С. 3–19. doi: 10.25728/pu.2020.2.1.
17. *Буй Т. Т. Ч., Спицын В. Г.* Разложение цифровых изображений с помощью двумерного дискретного вейвлет-преобразования и быстрого преобразования Хаара // Известия Томского политехнического университета, 2011. Т. 318. № 5. С. 73–76.
18. *Павлов А. Н.* Детектирование информационных сигналов на основе реконструкции динамических систем и дискретного вейвлет-преобразования // Известия высших учебных заведений. Прикладная нелинейная динамика, 2008. Т. 16. № 6. С. 3–17.
19. *Пронькин А.* Оценивание уровня шума в составе изображения с использованием вейвлетов Хаара // Труды 22-й Международной конференции по компьютерной графике и зрению. — М.: ИПМ им. М. В. Келдыша, 2022. Т. 32. С. 442–448. doi: 10.20948/graphicon-2022-442-448.
20. *Bradski G.* The OpenCV Library // Dr. Dobbs J., 2000. Vol. 25. Iss. 11. P. 120–125.
21. *Abdelhamed A., Lin S., Brown M. S.* A high-quality denoising dataset for smartphone cameras // IEEE/CVF Conference on Computer Vision and Pattern Recognition Proceedings. — Los Alamitos, CA, USA:

IEEE Computer Society, 2018. P. 1692–1700. doi: 10.1109/CVPR.2018.00182.

22. *Huang J.-B., Singh A., Ahuja N.* Single image super-resolution from transformed self-exemplars // Conference on Computer Vision and Pattern Recognition. — Los Alamitos, CA, USA: IEEE Computer Society, 2015. P. 5197–5206.
23. *Agustsson E., Timofte R.* NTIRE 2017 challenge on single image super-resolution: Dataset and study // Conference on Computer Vision and Pattern Recognition Workshops Proceedings. — Los Alamitos, CA, USA: IEEE Computer Society, 2017. P. 1110–1121. doi: 10.1109/CVPRW.2017.149.
24. *Martin D., Fowlkes C., Tal D., Malik J.* A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics // 8th Conference (International) on Computer Vision Proceedings. — Piscataway, NJ, USA: IEEE, 2001. Vol. 2. P. 416–423. doi: 10.1109/ICCV.2001.937655.
25. *Plotz T., Roth S.* Benchmarking denoising algorithms with real photographs // Conference on Computer Vision and Pattern Recognition Proceedings. — Los Alamitos, CA, USA: IEEE Computer Society, 2017. P. 2750–2759. doi: 10.1109/CVPR.2017.294.
26. *Paszke A., Gross S., Massa F., Lerer A.* PyTorch: An Imperative style, high-performance deep learning library // Adv. Neur. Inf., 2019. Vol. 32. P. 8024–8035.
27. *Wightman R.* PyTorch image models, 2019. https://github.com/rwightman/pytorch-image-models.
28. *Wang Z., Cun X., Bao J., Zhou W., Liu J., Li H.* Uformer: A general U-shaped transformer for image restoration // IEEE/CVF Conference on Computer Vision and Pattern Recognition Proceedings. — Los Alamitos, CA, USA: IEEE Computer Society, 2022. P. 17683–17693. doi: 10.1109/CVPR52688.2022.01716/
29. *Lu J.* AdaSmooth: An adaptive learning rate method based on effective ratio // Sentiment analysis and deep learning. — Singapore: Springer Nature Singapore, 2023. P. 273–293.
30. *Girshick R. B.* Fast R-CNN. — Cornell University, 2015. 9 p. arXiv:1504.08083.
31. *Dabov K., Foi A., Katkovnik V., Egiazarian K.* Image denoising by sparse 3-D transform-domain collaborative filtering // IEEE T. Image Process., 2007. Vol. 16. P. 2080–2095.
32. *Zhang K., Zuo W., Chen Y., Meng D., Zhang L.* Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising // IEEE T. Image Process., 2017. Vol. 26. Iss. 7. P. 3142–3155. doi: 10.1109/TIP.2017.2662206.

Поступила в редакцию 21.12.23

IMAGE DECOMPOSITION WITH DISCRETE WAVELET TRANSFORM TO DESIGN A DENOISING NEURAL NETWORK

A. S. Kovalenko

Institute of Mathematics, Mechanics, and Computer Science named after I. I. Vorovich, Southern Federal University, 105/42 Bolshaya Sadovaya Str., Rostov-on-Don 344006, Russian Federation

Abstract: Reducing noise in digital images is one of the most common tasks in image processing. At the moment, noise reduction approaches based on the applying of convolutional neural networks are widely used. In this case, as a rule, model training is based on minimizing the error function between the result of the network operation and the expected reference image and, additionally, various representations of the two-dimensional image signal and their properties are not used to optimize the training of noise reduction network architectures. The paper proposes an approach to training neural networks to suppress noise. The described approach is based on the usage of the N-fold fast Haar wavelet transform. This representation of a discrete image signal allows one to discard the classical architecture of the autoencoder and to use only its part that encodes the signal which leads to a significant reduction in model parameters and speeds up the network.

Keywords: neural networks; deep learning; image denoising; image processing

DOI: 10.14357/19922264240209 **EDN:** UEQSXP

References

1. Ronneberger, O., P. Fischer, and T. Brox. 2015. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention Proceedings*. Cham: Springer International Publishing. 234–241.
2. Komatsu, R., and T. Gonsalves. 2020. Comparing U-Net based models for denoising color images. *AI* 1(4):465–486. doi: 10.3390/ai1040029.
3. Hasinoff, S. W. 2014. Saturation (imaging). *Computer vision: A reference guide*. Boston, MA: Springer. 699–701. doi: 10.1007/978-0-387-31439-6_483.

4. Wang, Z., A. Bovic, H. Sheikh, and E. Simoncelli. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE T. Image Process.* 13(4):600–612. doi: 10.1109/TIP.2003.819861.
5. Zhao, H., O. Gallo, I. Frosio, and J. Kautz 2017. Loss functions for image restoration with neural networks. *IEEE Trans. Computational Imaging* 3(1):47–57. doi: 10.1109/TCI.2016.2644865.
6. Woo, S., P. Jongchan, L. Joon-Young, and K. In-So. 2018. CBAM: Convolutional block attention module. 17 p. Available at: https://arxiv.org/abs/1807.06521 (accessed April 28, 2024).
7. Jiang, J., H. Xiangming, Y. Zhao, X. Xu, and Y. Cui. 2023. SDAUNet: A simple dual attention mechanism UNet for mixed noise removal. *IET Image Process.* 17(13):3884–3896. doi: 10.1049/IPR2.12905.
8. Liu, P., H. Zhang, K. Zhang, L. Lin, and W. Zuo. 2018. Multi-level wavelet-CNN for image restoration. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops Proceedings*. Los Alamitos, CA: IEEE Computer Society. 886–895. doi: 10.1109/CVPRW.2018.00121.
9. Batziou, E., K. Ioannidis, I. Patras, S. Vrochidis, and I. Kompatsiaris. 2023. Low-light image enhancement based on U-Net and Haar wavelet pooling. *MultiMedia modeling*. Cham: Springer Nature Switzerland. 510–522.
10. Tian, C., M. Zheng, W. Zuo, B. Zhang, Y. Zhang, and D. Zhang. 2023. Multi-stage image denoising with the wavelet transform. *Pattern Recogn.* 134:109050. doi: 10.1016/j.patcog.2022.109050.
11. Huang, J.-J., and P. L. Dragotti. 2021. WINNet: Wavelet-inspired invertible network for image denoising. *IEEE T. Image Process.* 31:4377–4392.
12. Guo, T., H. Mousavi, T. Vu, and V. Monga. 2017. Deep wavelet prediction for image super-resolution. *Conference on Computer Vision and Pattern Recognition Workshops Proceedings*. Los Alamitos, CA: IEEE Computer Society. 1100–1109. doi: 10.1109/CVPRW.2017.148.
13. Qin, X., H. Dai, X. Hu, D.-P. Fan, L. Shao, and G. Van. 2022. Highly accurate dichotomous image segmentation. *Computer vision*. Cham: Springer Nature Switzerland. 38–56.
14. He, K., H. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. *Conference on Computer Vision and Pattern Recognition Proceedings*. Los Alamitos, CA. 770–778.
15. Scherer, D., A. Muller, and S. Behnke. 2010. Evaluation of pooling operations in convolutional architectures for object recognition. *Artificial neural networks*. Berlin, Heidelberg: Springer. 92–101. doi: 10.1007/978-3-642-15825-4_10.
16. Makarenko, A. V. 2020. Glubokie neyronnye seti: zarozhdenie, stanovlenie, sovremennoe sostoyanie [Deep neural networks: Origins, development, and current status]. *Problemy upravleniya* [Control Sciences] 2:3–19. doi: 10.25728/pu.2020.2.1.
17. Buy, T. T. C., and V. G. Spitsyn. 2011. Razlozhenie tsifrovykh izobrazheniy s pomoshch'yu dvumernogo diskretnogo veyvlet-preobrazovaniya i bystrogo preobrazovaniya Khaara [Digital image decomposition using two-dimensional discrete wavelet transform and fast Haar transform]. *Izvestiya Tomskogo politekhnicheskogo universiteta* [Bulletin of the Tomsk Polytechnic University] 318(5):73–76.
18. Pavlov, A. N. 2008. Detektirovanie informatsionnykh signalov na osnove rekonstruktsii dinamicheskikh sistem i diskretnogo veyvlet-preobrazovaniya [Detection of information signals based on reconstruction of dynamical systems and discrete wavelet-transform]. *Izvestiya vysshikh uchebnykh zavedeniy. Prikladnaya nelineynaya dinamika* [Bulletin of Higher Educational Institutions. Applied Nonlinear Dynamics] 16(6):3–17.
19. Pronkin, A. V. 2022. Otsenivanie urovnya shuma v sostave izobrazheniya s ispol'zovaniem veyvletov Khaara [Noise level estimation in images using Haar wavelets]. *22nd Conference (International) on Computer Graphics and Computer Vision Proceedings*. Moscow: IPM im. M. V. Keldysha. 32:442–448. doi: 10.20948/graphicon-2022-442-448.
20. Bradski, G. 2000. The OpenCV Library. *Dr. Dobbs J.* 25(11):120–125.
21. Abdelhamed, A., S. Lin, and M. S. Brown. 2018. A high-quality denoising dataset for smartphone cameras. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Proceedings*. Los Alamitos, CA: IEEE Computer Society. 1692–1700. doi: 10.1109/CVPR.2018.00182.
22. Huang, J.-B., A. Singh, and N. Ahuja. 2015. Single image super-resolution from transformed self-exemplars. *Conference on Computer Vision and Pattern Recognition Proceedings*. Los Alamitos, CA: IEEE Computer Society. 5197–5206.
23. Agustsson, E., and R. Timofte. 2017. NTIRE 2017 challenge on single image super-resolution: Dataset and study. *Conference on Computer Vision and Pattern Recognition Workshops Proceedings*. Los Alamitos, CA: IEEE Computer Society. 1110–1121. doi: 10.1109/ CVPRW.2017.149.
24. Martin, D., C. Fowlkes, D. Tal, and J. Malik. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *8th Conference (International) on Computer Vision Proceedings*. Piscataway, NJ: IEEE Computer Society. 2:416–423. doi: 10.1109/ICCV.2001.937655.
25. Plotz, T., and S. Roth. 2017. Benchmarking denoising algorithms with real photographs. 2017. *Conference on Computer Vision and Pattern Recognition Proceedings*. Los Alamitos, CA: IEEE Computer Society. 2750–2759. doi: 10.1109/CVPR.2017.294.
26. Paszke, A., S. Gross, F. Massa, and A. Lerer. 2019. PyTorch: An imperative style, high-performance deep learning library. *Adv. Neur. Inf.* 32:8024–8035.

27. Wightman, R. 2019. PyTorch image models. Available at: https://github.com/rwightman/pytorch-image-models (accessed April 28, 2024).
28. Wang, Z., X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li. 2022. Uformer: A general U-shaped transformer for image restoration. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Proceedings*. Los Alamitos, CA: IEEE Computer Society. 17683–17693.
29. Lu, J. 2023. AdaSmooth: An adaptive learning rate method based on effective ratio. *Sentiment analysis and deep learning*. Singapore: Springer Nature Singapore. 273–293.
30. Girshick, R. B. 2015. Fast R-CNN. 9 p. Available at: https://arxiv.org/abs/1504.08083 (accessed April 28, 2024).
31. Dabov, K., A. Foi, V. Katkovnik, and K. Egiazarian. 2007. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE T. Image Process.* 16:2080–2095.
32. Zhang, K., W. Zuo, Y. Chen, D. Meng, and L. Zhang. 2017. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE T. Image Process.* 26(7):3142–3155. doi: 10.1109/TIP.2017.2662206.

Received December 21, 2023

Contributor

Kovalenko Alexey S. (b. 1996) — assistant, Department of Applied Mathematics and Programming, Institute of Mathematics, Mechanics, and Computer Science named after I. I. Vorovich, Southern Federal University, 105/42 Bolshaya Sadovaya Str., Rostov-on-Don 344006, Russian Federation; akov@sfedu.ru

О ПРИМЕНЕНИИ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ В СИСТЕМЕ ЭЛЕКТРОННОГО ОБУЧЕНИЯ МАТЕМАТИЧЕСКИМ ДИСЦИПЛИНАМ*

А. В. Босов[1], А. В. Иванов[2]

Аннотация: Существующие инструменты динамического формирования индивидуальной траектории обучения дополнены технологией генерации аттестационных заданий и экзаменационных билетов. В качестве источника качественных, сбалансированных наборов заданий использован комплект экзаменационных билетов, специально подготовленный экспертами по вузовскому курсу теории функций комплексного переменного. Этот значительный обучающий массив качественных аттестационных заданий существенно расширил имеющиеся данные, созданные на предыдущих этапах. Цель выполненного исследования состояла в создании методов, позволяющих учитывать знания экспертов, заложенные в имеющемся комплекте заданий. Реализованная модель генерации при обработке образовательного контента использует в качестве параметров атрибуты, назначенные экспертами задачам: тематику, сложность, формируемые компетенции. Предложены два метода генерации. Первый — вероятностный — использует только частотные характеристики обучающего комплекта, аппроксимируя распределение вероятностей. Второй базируется на генеративно-состязательных нейронных сетях. Особое внимание уделено обсуждению трудностей реализации сети, связанных в числе прочего со специфическим характером генеративной модели.

Ключевые слова: электронная обучающая система; образовательный контент; машинное обучение; генеративные модели; имитационное компьютерное моделирование; генеративно-состязательные сети

DOI: 10.14357/19922264240210 **EDN:** UWKQLN

1 Введение

Электронные обучающие системы (ЭОС) помимо того, что стали типовым инструментом образовательного процесса, оказались богатым источником постановок задач для прикладных и фундаментальных исследований. Общеизвестный пример — управление тестированием обучаемых, оформившееся в самостоятельное направление (теория тестирования, item response theory, IRT) [1–6], небольшой вклад в которое внесен и авторами [7, 8]. Практическим результатом методов IRT становится индивидуальная траектория обучения, сформированная из подобранных системой заданий — образовательного контента, адаптированного под контингент обучаемых. Не будет ошибкой считать теорию тестирования частью более широкого направления рекомендательных систем [9, 10], известных приложениями не только в области электронного обучения. Методы рекомендаций, в частности самообучающиеся карты, также исследовались в связи с применениями в ЭОС как авторами [11], так и другими исследователями [12, 13]. Если инструментами IRT, как правило, служат методы теории вероятностей, то рекомендательным системам в целом более свойственны методы машинного обучения.

Индивидуальная траектория, вне зависимости от того, какой конкретный смысл в нее вкладывается и какие инструменты ее формируют, несет еще и специфическую окраску уникальности, новизны. Используя имеющийся образовательный контент, взаимодействуя со средой, с обучаемыми и преподавателями, система формирует новый контент, до сих пор не существовавший. Это может быть очень простой контент, например последовательность тестов очередного зачета, или более сложный — прогноз результатов обучения группы из сотни учащихся. Но это контент новый, а значит, в области электронного обучения есть место для применения современных инструментов машинного обучения, генерирующих контент. Классификаторы, реализующие упомянутый выше функционал индивидуального обучения как методики, имеют выраженный дискриминационный характер, так как результат классификации определен

*Исследование выполнено за счет гранта Российского научного фонда № 22-28-00588, https://rscf.ru/project/22-28-00588/. Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, avbosov@ipiran.ru

[2]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, aivanov@ipiran.ru

однозначно, никакой вариативности решений не предусматривается. Действительно, генеративные методы порождают новый контент «из ничего», без ограничений, без очевидных повторений. Простые модели — гауссовские смеси, скрытые марковские процессы, скрытое распределение Дирихле — и вычислительно сложные — ограниченные машины Больцмана и глубокие сети доверия, а также самые совершенные генеративно-состязательные сети [14] — концептуально направлены не на выбор из имеющихся альтернатив, а на создание новой, другой альтернативы. Процесс создания такого контента имеет характер креативный, что подтверждают и приложения, в которых эти модели применяются. Это искусство, синтез текстов и речи, дизайн, игры и т. п. Сейчас всему этому уделяется внимания едва ли не больше, чем традиционным приложениям нейронных сетей, компьютерному зрению, обработке текстов, биометрии. Цель данной статьи — исследовать применимость генеративных моделей в прикладной области электронного обучения и провести эксперимент с теми данными, что были сформированы в предыдущих работах [15, 16].

2 Генеративная модель экзаменационного билета

Формально ставится задача из имеющегося комплекта наборов заданий — задач для самостоятельного решения — сформировать экзаменационный билет (аттестационное задание). Используется подготовленный ранее набор заданий по курсу «Теория функций комплексного переменного» [15] (ТФКП, 750 задач, 9 тематических разделов, числовая разметка по уровням сложности от 1 до 7 и набор формируемых компетенций — множество компетенций пронумеровано числовыми значениями от 1 до 20, каждое задание имеет набор из нескольких компетенций, всего получается 109 уникальных комбинаций). Нужно сгенерировать билет, состоящий из 5 задач. Набор задач билета должен быть «сбалансированным», а в качестве описания сбалансированности экспертами (авторами пособия [17]) подготовлены 100 образцов билетов.

2.1 Применение языковой модели

На первый взгляд довольно легко получить требуемый результат, воспользовавшись готовыми технологическими решениями. Например, известной языковой моделью и генеративной сетью ChatGPT (https://chat.openai.com). Версия 3.5 общедоступна и обучена на огромном объеме текстов, содержащем и много математических пособий, в том числе по ТФКП. Итоги экспериментирования с ChatGPT по вопросу подготовки экзаменационных билетов из пяти задач иллюстрирует рис. 1. На этом рисунке показаны два (первый и четвертый) образцы, подготовленные в диалоге с ChatGPT. Как видно, первый результат выглядит вполне содержательным билетом по ТФКП, по крайней мере содержит задачи именно по этой дисциплине, и они достаточно разнообразны. Можно обсуждать, что билет не

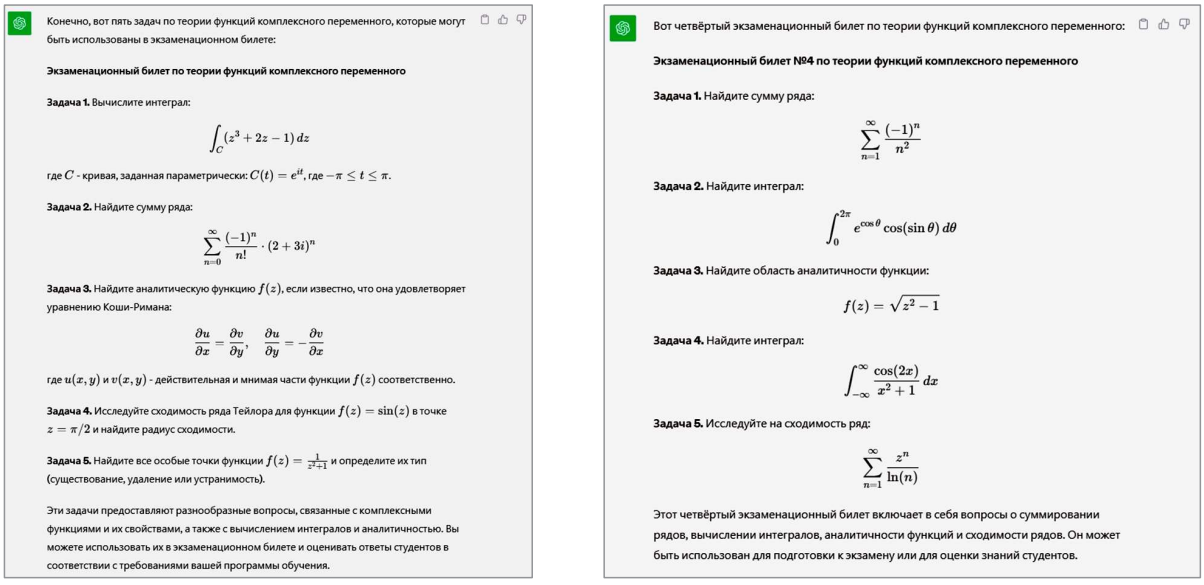


Рис. 1 Билеты, подготовленные сетью ChatGPT 3.5

слишком сложный, хотя понятие «сложность билета» требует уточнения. Задание усложнить и разнообразить задачи быстро привело к результату — второму билету на рис. 1. Здесь сразу обращает на себя внимание то, что это уже билет не по ТФКП.

Таким образом, несмотря на колоссальный объем обучающей выборки, на грандиозный размер сети, на тематические «подкрепления» в формируемых запросах, предложить билеты чат-бот не смог. Возможная проблема здесь состоит том, что в распоряжении сети нет размеченного набора образцов заданий. Но более глубокой и сложной проблемой оказалось то, что сеть не обучена (и, наверное, с используемой ею языковой моделью, основанной на частотном анализе, не может быть обучена) пониманию «сбалансированного билета».

Вместе с тем данная иллюстрация не дает основания отказаться от самой идеи применять в рассматриваемой задаче генеративные модели. В тех приложениях, где эти модели дают хороший результат, можно увидеть общее свойство — отсутствие четкого критерия оценки результата генерации. Новый текст, новое изображение, новая мелодия одинаково ценны и новизной, и «сбалансированностью». Объективной оценки результата не может быть, решающий вклад вносит дискриминационный компонент, обученный на «удачных» примерах. Ровно такая же ситуация, хотя и с гораздо более скромными объемными характеристиками, имеет место и в рассматриваемой задаче. Поэтому генеративные модели, способные обучаться «сбалансированности» на подготовленных экспертами примерах билетов, должны давать гораздо лучший результат.

2.2 Применение вероятностной модели

К рассматриваемой задаче можно применить традиционные методы статистического моделирования типа гауссовской смеси. Непосредственно гауссовские распределения применить нельзя из-за дискретного характера реализаций всех имеющихся атрибутов, но аналогичные частотные методы технологически дадут такое же решение. Для построения вероятностной модели проиллюстрируем имеющиеся образцы «сбалансированных» билетов некоторыми статистическими показателями (см. таблицу).

В таблице указаны: (1) число заданий по теме, использованных в билетах, и общее число заданий по теме в имеющемся наборе 750 (в задаче 1 по теме 2 дополнительно число уникальных, неповторяющихся заданий); (2) сложности, встречающиеся в билетах, и в скобках — сложности по всем заданиям набора 750 по данной теме; (3) компетенции, встречающиеся в билетах, и в скобках — все компетенции по теме по всем заданиям набора 750. Отметим, что все задания, выбранные экспертами по темам 1, 3–9, встречаются ровно по одному разу. Повторное использование заданий встречается только по теме 2 в первой задаче билета.

Показанные характеристики можно уточнять далее. Так, интерес представляют распределения заданий в рамках каждой темы по отдельным ком-

Частотные характеристики обучающей выборки билетов

Задача 1		Задача 2	Задача 3		Задача 4		Задача 5	
Тема 1	Тема 2	Тема 3	Тема 4	Тема 5	Тема 6	Тема 7	Тема 8	Тема 9
11 (из 56)	89 (74) (из 99)	100 (из 151)	50 (из 92)	50 (из 72)	16 (из 30)	84 (из 109)	39 (из 39)	61 (из 102)
Сложность по корпусу 100 билетов								
1, 2 (2, 9)	2 (89)	2, 3 (28, 72)	2, 3, 4 (4, 35, 11)	4, 5 (6, 44)	3, 4 (4, 12)	5, 6 (78, 6)	4, 5, 6 (1, 19, 19)	5, 6, 7 (2, 33, 26)
Сложность по корпусу 750 заданий								
1, 2 (39, 17)	2 (99)	2, 3 (69, 82)	2, 3, 4 (7, 63, 22)	4, 5 (6, 66)	3, 4 (17, 13)	5, 6 (99, 10)	4, 5, 6 (1, 19, 19)	5, 6, 7 (2, 61, 39)
Компетенции по корпусу 100 билетов								
1, 2, 3 (3, 3, 9)	1, 2, 3, 4 (23, 3, 63, 86)	1, 3, 4, 5, 6 (52, 4, 36, 79, 32)	7, 8, 9, 10, 14 (27, 26, 4, 20, 2)	11 50	11, 12 (5, 16)	11, 12, 13, 14 (21, 79, 84, 71)	15, 16 (27, 12)	17, 18, 19 (49, 10, 2)
Компетенции по корпусу 750 заданий								
1, 2, 3 (13, 23, 40)	1, 2, 3, 4 (28, 3, 63, 96)	1, 3, 4, 5, 6 (71, 8, 42, 114, 52)	7, 8, 9, 10, 14 (45, 46, 7, 40, 4)	11 72	11, 12 (7, 30)	11, 12, 13, 14 (31, 100, 109, 89)	15, 16 (27, 12)	17, 18, 19, 20 (72, 11, 4, 15)

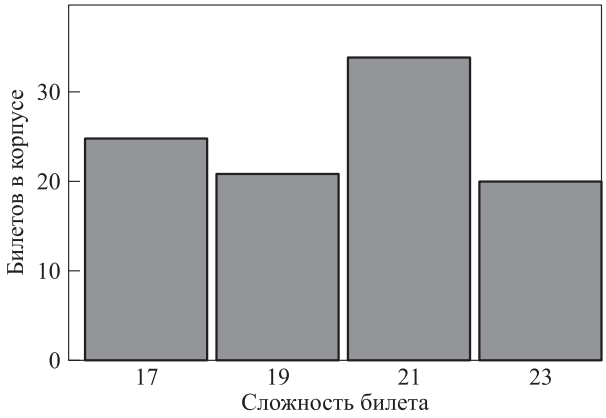


Рис. 2 Распределение билетов обучающей выборки по сложности

петенциям: можно сравнивать эти распределения на всех заданиях, имеющихся в корпусе, и на заданиях, вошедших в билеты. Также интересно общее распределение билетов по суммарной сложности задач, которое можно считать сложностью билета (рис. 2).

Для моделирования билета сначала нужно определить весь набор определяющих билет атрибутов, т. е. тематику–сложность–компетенции всех пяти задач. Для этого строится случайный вектор. Для каждой i-й задачи билета имеем атрибуты T_i, D_i и C_i, $i = \overline{1,5}$. Здесь $T_i \in \{t_j\}_{j=1}^{9} = \{1,\ldots,9\}$ — одна из имеющихся 9 тематик; $D_i \in \{d_j\}_{j=1}^{7} = \{1,\ldots,7\}$ — один из имеющихся 7 уровней сложности; $C_i \in \{c_j\}_{j=1}^{109} = \{1,\ldots,109\}$ — одна из имеющихся 109 групп компетенций (данный параметр представлен 20-мерным вектором, каждый компонент которого, принимая значение 1 или 0, определяет принадлежность задачи к соответствующей компетенции).

Предполагая все эти атрибуты случайными, получим модель билета — случайный вектор

$$\text{E\_ticket} = \operatorname{col}(T_1, D_1, C_1, T_2, \ldots, C_5). \qquad (1)$$

В качестве упрощенной альтернативы можно будет при необходимости использовать упрощенную модель без компетенций

$$\text{E\_simple\_ticket} = \operatorname{col}(T_1, D_1, T_2, \ldots, D_5). \qquad (2)$$

Набор возможных значений E_ticket обозначим $\{e_j\}_{j=1}^{L_e}$, E_simple_ticket — через $\{s_j\}_{j=1}^{L_s}$. Величины L_e и L_s определяются имеющейся обучающей выборкой, т. е. набором билетов, подготовленных экспертами. Имеющийся набор дает $L_e = 81$ и $L_s = 26$. Далее очевидным образом вычисляем частоты p_j^e, $j = \overline{1, L_e}$, и p_j^s, $j = \overline{1, L_s}$, реализации значений e_j и s_j в обучающем наборе билетов. Соответственно, алгоритм генерации нового билета состоит из трех шагов.

Шаг 1. Моделирование реализации случайного вектора, имеющего распределение $\{(e_j, p_j^e)\}_{j=1}^{L_e}$ или $\{(s_j, p_j^s)\}_{j=1}^{L_s}$.

Шаг 2. Отбор заданий, удовлетворяющих смоделированной реализации. Для этого по значению e_j или s_j определяются реализовавшиеся значения T_i, D_i и C_i, затем из имеющегося блока 750 заданий выбираются множества возможных заданий последовательно для i-й задачи билета.

Шаг 3. Случайным образом выбираются задания для задач билета из определенного на шаге 2 множества возможных заданий.

На шаге 3 используется равновероятное распределение, во-первых, из-за того, что иных оснований вероятностный подход предоставить не может, во-вторых, затем, чтобы использовать весь корпус 750 заданий и не ограничиваться только примерами, выбранными экспертами.

При такой генерации билетов учитываются статистические зависимости между тематиками, сложностями, компетенциями, т. е. предполагается, что эти зависимости — это именно те знания, которые выразили эксперты в обучающем наборе 100 билетов, и то, что выявлено частотами p_j^e, $j = \overline{1, L_e}$, и p_j^s, $j = \overline{1, L_s}$.

Несомненным достоинством здесь выглядят простота и объяснимость метода. Ключевым недостатком — ограниченность интерпретации знаний экспертов в отношении понятия «сбалансированности» только размеченными атрибутами тематика–сложность–компетенции. На самом деле, даже для неспециалиста в ТФКП при изучении 100 образцов билетов становится очевидным, что при их формировании эксперт вкладывал много «смыслов», не формализуемых в терминах тематика–сложность–компетенции, и извлечь эти смыслы вероятностный подход не способен. Поэтому большего ожидается от применения более креативных методов.

2.3 Генеративно-состязательная сеть

Как известно [18], архитектура генеративно-состязательной сети состоит из двух взаимодействующих подсетей. Первая реализует собственно генеративную модель, вторая — дискриминационную модель. Исходными образцами для генеративной модели выступают все 100 подготовленных экспертами билетов. Модель состоит в свертке смоделированного образца до вектора атрибутов (1) (модель (2) для этого метода представляется совсем малоинтересной).

Для реализации использовалась библиотека Keras (https://keras.io), предоставляющая удобную надстройку над фреймворком Tenzorflow, и описание ее автора [19]. И генеративная, и дискриминационная сети представлены сверточными сетями. Обращает внимание то, что в Keras в сравнении с использованной в предыдущих работах [15, 16] более простой библиотекой Scikit-learn (https://scikit-learn.org) уровень абстракции ниже, так что реализация сети требует существенных усилий разработчика.

После проведения серии экспериментов размерность вектора скрытого пространства была определена равной 5, значения получаются от генератора псевдослучайных чисел со стандартным нормальным распределением. Попытки увеличения числа латентных переменных до 20 не привели к сколь-нибудь значимому улучшению качества модели.

В отличие от вероятностной модели, использующей в модели (1) 20-мерный вектор для представления атрибутов компетенций, обучающие данные и данные на выходе генеративной сети представлены пятью 12-мерными векторами (каждый вектор соответствует одному заданию билета). Первый элемент вектора соответствует теме T_i, второй — сложности задания D_i, а для третьего использована свертка: каждая пара соседних бинарных элементов вектора компетенций заменена одним четырехзначным числом по формуле:

$$C_k = 0{,}25C_i + 0{,}5C_{i+1}, \quad k = \overline{1,10}\,.$$

Это сделано, чтобы уменьшить размерность, что очень важно для генеративной сети. После подготовки все элементы обучающей выборки были нормализованы.

Поскольку значения элементов обучающей выборки фактически дискретны, а значения на выходе генератора непрерывны, для последующей интерпретации результатов работы модели требуется дискретизация выходных значений генератора.

Генератор содержит 4 слоя:

(1) полносвязный слой с 60 выходами и функцией активации «линейный выпрямитель с утечкой» (LeakyReLU, leaky rectified linear unit) [20];

(2) одномерный сверточный слой с размером ядра свертки 2, содержащий 12 фильтров и использующий функцию активации LeakyReLU (для связи с предыдущим слоем полученный от него 60-мерный вектор преобразуется в пять 12-мерных векторов);

(3) одномерный слой транспонированной свертки с размером ядра 5, содержащий 12 фильтров и использующий функцию активации LeakyReLU;

(4) одномерный слой транспонированной свертки с размером ядра 5, содержащий 12 фильтров и использующий в качестве функции активации гиперболический тангенс.

Дискриминатор содержит три слоя:

(1) одномерный сверточный слой с размером ядра свертки 2, содержащий 48 фильтров и использующий функцию активации LeakyReLU;

(2) одномерный сверточный слой с размером ядра свертки 5, содержащий 16 фильтров и использующий функцию активации LeakyReLU;

(3) полносвязный слой с одним выходом, использующий сигмоидную функцию активации.

В ходе работы с сетью проявились две наиболее известные трудности, характерные для обучения генеративных моделей [21, 22]:

– отказ сходимости модели, проявляющийся в постоянном росте потерь генератора по мере роста числа эпох обучения при крайне низком качестве генерируемых данных;

– коллапс режима, проявляющийся в низкой вариативности генерируемых данных, а именно: склонности модели генерировать одни и те же данные при разных значениях вектора скрытого пространства.

Наибольшую эффективность при борьбе с отказом сходимости модели показали следующие приемы:

(1) замена оптимизатора с RMSProp (https://keras.io/api/optimizers/rmsprop) на Adam (https://keras.io/api/optimizers/adam);

(2) замена инициализатора весов модели с Glorot-Uniform [22] на HeNormal [23];

(3) применение прореживания в дискриминаторе [24];

(4) ограничение общей сложности модели (эксперименты показали, что на имеющихся данных крайне трудно обеспечить сходимость модели, если число параметров превышает 8000).

Также можно отметить, что попытка применить l_1-регуляризацию к каждому слою хотя и обеспечивает сходимость модели, но одновременно приводит к коллапсу режима и генерации моделью одного единственного задания.

Наибольшая трудность была связана с возникновением полного или частичного коллапса режима модели. Это может быть обусловлено как особенностями обучающих данных, так и выбранной

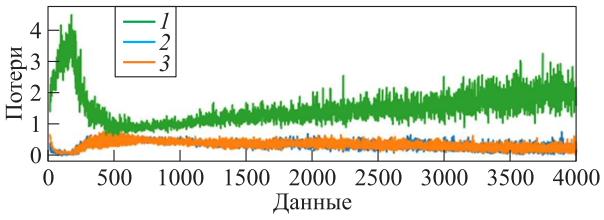


Рис. 3 Потери генератора (*1*) и дискриминатора (*2* — сгенерированные данные; *3* — обучающие данные)

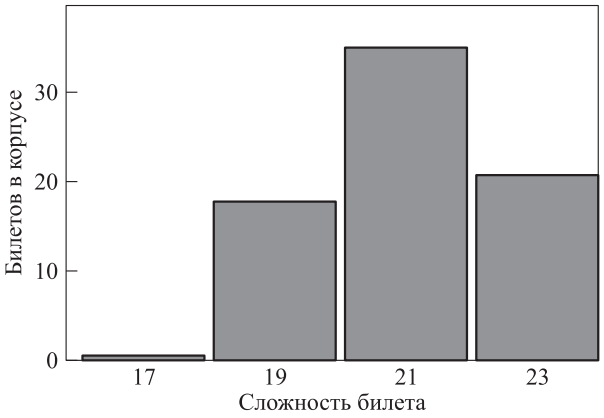


Рис. 4 Распределение билетов смоделированных выборок по сложности

архитектурой модели. В качестве приемов, помогающих повысить вариативность выходных данных модели, можно отметить:

(1) использование в генераторе слоев транспонированной свертки;

(2) увеличение числа слоев генератора;

(3) увеличение числа эпох обучения, что, в свою очередь, требует обеспечения высокой устойчивости модели в целом и предотвращения ее переобучения.

На рис. 3 приведены графики потерь генератора и дискриминатора на обучающих и сгенерированных данных для наиболее «продуктивного» варианта модели. Видно, что после первоначальной сходимости на 500-м цикле обучения (100 эпох) наблюдается медленный рост потерь генератора, т. е. данная модель оказывается неустойчивой. Тем не менее ее «продуктивность» растет. Поведение, схожее демонстрируемому генеративной моделью в ходе обучения, имеет один известный радиотехнический прибор — регенеративный радиоприемник. Такой приемник содержит один усилительный каскад, охваченный положительной обратной связью. Этот каскад может быть настолько же эффективен, как несколько усилительных каскадов без обратной связи. Однако его устойчивость крайне низка, так как определенная глубина обратной связи превращает усилитель в генератор. При этом наибольшая эффективность такого усилителя достигается именно на пороге генерации.

В итоге, используя имеющийся обучающий набор из 100 билетов и размеченный набор 750 задач, удается получить рабочую генеративную модель. Обсуждать ее качество, по-видимому, имеет смысл только экспертам. Некоторые вспомогательные показатели обсуждаются в следующем разделе статьи. Для примера и сравнения с вероятностной моделью подразд. 2.2 для смоделированного генеративной моделью набора билетов на рис. 4 приведена гистограмма распределения суммарной сложности билетов. Ее надо сравнивать с рис. 2, и отличия очевидны. Генеративная модель не дает свойственного вероятностной модели равномерного распределения сложностей. Вообще диапазон сложностей заужен и в «середине» есть сгущение. Объяснить причины, по которой генеративная модель игнорирует самые простые задания и увеличивает частоту заданий «средней» сложности, не представляется возможным. Но такое поведение выглядит более «честным» по отношению к студентам, хотя понять, как об этом «догадалась» сеть, не представляется возможным.

3 Варианты экспертной интерпретации результатов

Последним этапом для генерации должна быть оценка качества сгенерированных билетов. Формально этот этап должен проанализировать два набора:

(1) реализации обучающей выборки $\{\mathrm{E\_ticket}_k^{(1)}\}_{k=1}^{100}$;

(2) смоделированные билеты $\{\mathrm{E\_ticket}_k^{(2)}\}_{k=1}^{N}$ численностью N, потенциально много большей, чем 100.

Уже упоминалось, что объективных сравнительных показателей для этих наборов нет. Но можно предложить ряд типовых статистических характеристик, которые могут помочь в оценке. При необходимости будем считать, что $\mathrm{E\_ticket}_k^{(1)}$ и $\mathrm{E\_ticket}_k^{(2)}$ заменены соответствующими скалярами $\{e_j\}_{j=1}^{L_e}$.

Непарные корреляции — самый простой способ качественной оценки результатов моделирования с точки зрения сохранения важных зависимостей. Коэффициенты корреляции можно вычислять для значений любых двух атрибутов, например между

компетенциями пятой задачи C_5 и сложностями первой T_1:

$$\begin{aligned}&\text{корр}(C_5, T_1) = \\ &= \left(\sum_{k=1}^{K}\langle C_5\rangle_k\langle T_1\rangle_k - \left(\sum_{i=1}^{N}\langle C_5\rangle_k\right)\left(\sum_{i=1}^{N}\langle T_1\rangle_k\right)\right)\times \\ &\quad\times\left(\left(\sum_{k=1}^{K}\left(\langle C_5\rangle_k\right)^2 - \left(\sum_{k=1}^{K}\langle C_5\rangle_k\right)^2\right)\times\right. \\ &\quad\left.\times\left(\sum_{k=1}^{K}\left(\langle T_1\rangle_k\right)^2 - \left(\sum_{k=1}^{K}\langle T_1\rangle_k\right)^2\right)\right)^{-1/2}. \quad (3)\end{aligned}$$

Здесь $K = 100$, если корреляция $\text{корр}^{(1)}(C_5, T_1)$ считается на обучающем наборе, тогда $\langle C_5\rangle_k$ и $\langle T_1\rangle_k$ — реализации соответствующих компонентов векторов $\text{E\_ticket}_k^{(1)}$. Для смоделированного набора $K = N$ и корреляция $\text{корр}^{(2)}(C_5, T_1)$ считается на значениях $\langle C_5\rangle_k$ и $\langle T_1\rangle_k$ — реализациях компонентов векторов $\text{E\_ticket}_k^{(2)}$. Такие парные корреляция могут быть вычислены для любых пар атрибутов T_i, D_i и C_i, $i = \overline{1,5}$. Эти величины, принимающие значения от 0 до 1, характеризуют степень линейной зависимости значений, принимаемых атрибутами (в приведенном примере C_5 и T_1). Вычисленные на двух наборах величины $\text{корр}^{(1)}(C_5, T_1)$ и $\text{корр}^{(2)}(C_5, T_1)$ должны сравниваться, их близость можно интерпретировать как подтверждение сбалансированности смоделированного набора билетов по отношению к обучающему. Таких парных корреляций можно выбрать много. Каким отдавать предпочтение, не вполне понятно, поэтому предлагается использовать в анализе непарные корреляции, а именно: из значений тематик сформировать вектор тематик $T = \text{col}\,(T_1, \ldots, T_5)$, аналогично сложностей $D = \text{col}\,(D_1, \ldots, D_5)$ и компетенций $C = \text{col}\,(C_1, \ldots, C_5)$. Далее определить множества возможных значений каждого из трех векторов и заменить их нумерующими скалярами $\{t_j\}$, $\{d_j\}$ и $\{c_j\}$ так же, как при формировании скалярного варианта E_ticket. Теперь можно снова вернуться к парным корреляциям (3) и вычислить $\text{корр}^{(1)}(T, D)$, $\text{корр}^{(1)}(T, C)$ и $\text{корр}^{(1)}(D, C)$ и сравнить с $\text{корр}^{(2)}(T, D)$, $\text{корр}^{(2)}(T, C)$ и $\text{корр}^{(2)}(D, C)$.

Расстояние между распределениями. Более интегрированный анализ дает расстояние Кульбака–Лейблера [25], которое традиционно используют как числовую оценку меры сходства/разнообразия между распределениями вероятностей двух случайных векторов.

Распределение по обучающему набору билетов $\text{E\_ticket}_k^{(1)}$ дает значения и частоты $\{(e_j, p_j^{(1)})\}_{j=1}^{L_e}$, $p_j^{(1)} = p_j^e$, распределение смоделированного набора $\text{E\_ticket}_k^{(2)}$ дает значения и частоты $\{(e_j, p_j^{(2)})\}_{j=1}^{Le}$, частоты $p_j^{(2)}$ определяются аналогично p_j^e, но уже после моделирования и изменяются от эксперимента к эксперименту, зависят от объема N смоделированных данных так, что возможно для некоторых j $p_j^{(2)} = 0$, в то время как $p_j^{(1)} \neq 0\ \forall\, j$.

Расстоянием Кульбака–Лейблера по отношению к распределению $\{(e_j, p_j^{(1)})\}_{j=1}^{L_e}$ называется величина

$$\text{раст}\left(\text{E\_ticket}^{(2)}, \text{E\_ticket}^{(1)}\right) = \sum_{l=0}^{L-1} p_j^{(2)} \ln\left(\frac{p_j^{(2)}}{p_j^{(1)}}\right). \quad (4)$$

Заметим, что эта величина несимметрична, неотрицательна и равенство

$$\text{раст}\left(\text{E\_ticket}^{(2)}, \text{E\_ticket}^{(1)}\right) = \text{раст}\left(\text{E\_ticket}^{(1)}, \text{E\_ticket}^{(2)}\right)$$

означает $p_j^{(1)} = p_j^{(2)}\ \forall\, j$ и равенство расстояния нулю. Масштабной характеристики величина (4) не имеет, так что степень близости к 0 определяется экспериментально.

В качестве показателя качества смоделированного набора билетов можно использовать в дополнение к $\text{раст}\,(\text{E\_ticket}^{(2)}, \text{E\_ticket}^{(1)})$ и величину $\text{раст}\,(\text{E\_simple\_ticket}^{(2)}, \text{E\_simple\_ticket}^{(1)})$, вычисляемую по набору $\{(s_j, p_j^s)\}_{j=1}^{L_s}$.

В отношении любых предложенных показателей качественной оценки результатов моделирования надо учитывать их ограниченность. В случае использования вероятностной модели подразд. 2.2 эти величины просто показывают качество компьютерного моделирования заданного дискретного распределения, т. е. дадут заведомо хорошие результаты, превосходящие генеративную модель разд. 2.3. С другой стороны, потенциал и учитываемые «смыслы» генеративной модели значительно больше, поэтому основной объективный вывод, который дадут эти расчеты, — насколько важно для генеративной модели учитывать частотные зависимости, вложенные экспертами, составившими обучающий набор.

Литература

1. *Rasch G.* Probabilistic models for some intelligence and attainment tests. — Chicago, IL, USA: The University of Chicago Press, 1980. 224 p.

2. *Van der Linden W. J., Scrams D. J., Schnipke D. L., et al.* Using response-time constraints to control for differential speededness in computerized adaptive testing // Appl. Psych. Meas., 1999. Vol. 23. Iss. 3. P. 195–210. doi: 10.1177/01466219922031329.
3. *Chen C.-M., Lee H.-M., Chen Y.-H.* Personalized e-learning system using Item Response Theory // Computers Education, 2005. Vol. 44. No. 3. P. 237–255.
4. *Кибзун А. И., Иноземцев А. О.* Оценивание уровней сложности тестов на основе метода максимального правдоподобия // Автоматика и телемеханика, 2014. № 4. С. 20–37.
5. *Kuravsky L. S., Margolis A. A., Marmalyuk P. A., Panfilova A. S., Yuryev G. A., Dumin P. N.* A probabilistic model of adaptive training // Applied Mathematical Sciences, 2016. Vol. 10. Iss. 48. P. 2369–2380. doi: 10.12988/ams.2016.65168.
6. *Наумов А. В., Мхитарян Г. А.* О задаче вероятностной оптимизации для ограниченного по времени тестирования // Автоматика и телемеханика, 2016. № 9. С. 124–135.
7. *Босов А. В., Мхитарян Г. А., Наумов А. В., Сапунова А. П.* Использование модели гамма-распределения в задаче формирования ограниченного по времени теста в системе дистанционного обучения // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 12–18. doi: 10.14357/19922264190402. EDN: XUBLZX.
8. *Босов А. В., Мартюшова Я. Г., Наумов А. В., Сапунова А. П.* Байесовский подход к построению индивидуальной траектории пользователя в системе дистанционного обучения // Информатика и её применения, 2020. Т. 14. Вып. 3. С. 89–96. doi: 10.14357/19922264200313. EDN: WAKFJR.
9. *Adomavicius G., Tuzhilin A.* Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions // IEEE T. Knowl. Data En., 2005. Vol. 17. No. 6. P. 734–749. doi: 10.1109/TKDE.2005.99.
10. *Verbert K., Manouselis N., Ochoa X., Wolpers M., Drachsler H., Bosnic I., Duval E.* Context-aware recommender systems for learning: A survey and future challenges // IEEE T. Learn. Technol., 2012. Vol. 5. No. 4. P. 318–335. doi: 10.1109/TLT.2012.11.
11. *Босов А. В.* Применение самоорганизующихся нейронных сетей к процессу формирования индивидуальной траектории обучения // Информатика и её применения, 2022. Т. 16. Вып. 3. С. 7–15. doi: 10.14357/19922264220302. . EDN: HJQANN.
12. *Tai D. W. S., Wu H. J., Li P. H.* Effective e-learning recommendation system based on self-organizing maps and association mining // Electron. Libr., 2008. Vol. 26. No. 3. P. 329–344.
13. *Bhaskaran S., Marappan R., Santhi B.* Design and analysis of a cluster-based intelligent hybrid recommendation system for e-learning applications // Mathematics, 2021. Vol. 9. Art. 197. 21 p. doi: 10.3390/math9020197.
14. *Harshvardhan G. M., Gourisaria M. K., Pandey M., Rautaray S. S.* A comprehensive survey and analysis of generative models in machine learning // Computer Science Review, 2020. Vol. 38. Art. 100285. 29 p. doi: 10.1016/j.cosrev.2020.100285.
15. *Босов А. В., Иванов А. В.* Технология классификации типов контента электронного учебника // Информатика и её применения, 2022. Т. 16. Вып. 4. С. 63–72. doi: 10.14357/19922264220410. EDN: YERCNH.
16. *Босов А. В., Иванов А. В.* Технология многофакторной классификации математического контента электронной системы обучения // Информатика и её применения, 2023. Т. 17. Вып. 4. С. 32–41. doi: 10.14357/19922264230405. EDN: LISHHZ.
17. *Битюков Ю. И., Мартюшова Я. Г.* Решение задач по теории функций комплексного переменного. — М.: МАИ, 2022. 87 с.
18. *Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Bengio Y.* Generative adversarial nets // Adv. Neur. Inf., 2014. Vol. 27. No. 3. P. 2672–2680. doi: 10.1007/978-3-658-40442-0_9.
19. *Chollet F.* Deep learning with Python. — 2nd ed. — Shelter Island, NY, USA: Manning, 2021. 504 p.
20. *Maas A. L., Hannun A. Y., Ng A. Y.* Rectifier nonlinearities improve neural network acoustic models // Proc. ICML, 2013. Vol. 30. No. 1. Art. 3. 6 p.
21. *Arjovsky M., Bottou L.* Towards principled methods for training generative adversarial networks. — Cornell University, 2017. 17 p. arXiv:1701.04862 [stat.ML].
22. *Saatci Y., Wilson A. G.* Bayesian GAN. — Cornell University, 2017. 16 p. arXiv:1705.09558 [stat.ML].
23. *Glorot X., Bengio Y.* Understanding the difficulty of training deep feedforward neural networks // J. Mach. Learn. Res., 2010. Vol. 9. P. 249–256.
24. *He K., Zhang X., Ren S., Sun J.* Delving deep into rectifiers: Surpassing human-level performance on ImageNet Classification // Conference (International) on Computer Vision Proceedings. — Piscataway, NJ, USA: IEEE, 2015. P. 1026–1034. doi: 10.1109/ICCV.2015.123.
25. *Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R.* Dropout: A simple way to prevent neural networks from overfitting // J. Mach. Learn. Res., 2014. Vol. 15. No. 56. P. 1929–1958.
26. *Kullback S., Leibler R. A.* On information and sufficiency // Ann. Math. Stat., 1951. Vol. 22. No. 1. P. 79–86. doi: 10.1214/aoms/1177729694.

Поступила в редакцию 06.02.24

ON THE APPLICATION OF GENERATIVE MODELS IN THE E-LEARNING SYSTEM OF MATHEMATICAL DISCIPLINES

A. V. Bosov and A. V. Ivanov

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

Abstract: The existing tools for individual learning trajectory dynamic design are complemented by the generating technology of certification tasks and exam tickets. A set of exam tickets specially prepared by experts in the university course of the theory of functions of a complex variable was used as a source of high-quality, balanced sets of tasks. This significant training array of high-quality attestation tasks has significantly expanded the available data created at previous stages. The purpose of the performed research was to create methods that allow taking into account the experts' knowledge embedded in the available set of tasks. The implemented generation model when processing educational content uses as parameters the attributes assigned by experts to tasks: topic, complexity, and formed competencies. Two generation methods are proposed. The first one, probabilistic, uses only the frequency characteristics of the training set, approximating the probability distribution. The second one is based on generative-adversarial neural networks. Particular attention is paid to the discussion of the difficulties of the network implementation, including those related to the specific nature of the generative model.

Keywords: e-learning system; educational content; machine learning; generative models; computer simulation; generative-adversarial networks

DOI: 10.14357/19922264240210 **EDN:** UWKQLN

Acknowledgments

The research was supported by the Russian Science Foundation, project No. 22-28-00588, https://rscf.ru/project/22-28-00588/. 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. Rasch, G. 1980. *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press. 224 p.
2. Van der Linden, W. J., D. J. Scrams, and D. L. Schnipke. 1999. Using response-time constraints to control for differential speededness in computerized adaptive testing. *Appl. Psych. Meas.* 23(3):195–210. doi: 10.1177/01466219922031329.
3. Chen, C.-M., H.-M. Lee, and Y.-H. Chen. 2005. Personalized e-learning system using Item Response Theory. *Computers Education* 44(3):237–255. doi: 10.1016/j.compedu.2004.01.006.
4. Kibzun, A. I., and A. O. Inozemtsev. 2014. Using the maximum likelihood method to estimate test complexity levels. *Automat. Rem. Contr.* 75(4):607–621. doi: 10.1134/S000511791404002X. EDN: SKRJCX.
5. Kuravsky, L. S., A. A. Margolis, P. A. Marmalyuk, A. S. Panfilova, G. A. Yuryev, and P. N. Dumin. 2016. A probabilistic model of adaptive training. *Applied Mathematical Sciences* 10(48):2369–2380. doi: 10.12988/ams.2016.65168.
6. Naumov, A. V., and G. A. Mkhitaryan. 2016. On the problem of probabilistic optimization of time-limited testing. *Automat. Rem. Contr.* 77(9):1612–1621. doi: 10.1134/S0005117916090083. EDN: XFMWHF.
7. Bosov, A. V., G. A. Mkhitaryan, A. V. Naumov, and A. P. Sapunova. 2019. Ispol'zovanie modeli gamma-raspredeleniya v zadache formirovaniya ogranichennogo po vremeni testa v sisteme distantsionnogo obucheniya [Using the model of gamma distribution in the problem of forming a time-limited test in a distance learning system]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):12–18. doi: 10.14357/19922264190402. EDN: XUBLZX.
8. Bosov, A. V., Ya. G. Martyushova, A. V. Naumov, and A. P. Sapunova. 2020. Bayesovskiy podkhod k postroeniyu individual'noy traektorii pol'zovatelya v sisteme distantsionnogo obucheniya [Bayesian approach to the construction of an individual user trajectory in the system of distance learning]. *Informatika i ee Primeneniya — Inform. Appl.* 14(3):89–96. doi: 10.14357/19922264200313. EDN: WAKFJR.
9. Adomavicius, G., and A. Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE T. Knowl. Data En.* 17(6):734–749. doi: 10.1109/TKDE.2005.99.
10. Verbert, K., N. Manouselis, X. Ochoa, M. Wolpers, H. Drachsler, I. Bosnic, and E. Duval. 2012. Context-aware recommender systems for learning: A survey and future challenges. *IEEE T. Learn. Technol.* 5(4):318–335. doi: 10.1109/TLT.2012.11.

11. Bosov, A. V. 2022. Primenenie samoorganizuyushchikhsya neyronnykh setey k protsessu formirovaniya individual'noy traektorii obucheniya [Application of self-organizing neural networks to the process of forming an individual learning path]. *Informatika i ee Primeneniya — Inform. Appl.* 16(3):7–15. doi: 10.14357/19922264220302. EDN: HJQANN.
12. Tai, D. W. S., H. J. Wu, and P. H. Li. 2008. Effective e-learning recommendation system based on self-organizing maps and association mining. *Electron. Libr.* 26(3):329–344.
13. Bhaskaran, S., R. Marappan, and B. Santhi. 2021. Design and analysis of a cluster-based intelligent hybrid recommendation system for e-learning applications. *Mathematics* 9:197. 21 p. doi: 10.3390/math9020197.
14. Harshvardhan, G. M., M. K. Gourisaria, M. Pandey, and S. S. Rautaray. 2020. A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review* 38:100285. 29 p. doi: 10.1016/j.cosrev.2020.100285.
15. Bosov, A. V., and A. V. Ivanov. 2022. Tekhnologiya klassifikatsii tipov kontenta elektronnogo uchebnika [Technology for classification of content types of e-textbooks]. *Informatika i ee Primeneniya — Inform. Appl.* 16(4):63–72. doi: 10.14357/19922264220410. EDN: YERCNH.
16. Bosov, A. V., and A. V. Ivanov. 2023. Tekhnologiya mnogofaktornoy klassifikatsii matematicheskogo kontenta elektronnoy sistemy obucheniya [Multifactor classification technology of mathematical content of e-learning system]. *Informatika i ee Primeneniya — Inform. Appl.* 17(4):32–41. doi: 10.14357/19922264230405. EDN: LISHHZ.
17. Bityukov, Yu. I., and Ya. G. Martyushova. 2022. *Reshenie zadach po teorii funktsiy kompleksnogo peremennogo* [Solving problems on the theory of functions of a complex variable]. Moscow: MAI. 87 p.
18. Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, and Y. Bengio. 2014. Generative adversarial nets. *Adv. Neur. Inf.* 27(3):2672–2680. doi: 10.1007/978-3-658-40442-0_9.
19. Chollet, F. 2021. *Deep learning with Python*. 2nd ed. Shelter Island, NY: Manning. 504 p.
20. Maas, A. L., A. Y. Hannun, and A. Y. Ng. 2013. Rectifier nonlinearities improve neural network acoustic models. *Proc. ICML* 30(1):3. 6 p.
21. Arjovsky, M., and L. Bottou. 2017. Towards principled methods for training generative adversarial networks. Cornell University. 17 p. Available at: https://arxiv.org/abs/1701.04862 (accessed May 6, 2024).
22. Saatci, Y., and A. G. Wilson. 2017. Bayesian GAN. Cornell University. 16 p. Available at: https://arxiv.org/abs/1705.09558 (accessed May 6, 2024).
23. Glorot, X., and Y. Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. *J. Mach. Learn. Res.* 9:249–256.
24. He, K., X. Zhang, S. Ren, and J. Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. *Conference (International) on Computer Vision Proceedings*. Piscataway, NJ: IEEE. 1026–1034. doi: 10.1109/ICCV.2015.123.
25. Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15(56):1929–1958.
26. Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. *Ann. Math. Stat.* 22(1):79–86. doi: 10.1214/aoms/1177729694.

Received February 6, 2024

Contributors

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

Ivanov Alexey V. (b. 1976) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; aivanov@ipiran.ru

ТРАНСФОРМАЦИИ ОБЪЕКТОВ ПЕРВОГО И ВТОРОГО ПОРЯДКА В ЛЕКСИКОГРАФИЧЕСКОЙ ИНФОРМАЦИОННОЙ СИСТЕМЕ*

И. М. Зацман[1]

Аннотация: Рассматриваются теоретические основания проектирования информационных технологий (ИТ) интеграции двуязычных словарей и параллельных корпусов. Дано описание первых результатов создания третьего уровня классификации трансформаций объектов предметной области информатики, которую предполагается использовать при создании концепции лексикографической информационной системы, обеспечивающей интеграцию. Все сущности информатики в статье разделены на два глобальных класса: объекты и их трансформации. Для каждого такого класса конструируется своя классификация. Ранее были описаны два верхних уровня классификации трансформаций объектов предметной области. В данной статье рассматривается третий уровень этой классификации. Основанием для построения самого верхнего ее уровня служило деление предметной области информатики на среды (ментальная, сенсорно воспринимаемая, цифровая и ряд других сред), каждая из которых по определению включает объекты одной природы. Основанием для построения второго уровня классификации трансформаций объектов служила типология знаковых систем А. Соломоника. Цель статьи состоит в систематизации трансформаций первого и второго порядка объектов предметной области на третьем уровне этой классификации. Основанием для систематизации служит средовая версия иерархии Акоффа.

Ключевые слова: объекты предметной области; трансформации объектов; классификация; данные; информация; знание; лексикографическая информационная система

DOI: 10.14357/19922264240211 **EDN:** VZTGVV

1 Введение

Возникновение параллельных корпусов, в которых предложениям оригинального текста сопоставлены предложения его перевода, обеспечило возможность контрастивного лингвистического анализа на принципиально новом уровне полноты и точности, недостижимом в докорпусную эпоху. Пионерскими в этой области стали работы 1990-х гг. Стига Йоханссона с англо-норвежским корпусом [1]. В России параллельные корпусы стали формироваться в начале XXI века в рамках Национального корпуса русского языка [2].

Создатели двуязычных словарей используют параллельные корпусы для сбора материала и эмпирической проверки своих гипотез, касающихся межъязыковой эквивалентности. Ценность параллельных корпусов определяется тем, что в лингвистике этап сбора исходного материала считается наиболее трудоемким и наименее творческим, а параллельные корпусы позволяют значительно сэкономить время и силы для творческого этапа создания словарей [3]. При этом двуязычные словари, создаваемые на основе исходного материала, извлеченного из параллельных корпусов, сейчас формируются без связей с их текстами. Другими словами, онлайновые связи созданных словарей с параллельными корпусами, которые служили источниками исходного материала, отсутствуют.

Параллельные корпусы постоянно пополняются новыми текстами, в предложениях которых можно обнаружить новые значения слов и устойчивых словосочетаний. Однако при этом отсутствуют методы и средства оперативного обновления словарей по корпусным данным. В настоящее время проблема установления связей между двуязычными словарями и параллельными корпусами (далее — проблема интеграции) находится на стадии поиска концептуальных подходов к их интеграции на уровне значений.

Подход к решению проблемы интеграции, предлагаемый в статье, учитывает и появление новых значений слов и устойчивых словосочетаний, и динамику смысловых значений, которая обусловлена развитием и пополнением знания лингвистов, фиксирующих эти значения в результате семантического анализа пополняемых корпусных данных. Проведенные эксперименты показали, что обнаружение нового лингвистического знания обусловливает и формирование дефиниций новых значений, и пересмотр уже существующих дефиниций [4, 5].

*Исследование выполнено в ФИЦ ИУ РАН за счет гранта Российского научного фонда № 24-18-00155, https://rscf.ru/project/24-18-00155. Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, izatsman@yandex.ru

Например, в проведенных экспериментах с использованием ЦКП «Информатика» ФИЦ ИУ РАН фиксировалась эволюция значений немецких модальных глаголов, исходное состояние значений которых было описано в немецко-русском словаре. В экспериментальном массиве текстов как потенциальных источниках нового знания 16 268 предложений содержали немецкие модальные глаголы и в 2041 из них встречался глагол sollen. В начале эксперимента в словаре были описаны 12 значений этого модального глагола. По окончании эксперимента лингвисты обнаружили два новых его значения, согласовали их дефиниции и описали эволюцию дефиниций [6, 7].

Таким образом, для решения проблемы интеграции требуется фиксировать новое знание, обнаруженное лингвистами в текстовых данных параллельных корпусов, отслеживать эволюцию знания, представленного в виде дефиниций значений слов и устойчивых словосочетаний, и, соответственно, актуализировать электронные двуязычные словари. Предлагаемый концептуальный подход к интеграции, который планируется реализовать в процессе проектирования лексикографической информационной системы, фиксирующей эволюцию лингвистического знания, основан на решении следующих задач:

– категоризация трех базовых понятий информатики, включенных в иерархию Акоффа [8] (данные, информация, знание), на объекты проектируемой системы, которая необходима, чтобы фиксировать «кванты» нового знания и отслеживать его эволюцию в этой системе;

– систематизация трансформаций объектов этой системы.

Цель статьи и состоит в решении двух задач: категоризации трех базовых понятий информатики на объекты лексикографической информационной системы и систематизации трансформаций первого и второго порядка ее объектов.

Трансформациями первого порядка, о которых сказано в формулировке цели статьи, называются взаимные преобразования между двумя объектами системы одной природы. Например, перевод в системе текста с русского языка на английский относится к ним. Трансформациями второго порядка и выше называются взаимные преобразования между двумя и более объектами разной природы. Например, кодирование символов текста компьютерными кодами и их декодирование относятся по определению к трансформациям второго порядка.

2 Процессы трансформаций в информатике

Процессы трансформаций, рассматриваемые в статье, относятся к теоретическому ядру информатики, а не только к проектированию лексикографической информационной системы. Например, из трех основных подходов к описанию предметной области информатики[1] (объектный, трансформационный и синтетический) систематизация трансформаций ближе всего ко второму подходу. Примерами первого подхода, в рамках которого основное внимание уделяется объектам предметной области информатики и в меньшей степени отношениям между ними, могут служить работы [8, 10, 11]; примерами второго подхода, в рамках которого основное внимание уделяется трансформациям и в меньшей степени трансформируемым объектам, — работы [12, 13]; примерами третьего, синтетического подхода, в котором уделяется внимание и объектам предметной области информатики, и отношениям между ними, могут служить работы [14–18].

Таким образом, для описания трансформаций объектов лексикографической информационной системы предпочтительнее всего трансформационный подход, который упоминается и в определениях информатики. Например, в 2009 г. П. Деннинг и П. Розенблум сформулировали суть информатики как компьютинга следующим образом: «. . .информатика — это не просто алгоритмы и структуры данных; это преобразования [трансформации] представлений» [12]. Чуть позже, в контексте краткого описания парадигмы информатики как компьютинга, П. Деннинг и П. Фриман изменили эту формулировку на такую: «Центральный объект внимания в информатике можно определить как информационные процессы — *естественные или искусственные процессы, преобразующие информацию* (курсив мой — И. З.)» [13]. Согласно парадигме, предлагаемой авторами этой статьи, на начальном этапе проектирования автоматизированных систем базовыми элементами моделей их функционирования служат *информационные процессы*.

Однако если 15 лет назад в формулировке из работы [13] шла речь о процессах, преобразующих информацию, то в последние 10 лет в спектр процессов трансформаций все чаще стали включать процессы, преобразующие не только информацию, но также и другие объекты автоматизированных систем, в первую очередь данные и знания [19–21]. Например, Виктория Стодден, позиционируя

[1] В статье предметная область информатики трактуется согласно концепции полиадического компьютинга Пола Розенблума [9].

науку о данных как одну из дисциплин информатики, говорит, что центральный объект исследований в науке о данных — это «изучение обобщаемого извлечения знания из данных» [21]. Увеличение и числа объектов, и спектра процессов их трансформаций в автоматизированных системах обусловливает необходимость систематизации и объектов, и процессов их трансформаций на начальном этапе проектирования систем.

Для создания концепции лексикографической информационной системы и проектирования ИТ, обеспечивающих интеграцию двуязычных словарей и параллельных корпусов, сначала выполним категоризацию на объекты этой системы трех базовых понятий информатики (данные, информация, знание) в контексте построения классификаций сущностей ее предметной области.

Необходимость использования классификаций информатики в процессе создания концепции проиллюстрируем, используя иерархию Акоффа [8]. Он использовал принцип их вертикального размещения в иерархии снизу вверх: данные, информация и знание. Еще в ней есть термин «мудрость», который в статье не рассматривается. Такое размещение Акофф прокомментировал так: «Каждое из перечисленных понятий [кроме данных] содержит в себе нижестоящие. . .» [8].

Этому принципу размещения и комментарию Акоффа свойственны недостатки, проанализированные, в частности, в работе [10]. Главный вывод, к которому пришла Роули после изучения иерархии Акоффа, заключается в следующем: «. . .информация определяется в терминах данных, знание — в терминах информации. . . но существует меньше консенсуса в описании трансформаций, которые преобразуют сущности, расположенные ниже в иерархии, в те, которые находятся над ними, что приводит к их терминологической неопределенности» [10]. Причина этой неопределенности, скорее всего, в том, что базовые понятия информатики включены в иерархию Акоффа изолированно от общего контекста классификаций сущностей ее предметной области.

3 Классификации сущностей информатики

Все сущности предметной области информатики в работах [22, 23] разделены на два глобальных класса: ее объекты и их трансформации. Для каждого такого класса была предложена своя классификация. В работе [23] дано описание классификации объектов предметной области информатики, первый уровень которой содержит базовые понятия ее предметной области (данные, информация, знания и др.). В работе [22] дано описание двух верхних уровней классификации трансформаций объектов предметной области (см. рисунок в работе [22]). Основанием для построения самого верхнего ее уровня послужило деление предметной области информатики на среды[1] и степень разнообразия природы объектов, вовлеченных в трансформации:

- первый класс верхнего уровня классификации включает трансформации объектов в пределах среды только одной природы (трансформации первого порядка);
- второй класс включает трансформации объектов, относящихся к двум средам разной природы (трансформации второго порядка);
- третий и последующие классы включают трансформации объектов, относящихся к трем и более средам разной природы (трансформации третьего и более высоких порядков).

В работе [22] были приведены примеры для трех первых классов трансформаций, включая пример трансформаций объектов, относящихся к двум средам разной природы (компьютерное кодирование символов текстов с помощью таблиц Unicode).

Основанием для построения второго уровня классификации трансформаций объектов послужила типология знаковых систем А. Соломоника [25, с. 131]: естественные знаковые системы, образные, естественно-языковы́е, вербально-несловесные системы записи[2] и формализованные знаковые системы, включая математические. Введем понятие обобщенного текста — это текст, который может быть создан в любой из перечисленных знаковых систем. Тогда обобщенные тексты могут быть естественными, образными, естественно-языковы́ми, вербально-несловесными и формализованными. Второй уровень классификации трансформаций охватывает не все виды объектов предметной области информатики, а только перечисленные 5 видов текстов и их представления, вовлеченные в процессы трансформаций в одной или более средах вместе с данными, знанием и его концептами.

[1] В работе [24] дано описание пяти сред предметной области информатики (ментальная; сенсорно воспринимаемая, или информационная; цифровая; нейро- и ДНК-среда), каждая из которых по определению включает объекты одной и той же природы.

[2] Под системой записи понимается знаковая система, сочетающая вербальные знаки с несловесными (языки нотной записи, карт, таблиц и др.).

4 Классификация трансформаций: построение третьего уровня

Основанием для систематизации трансформаций первого и второго порядка на третьем уровне этой классификации служит иерархия Акоффа [8], на основе которой и была создана ее средова́я версия [26, 27]. Для создания средово́й версии была выполнена категоризация трех базовых понятий информатики (данные, информация, знания) на объекты лексикографической информационной системы в процессе создания ее концепции (рис. 1).

В отличие от классической иерархии Акоффа, в ее средово́й версии различаются три вида данных: сенсорно воспринимаемые, цифровые и те данные, которые генерируются искусственными нейронными сетями (ИНС) в системах искусственного интеллекта (далее — ИИ-данные). Последний вид данных необходим, например, для различения входа и выхода процесса применения обученной ИНС в цифровой модели генерации знания, описанию которой посвящена работа [27].

Также предлагается различать два вида информации: сенсорно воспринимаемая и цифровая. Кроме знания в средову́ю версию добавлены концепты и ментальные образы сенсорно воспринимаемых данных. Последние служат промежуточной сущностью между сенсорно воспринимаемыми данными и генерируемым знанием при описании процессов извлечения знания из текстовых данных лексикографической информационной системы. Описание объектов средово́й версии иерархии Акоффа (см. рис. 1) и отношений между ними дано в работах [26, 28].

В средово́й версии число объектов равно восьми. Если учитывать направления трансформаций, то между восемью объектами на рис. 1 она включает 16 их видов (трансформации на границе между сенсорно воспринимаемыми данными и информацией, обозначенные символом «?», в статье не рассматриваются). В будущем число объектов в средово́й версии, которая выбрана как основание для систематизации трансформаций первого и второго порядка, может быть увеличено. Для построения классификации трансформаций важно не возможное увеличение числа объектов и трансформаций между ними, а то, что их виды в средово́й версии распределены между трансформациями первого и второго порядка. Из 16 видов на рис. 1 шесть относятся к трансформациям первого порядка, это виды с номерами 7, 8, 13–16 (далее — типология трансформаций первого порядка), а десять — к трансформациям второго порядка, это виды с номерами 1–6 и 9–12 (далее — типология трансформаций второго порядка). Разместим обе типологии на третьем уровне классификации (см. ее схему на рис. 2). Перечислим виды трансформаций первой типологии, вводя в скобках их краткие названия, используемые ниже на рис. 3:

7 — членение знания на концепты с помощью одной или нескольких знаковых систем (далее — членение знания);

8 — формирование знания на основе концептов (формирование знания);

13 — обучение ИНС;

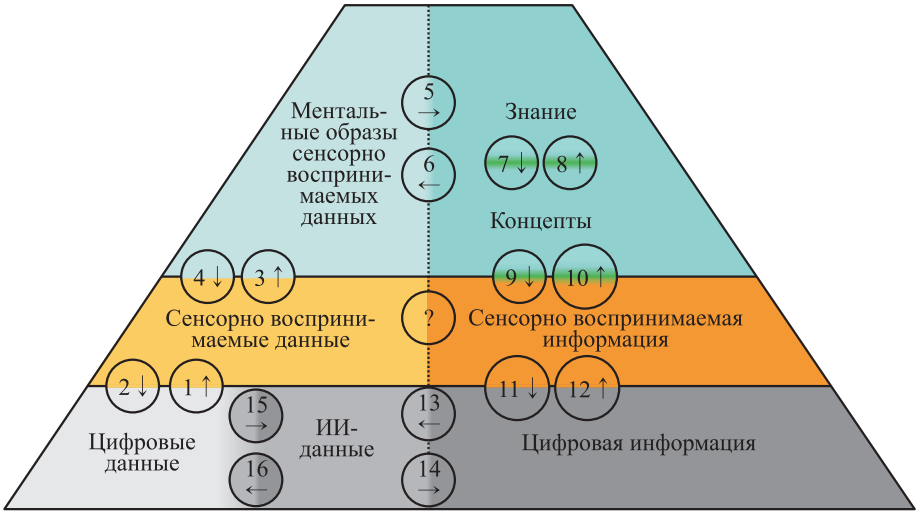


Рис. 1 Средовая версия иерархии Акоффа

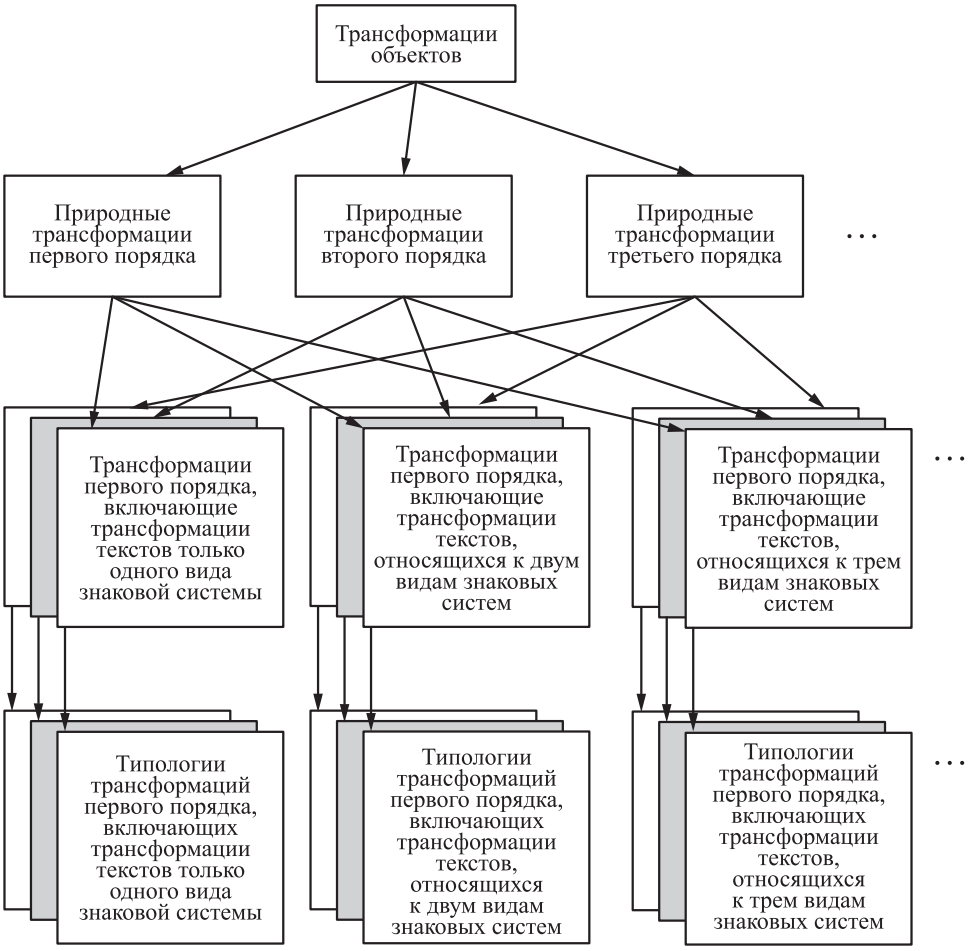


Рис. 2 Схема трех верхних уровней классификации трансформаций объектов (объединены по три слоя и для второго, и для третьего уровней этой классификации)

14 — восстановление обучающей информации на основе содержания обученной ИНС (обращение ИНС);

15 — использование обученной ИНС (использование ИНС);

16 — восстановление исходных данных, соответствующих полученным результатам работы обученной ИНС (восстановление исходных данных по результатам ИНС).

Не все виды трансформаций 13–16 поддерживаются в конкретных системах искусственного интеллекта, но с теоретической точки зрения все их предлагается включить в первую типологию для полноты спектра видов трансформаций.

Перечислим виды трансформаций второй типологии:

1 — декодирование цифровых данных в компьютерных системах (декодирование данных);

2 — кодирование сенсорно воспринимаемых данных (кодирование данных);

3 — ментальное копирование сенсорно воспринимаемых данных (ментальное копирование);

4 — восстановление сенсорно воспринимаемых данных по ментальным образам (восстановление по образам);

5 — смысловая интерпретация без деления на концепты ментальных образов сенсорно воспринимаемых данных (смысловая интерпретация);

6 — восстановление ментальных образов (восстановление образов);

9 — представление концептов в виде сенсорно воспринимаемой информации, например текстами, формулами, таблицами, рисунками и т. д. (представление концептов);

10 — понимание смысла сенсорно воспринимаемой информации (понимание смысла);

11 — кодирование сенсорно воспринимаемой информации (кодирование информации);

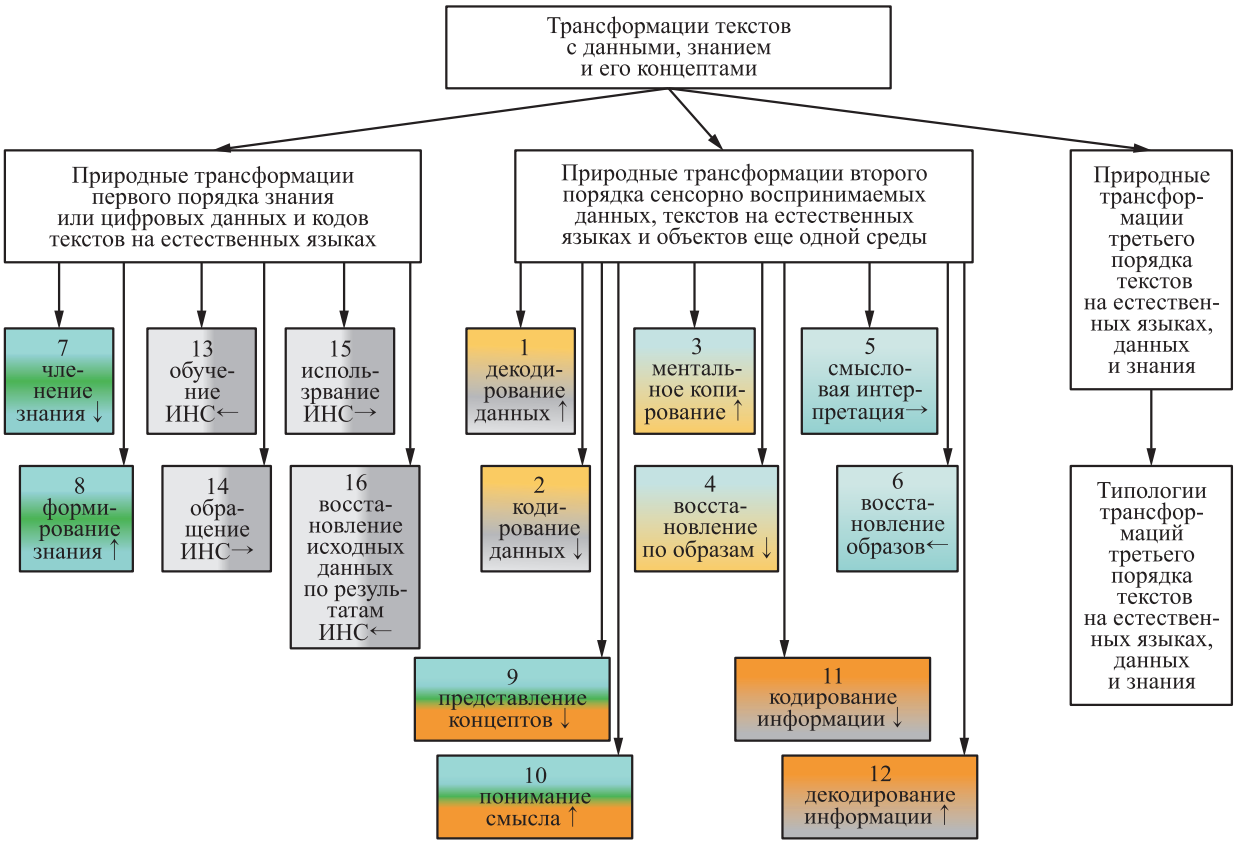


Рис. 3 Схема частного случая классификации трансформаций объектов (трансформации пронумерованы согласно рис. 1)

12 — декодирование цифровой информации (декодирование информации).

Отметим, что в существующих ИТ и компьютерных системах наиболее часто используются виды трансформаций 13 и 15 типологии первого порядка и 1, 2, 11 и 12 типологии второго порядка. На рис. 2 в первом слое третьего уровня классификации показаны типологии первого порядка без указания числа трансформаций в них и без детализации трансформируемых объектов.

Во втором слое третьего уровня классификации условно (без названий) показаны типологии второго порядка. Также на рис. 2 в третьем слое третьего уровня классификации условно (также без названий) показаны типологии третьего порядка, которые планируется рассмотреть в отдельной статье. По определению они должны включать трансформации между тремя объектами разной природы, но средова́я версия иерархии Акоффа включает трансформации только между двумя объектами разной природы. Поэтому потребуется другое основание для их систематизации (ранее были рассмотрены отдельные примеры трансформаций третьего порядка[1] [29]).

5 Классификация трансформаций: частный случай

Выше было отмечено, что в будущем число объектов в средово́й версии иерархии Акоффа может быть увеличено. Это означает, что увеличатся и число объектов, и число трансформаций между ними в классификации трансформаций, так как эта средова́я версия служит по определению основанием для систематизации трансформаций первого и второго порядка. Поэтому на третьем уровне рис. 2 указаны типологии без детализации объектов и без указания числа трансформаций в каждой из

[1]Далеко не всегда трансформации третьего и более высоких порядков можно рассматривать как последовательность трансформаций второго порядка. Примером этого могут служить трансформации в процессе обучения пациента пользованию роботизированной рукой, охватывающие личностные концепты пациента, релевантные его намерениям, сигналы активности мозга как объекты нейросреды и компьютерные коды [29].

них. С одной стороны, при таком подходе получаем достаточно общий вид этой классификации, так как она не зависит от числа объектов в том или ином варианте средово́й версии (и это существенно упрощает рис. 2). С другой стороны, на третьем уровне такой общей классификации подразумевается, но не эксплицируется природа трансформируемых объектов и их возможные сочетания в трансформациях.

При проектировании лексикографической информационной системы важно эксплицировать природу трансформируемых объектов и их возможные сочетания. Поэтому в парадигму информатики [30] кроме общей классификации трансформаций предлагается включать и ее частные случаи, эксплицирующие природу трансформируемых объектов.

В этом разделе рассмотрим один частный случай, когда используются только естественные знаковые системы из типологии А. Соломоника [25] вместе с данными, знанием и его концептами. Число естественных языков при этом не ограничено. И этот частный случай классификации включает только три класса природных трансформаций (первого, второго и третьего порядка, см. схему классификации на рис. 3).

Первый и второй уровни схемы общей классификации (см. рис. 2) можно объединить в один уровень в этом частном случае. Ниже этого уровня приведено содержание типологий первого и второго порядка без содержания типологий третьего порядка.

Наполнение типологий первого и второго порядка соответствует средово́й версии иерархии Акоффа на рис. 1, содержащей 6 видов трансформаций типологии первого порядка и 10 видов трансформаций типологии второго порядка (на рис. 3 стрелки указывают направления трансформаций согласно средово́й версии на рис. 1).

Таким образом, частный случай классификации содержит для этих двух типологий 16 теоретически возможных трансформаций, 6 из которых в настоящее время в существующих ИТ применяются наиболее часто: виды трансформаций 1, 2, 11 и 12 типологии второго порядка реализуются с помощью тех или иных методов кодирования/декодирования (например, с использованием таблиц Unicode), а виды трансформаций 13 и 15 в типологии первого порядка реализуются полностью с помощью процессов цифровой обработки компьютерами.

Остальные виды трансформаций или применяются намного реже (это виды 3, 5, 7, 9 и 10), или находятся в стадии поиска и разработки (14 и 16) или в настоящее время носят только теоретический характер, обеспечивая полноту первой и второй типологий (4, 6 и 8). Знаком «?» обозначены те виды трансформаций, которые по определению не существуют в используемой парадигме информатики [30]. Однако возможно, что в других будущих подходах к построению ее парадигмы эти виды трансформаций будут существовать.

6 Заключение

На сегодняшний день процесс построения классификаций объектов предметной области информатики [23] и их трансформаций [22] еще не завершен. Однако первые результаты их построения уже используются для создания концепции лексикографической информационной системы, обеспечивающей интеграцию двуязычных словарей и параллельных корпусов.

Автор признателен рецензентам за помощь в улучшении статьи.

Литература

1. *Aijmer K., Altenberg B.* Advances in corpus-based contrastive linguistics. Studies in honour of Stig Johansson. — Amsterdam: John Benjamins, 2013. 295 p. doi: 10.1075/scl.54.
2. *Добровольский Д. О., Кретов А. А., Шаров С. А.* Корпус параллельных текстов // Научная и техническая информация. Сер. 2: Информационные процессы и системы, 2005. № 6. С. 16–27.
3. *Добровольский Д. О.* Корпус параллельных текстов и сопоставительная лексикология // Труды Института русского языка им. В. В. Виноградова, 2015. № 6. С. 413–449. EDN: VJQBHP.
4. *Гончаров А. А., Зацман И. М., Кружков М. Г.* Эволюция классификаций в надкорпусных базах данных // Информатика и её применения, 2020. Т. 14. Вып. 4. С. 108–116. doi: 10.14357/19922264200415. EDN: GKWBZT.
5. *Гончаров А. А., Зацман И. М., Кружков М. Г.* Представление новых лексикографических знаний в динамических классификационных системах // Информатика и её применения, 2021. Т. 15. Вып. 1. С. 86–93. doi: 10.14357/19922264210112. EDN: OPEFXW.
6. *Zatsman I.* Finding and filling lacunas in linguistic typologies // 15th Forum (International) on Knowledge Asset Dynamics Proceedings. — Matera, Italy: Institute of Knowledge Asset Management, 2020. P. 780–793.
7. *Zatsman I.* Three-dimensional encoding of emerging meanings in AI-systems // 21st European Conference on Knowledge Management Proceedings. — Reading, U.K.: Academic Publishing International Ltd., 2020. P. 878–887.

8. *Ackoff R.* From data to wisdom // J. Applied Systems Analysis, 1989. Vol. 16. No. 1. P. 3–9.
9. *Rosenbloom P. S.* On computing: The fourth great scientific domain. — Cambridge, MA, USA: MIT Press, 2013. 307 p.
10. *Rowley J.* The wisdom hierarchy: Representations of the DIKW hierarchy // J. Inf. Sci., 2007. Vol. 33. Iss. 2. P. 163–180. doi: 10.1177/0165551506070706.
11. *Frické M. H.* Data–Information–Knowledge–Wisdom (DIKW) pyramid, framework, continuum // Encyclopedia of big data / Eds. L. Schintler, C. McNeely. — Cham: Springer, 2018. 4 p. doi: 10.1007/978-3-319-32001-4_331-1.
12. *Denning P., Rosenbloom P.* Computing: The fourth great domain of science // Commun. ACM, 2009. Vol. 52. Iss. 9. P. 27–29.
13. *Denning P., Freeman P.* Computing's paradigm // Commun. ACM, 2009. Vol. 52. Iss. 12. P. 28–30. doi: 10.1145/1610252.1610265.
14. *Farradane J.* Knowledge, information, and information science // J. Inf. Sci., 1980. Vol. 2. Iss. 2. P. 75–80. doi: 10.1177/016555158000020020.
15. *Шрейдер Ю. А.* Информация и знание // Системная концепция информационных процессов. — М.: ВНИИСИ, 1988. С. 47–52.
16. *Ingwersen P.* Information and information science // Enclyclopaedie of library and information science / Eds. J. D. McDonald, M. Levine-Clark. — New York, NY, USA: Marcel Dekker Inc., 1992. Vol. 56. Sup. 19. P. 137–174.
17. Информатика как наука об информации: Информационный, документальный, технологический, экономический, социальный и организационный аспекты / Под ред. Р. С. Гиляревского. — М.: Фаир-Пресс, 2006. 592 с.
18. *Hjorland B.* Library and information science: practice, theory, and philosophical basis // Inform. Process. Manag., 2000. Vol. 36. Iss. 3. P. 501–531. doi: 10.1016/S0306-4573(99)00038-2.
19. Deep shift — technology tipping points and societal impact. — Geneva: WE Forum, 2015. 44 p. http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf.
20. *Berman F., Rutenbar R., Hailpern B., Christensen H., Davidson S., Estrin D., Franklin M., Martonosi M., Raghavan P., Stodden V., Szalay A. S.* Realizing the potential of data science // Commun. ACM, 2018. Vol. 61. Iss. 4. P. 67–72. doi: 10.1145/3188721.
21. *Stodden V.* The data science life cycle: A disciplined approach to advancing data science as a science // Commun. ACM, 2020. Vol. 63. Iss. 7. P. 58–66. doi: 10.1145/3360646.
22. *Зацман И. М.* Научная парадигма информатики: классификация трансформаций объектов предметной области // Системы и средства информатики, 2023. Т. 33. № 4. С. 126–138. doi: 10.14357/08696527230412. EDN: ZIKUWO.
23. *Зацман И. М.* Научная парадигма информатики: классификация объектов предметной области // Информатика и её применения, 2023. Т. 17. Вып. 4. С. 96–103. doi: 10.14357/19922264230413. EDN: FIUQAT.
24. *Зацман И. М.* О научной парадигме информатики: верхний уровень классификации объектов ее предметной области // Информатика и её применения, 2022. Т. 16. Вып. 4. С. 73–79. doi: 10.14357/19922264220411. EDN: XZNKVI.
25. *Соломоник А. Б.* Философия знаковых систем и язык. — М.: ЛКИ, 2011. 408 с.
26. *Зацман И. М.* Трансформация иерархии Акоффа в научной парадигме информатики // Информатика и её применения, 2023. Т. 17. Вып. 3. С. 107–113. doi: 10.14357/19922264230315. EDN: UMVRRV.
27. *Zatsman I.* Building digital spiral models of knowledge generation // 19th Forum (International) on Knowledge Asset Dynamics Proceedings. — Matera, Italy: Arts for Business Institute, 2024. P. 2185–2196.
28. *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.
29. *Зацман И. М.* Интерфейсы третьего порядка в информатике // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 82–89. doi: 10.14357/19922264190312. EDN: EHRQLF.
30. *Зацман И. М.* Научная парадигма информатики как третьей культуры // Научно-техническая информация. Сер. 1: Организация и методика информационной работы, 2023. № 11. С. 1–14.

Поступила в редакцию 14.04.24

OBJECT TRANSFORMATIONS OF THE FIRST AND SECOND ORDER IN A LEXICOGRAPHIC INFORMATION SYSTEM

I. M. Zatsman

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

Abstract: The theoretical foundations of the design of information technologies used for the integration of bilingual dictionaries and parallel corpora are considered. The description of the first outcomes of the creation of the third

level of object transformations classification in the subject domain of informatics, which is supposed to be used in creating the lexicographic information system providing integration, is given. All the entities of informatics are divided into two global classes: objects and their transformations. For each such class, its own classification is constructed. Previously, the two upper levels of the object transformation classification in the subject domain have been described. The present paper discusses the third level of this classification. The basis for the construction of its highest level was the division of the subject domain of informatics into media (mental, sensory, digital, and a number of other media), each of which by definition includes objects of the same nature. The Solomonick's typology of sign systems served as the basis for constructing the second level of the object transformation classification. The aim of the paper is to systematize object transformations of the first and second orders at the third level of this classification. The basis for systematization is the medium version of the Ackoff's hierarchy.

Keywords: subject domain objects; object transformations; classification; data; information; knowledge; lexicographic information system

DOI: 10.14357/19922264240211 **EDN:** VZTGVV

Acknowledgments

The reported study was funded by the Russian Science Foundation, project No. 24-18-00155, https://rscf.ru/project/24-18-00155. 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. Aijmer, K., and B. Altenberg. 2013. *Advances in corpus-based contrastive linguistics. Studies in honour of Stig Johansson*. Amsterdam: John Benjamins. 295 p. doi: 10.1075/scl.54.
2. Dobrovolskiy, D. O., A. A. Kretov, and S. A. Sharov. 2005. Korpus parallel'nykh tekstov [Corpus of parallel texts]. *Nauchnaya i tekhnicheskaya informatsiya. Ser. 2. Informatsionnye protsessy i sistemy* [Scientific and Technical Information. Ser. 2: Information Processes and Systems] 6:16–27.
3. Dobrovolskiy, D. O. 2015. Korpus parallel'nykh tekstov i sopostavitel'naya leksikologiya [The corpus of parallel texts and contrastive lexicology]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova* [Proceedings of the V. V. Vinogradov Russian Language Institute] 6:413–449. EDN: VJQBHP.
4. Goncharov, A. A., I. M. Zatsman, and M. G. Kruzhkov. 2020. Evolyutsiya klassifikatsiy v nadkorpusnykh bazakh dannykh [Evolution of classifications in supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 14(4):108–116. doi: 10.14357/19922264200415. EDN: GKWBZT.
5. Goncharov, A. A., I. M. Zatsman, and M. G. Kruzhkov. 2021. Predstavlenie novykh leksikograficheskikh znaniy v dinamicheskikh klassifikatsionnykh sistemakh [Representation of new lexicographical knowledge in dynamic classification systems]. *Informatika i ee Primeneniya — Inform. Appl.* 15(1):86–93. doi: 10.14357/19922264210112. EDN: OPEFXW.
6. Zatsman, I. 2020. Finding and filling lacunas in linguistic typologies. *15th Forum (International) on Knowledge Asset Dynamics Proceedings*. Matera, Italy: Institute of Knowledge Asset Management. 780–793.
7. Zatsman, I. 2020. Three-dimensional encoding of emerging meanings in AI-systems. *21st European Conference on Knowledge Management Proceedings*. Reading, U.K.: Academic Publishing International Ltd. 878–887.
8. Ackoff, R. 1989. From data to wisdom. *J. Applied Systems Analysis* 16(1):3–9.
9. Rosenbloom, P. S. 2013. *On computing: The fourth great scientific domain*. Cambridge, MA: MIT Press. 307 p.
10. Rowley, J. 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. *J. Inf. Sci.* 33(2):163–180. doi: 10.1177/0165551506070706.
11. Frické, M. H. 2018. Data-Information-Knowledge-Wisdom (DIKW) pyramid, framework, continuum. *Encyclopedia of big data*. Eds. L. Schintler and C. McNeely. Cham: Springer. 4 p. doi: 10.1007/978-3-319-32001-4_331-1.
12. Denning, P., and P. Rosenbloom. 2009. Computing: The fourth great domain of science. *Commun. ACM* 52(9):27–29.
13. Denning, P., and P. Freeman. 2009. Computing's paradigm. *Commun. ACM* 52(12):28–30. doi: 10.1145/1610252.1610265.
14. Farradane, J. 1980. Knowledge, information, and information science. *J. Inf. Sci.* 2(2):75–80. doi: 10.1177/016555158000020020.
15. Shreyder, Yu. A. 1988. Informatsiya i znanie [Information and knowledge]. *Sistemnaya kontseptsiya informatsionnykh protsessov* [System concept of information processes]. Moscow: VNIISI. 47–52.
16. Ingwersen, P. 1995. Information and information science. *Encyclopedia of library and information science*. Eds. J. D. McDonald and M. Levine-Clark. New York, NY: Marcel Dekker Inc. 56(19):137–174.
17. Gilyarevskiy, R. S., ed. 2006. *Informatika kak nauka ob informatsii: informatsionnyy, dokumental'nyy, tekhnologicheskiy, ekonomicheskiy, sotsial'nyy i organizatsionnyy aspekty* [Informatics as information science: Informational,

documentary, technological, economic, social, and organizational dimensions]. Moscow: FAIR-PRESS. 592 p.

18. Hjorland, B. 2000. Library and information science: Practice, theory, and philosophical basis. *Inform. Process. Manag.* 36(3):501–531. doi: 10.1016/S0306-4573(99)00038-2.
19. Deep shift — technology tipping points and societal impact. 2015. *World Economic Forum*. Geneva. 44 p. Available at: http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf (accessed May 20, 2024).
20. Berman, F., R. Rutenbar, B. Hailpern, H. Christensen, S. Davidson, D. Estrin, M. Franklin, M. Martonosi, P. Raghavan, V. Stodden, and A. S. Szalay. 2018. Realizing the potential of data science. *Commun. ACM* 61(4):67–72. doi: 10.1145/3188721.
21. Stodden, V. 2020. The data science life cycle: A disciplined approach to advancing data science as a science. *Commun. ACM* 63(7):58–66. doi: 10.1145/3360646.
22. Zatsman, I. M. 2023. Nauchnaya paradigma informatiki: klassifikatsiya transformatsiy ob"ektov predmetnoy oblasti [Scientific paradigm of informatics: Transformation classification of domain objects]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 33(4):126–138. doi: 10.14357/08696527230412. EDN: ZIKUWO.
23. Zatsman, I. M. 2023. Nauchnaya paradigma informatiki: klassifikatsiya ob"ektov predmetnoy oblasti [Scientific paradigm of informatics: Classification of domain objects]. *Informatika i ee Primeneniya — Inform. Appl.* 17(4):96–103. doi: 10.14357/19922264230413. EDN: FIUQAT.
24. Zatsman, I. M. 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. EDN: XZNKVI.
25. Solomonick, A. B. 2011. *Filosofiya znakovykh system i yazyk* [Philosophy of sign systems and language]. Moscow: LKI. 408 p.
26. Zatsman, I. M. 2023. Transformatsiya ierarkhii Akoffa v nauchnoy paradigme informatiki [Transformation of the Ackoff's hierarchy in the scientific paradigm of informatics]. *Informatika i ee Primeneniya — Inform. Appl.* 17(3):107–113. doi: 10.14357/19922264230315. EDN: UMVRRV.
27. Zatsman, I. 2024. Building digital spiral models of knowledge generation. *19th Forum (International) on Knowledge Asset Dynamics Proceedings*. Matera, Italy: Arts for Business Institute. 2185–2196.
28. 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.
29. 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. EDN: EHRQLF.
30. Zatsman, I. 2023. Scientific paradigm of informatics as a third culture. *Scientific Technical Information Processing* 50(4):246–258. doi: 10.3103/S0147688223040111. EDN: CKHMYS.

Received April 14, 2024

Contributor

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, head of department, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; izatsman@yandex.ru

ОБ АВТОРАХ

Агаларов Явер Мирзабекович (р. 1952) — кандидат технических наук, доцент, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Бесчастный Виталий Александрович (р. 1992) — кандидат физико-математических наук, доцент кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов им. Патриса Лумумбы

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 Федерального исследовательского центра «Информатика и управление» Российской академии наук

Васильев Николай Семенович (р. 1952) — доктор физико-математических наук, профессор Московского государственного технического университета имени Н. Э. Баумана

Гайдамака Юлия Васильевна (р. 1971) — доктор физико-математических наук, профессор кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов им. Патриса Лумумбы; старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Голос Елизавета Сергеевна (р. 1998) — аспирант кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов им. Патриса Лумумбы

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Николай Александрович (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Забежайло Михаил Иванович (р. 1956) — доктор физико-математических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Федерального исследовательского центра «Информатика и управление» Российской академии наук

Иванов Алексей Владимирович (р. 1976) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Коваленко Алексей Сергеевич (р. 1996) — ассистент кафедры прикладной математики и программирования Института математики, механики и компьютерных наук им. И. И. Воровича Южного федерального университета

Кочеткова Ирина Андреевна (р. 1985) — кандидат физико-математических наук, доцент кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов им. Патриса Лумумбы; старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кульченков Владислав Владимирович (р. 1989) — заместитель начальника управления портфельного анализа Банка ВТБ

Ланге Андрей Михайлович (р. 1979) — кандидат физико-математических наук, научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ланге Михаил Михайлович (р. 1945) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Макеева Елена Дмитриевна (р. 1996) — аспирант кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов им. Патриса Лумумбы; младший научный сотрудник Института проблем управления им. В. А. Трапезникова Российской академии наук

Мачнев Егор Андреевич (р. 1996) — аспирант кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов им. Патриса Лумумбы

Острикова Дарья Юрьевна (р. 1988) — кандидат физико-математических наук, доцент кафедры теории вероятностей и кибербезопасности Российского университета дружбы народов им. Патриса Лумумбы

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Торшин Иван Юрьевич (р. 1972) — кандидат физико-математических наук, кандидат химических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шоргин Всеволод Сергеевич (р. 1978) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, главный научный сотрудник 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 представленной 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 редактуре должны быть присланы авторами в кратчайшие сроки.

6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или LaTeX (.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 — A4; 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 оформления представлены в интернете:

http://www.ipiran.ru/journal/template_iiep_ssi_2024.zip

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 и 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 hozyaystvo* [*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 resursosberegayushchie 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. Mamikoniants, 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: Friktsionnye 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:

Latypov, 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;
    - указывать в тексте письма название статьи, авторов и журнал, в который направляется статья;
    - 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: iiep@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 LaTeX (.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_iiep_ssi_2024.zip.
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 and
     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; and
- 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 resursosberegayushchie 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.

   Books and other monographs:

   Lindorf, L. S., and L. G. Mamikoniants, 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 slovarya 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;
    - the article title, authors' names, and the journal title, whereto the paper is being submitted, in the text of the email;
    - an article and additional materials have to be attached using the "attach" function; and
    - 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 (to Svetlana Strigina)
http://www.ipiran.ru/english/journal.asp