

# Информатика и её применения

Том 16 Выпуск 3 Год 2022

## СОДЕРЖАНИЕ

Критерий непустоты эпсилон-ядер для нечетких игр с нетрансферабельной полезностью и вычислительные процедуры <b>А. С. Шведов</b>	2
Применение самоорганизующихся нейронных сетей к процессу формирования индивидуальной траектории обучения <b>А. В. Босов</b>	7
Некоторые результаты анализа процесса изменения цены бивалютной корзины на основе методов статистики случайных процессов <b>П. В. Шнурков, М. А. Мигуля</b>	16
Повышение доходности торговли на FOREX с помощью LSTM-идентификации свечных паттернов и индикатора тиковых объемов <b>А. К. Горшенин, Е. И. Гусейнова</b>	26
О достаточных условиях экстремума в многомерных вариационных задачах <b>Н. С. Васильев</b>	39
Последовательный анализ и метрические оценки предельных распределений межузловых потоков в многопользовательской сети <b>Ю. Е. Малашенко</b>	45
Выбор модели при факторизации матрицы данных с пропусками <b>М. П. Кривенко</b>	52
Средовые модели информационных технологий: теоретические основания построения <b>И. М. Зацман</b>	59
Переводческий анализ текста с применением информационных ресурсов: редуцирование спектра моделей перевода в надкорпусных базах данных <b>В. А. Нуриев</b>	68
Об одном подходе к оцениванию скорости сходимости нестационарных марковских моделей систем обслуживания <b>И. А. Ковалёв, Я. А. Сатин, А. В. Синицина, А. И. Зейфман</b>	75
Гистерезисное управление нагрузкой в беспроводных сенсорных сетях <b>Б. Алию, Е. А. Мачнев, Е. В. Мокров</b>	83
Управляемая система массового обслуживания с эластичным трафиком и сигналами для анализа нарезки ресурсов в сети радиодоступа <b>А. С. Власкина, С. А. Бурцева, И. А. Кочеткова, С. Я. Шоргин</b>	90
Метаданные в защищенном электронном документообороте <b>А. А. Грушо, Н. А. Грушо, Е. Е. Тимонина</b>	97
Кинематические модели задач преследования на плоскости методами параллельного сближения и погони <b>А. А. Дубанов, В. А. Нефедова</b>	103
Об авторах	110
Правила подготовки рукописей	112
Requirements for manuscripts	115

# КРИТЕРИЙ НЕПУСТОТЫ ЭПСИЛОН-ЯДЕР ДЛЯ НЕЧЕТКИХ ИГР С НЕТРАНСФЕРАБЕЛЬНОЙ ПОЛЕЗНОСТЬЮ И ВЫЧИСЛИТЕЛЬНЫЕ ПРОЦЕДУРЫ

А. С. Шведов<sup>1</sup>

**Аннотация:** Критерий нового типа для проверки непустоты ядер кооперативных игр был опубликован Жао в 2001 г. Сначала необходимое и достаточное условие было получено Жао для частного случая, когда полезность трансферабельна. В этом случае, как показано Жао, данный критерий легко может быть использован для построения вычислительной процедуры, дающей ответ на вопрос, пусто или не пусто ядро игры, и позволяющей найти дележи, принадлежащие ядру, если ядро не пусто. Затем критерий был обобщен Жао для игр с нетрансферабельной полезностью (НТП-игр). В настоящей работе эти результаты развиваются в нескольких направлениях. Изучается вопрос о непустоте эpsilon-ядер — несколько более общий, чем вопрос о непустоте ядер. Рассматриваются игры с нечеткими выигрышами. Для некоторых классов НТП-игр показана возможность построения вычислительной процедуры, сходной с вычислительной процедурой для игр с трансферабельной полезностью (ТП-игр).

**Ключевые слова:** нечеткое множество; кооперативная игра; эpsilon-ядро; исчерпывающее множество

**DOI:** 10.14357/19922264220301

## 1 Введение

Хорошо известно, что дележи, относящиеся к ядру кооперативной игры, во многих случаях представляют собой оптимальное решение и имеют большое практическое значение (см., например, [1]). Напомним определение ядра НТП-игры для случая, когда выигрыши четкие. Пусть  $N = \{1, \dots, n\}$  — множество игроков. Коалицией называется любое подмножество множества  $N$ . Каждой коалиции  $S$  ставится в соответствие допустимое множество полезностей  $V(S)$ , множество  $V(S)$  принадлежит евклидову пространству, размерность которого равна числу игроков в коалиции  $S$ . Дележ  $(x_1, \dots, x_n) \in V(N)$  блокируется коалицией  $S$ , если каждый из игроков, входящих в коалицию  $S$ , может получить больше в случае отделения этой коалиции. Те дележи, которые не блокируются ни одной коалицией, относятся к ядру игры. Ядро игры может быть как непустым, так и пустым. Частный случай НТП-игр — это ТП-игры. В игре с трансферабельной полезностью каждое из множеств  $V(S)$  состоит из таких наборов выигрышей, что сумма этих выигрышей не больше некоторого значения. Критерий непустоты ядер для ТП-игр — это один из классических результатов теории кооперативных игр (см., например, [1]).

В книге [1] приводится также достаточное, но не необходимое условие для непустоты ядер НТП-игр. При этом в [1, с. 148] сказано следующее: «Теория НТП-игр намного беднее результатами и техниче-

ски гораздо сложнее теории ТП-игр». Вторая часть данного утверждения в определенной мере опровергается в [2]. В работе [2] найдено необходимое и достаточное условие непустоты ядер для НТП-игр, при этом используются относительно несложные математические средства. В работе [3] данный критерий представлен для частного случая, когда полезность трансферабельна, и показано, что этот критерий приводит к вычислительной процедуре, дающей ответ на вопрос, пусто или не пусто ядро игры, и позволяющей найти дележи, принадлежащие ядру, если ядро не пусто. Вопрос сводится к решению некоторой задачи линейного программирования. Другие два критерия непустоты ядер для НТП-игр получены в работах [4, 5]. Отметим, что эти критерии основаны на традиционном подходе изучения сбалансированности кооперативной игры и математически значительно более сложные. Однако алгоритмы для применения к конкретным НТП-играм в работах [4, 5] не даются, и, видимо, данные критерии остаются результатами, относящимися исключительно к теоретической математике.

Приведенное выше условие можно переформулировать так: дележ  $(x_1, \dots, x_n) \in V(N)$  0-блокируется коалицией  $S$ , если каждый из игроков, входящих в коалицию  $S$ , может обеспечить себе потери, меньшие 0, в случае отделения этой коалиции. Такая переформулировка позволяет дать определение  $\varepsilon$ -ядра НТП-игры при любом действительном  $\varepsilon$ . Во-первых, в предыдущей фразе надо заменить «0-

<sup>1</sup>Национальный исследовательский университет Высшая школа экономики, ashvedov@hse.ru

блокируется» на « $\varepsilon$ -блокируется» и «потери, меньшие 0» на «потери, меньшие  $\varepsilon$ ». Те дележи, которые не  $\varepsilon$ -блокируются ни одной коалицией, относятся к  $\varepsilon$ -ядру игры. Ясно, что  $\varepsilon$ -ядра будут непустыми при достаточно больших  $\varepsilon$  и пустыми при достаточно малых  $\varepsilon$ . Таким образом, все множество действительных чисел делится на две полупрямые. Вопрос состоит в нахождении точки, разделяющей эти две полупрямые. (В частности, вопрос о непустоте ядра решается тем, принадлежит 0 левой или правой полупрямой.) Введены  $\varepsilon$ -ядра в [6], детально изучаются в [7], а в [7, с. 307] приводится пример игры, для которой  $\varepsilon$ -ядро непустое при отрицательном  $\varepsilon$ . Понятие  $\varepsilon$ -ядра оказывается полезным в приложениях (см., например, [8]). Для прикладных задач важно не усложнять дело вопросом, существуют или не существуют объекты, которые предполагается использовать, а знать, что такие объекты существуют, и уметь их находить. То минимальное значение  $\varepsilon$ , при котором  $\varepsilon$ -ядро не пусто, несет в себе дополнительную информацию об игре.

Применения теории нечетких множеств в теории игр очень разнообразны. В работе [9] критерий из [5] обобщается для ядер игр с нетрансферабельной полезностью, четкими выигрышами и нечеткими коалициями. Насколько известно автору, ранее игры с нетрансферабельной полезностью и с нечеткими выигрышами не изучались, хотя, скажем, в некооперативной теории игр рассмотрение игр с нечеткими выигрышами широко распространено (см., например, [10]). Моделирование выигрышей нечеткими числами оказывается очень удобным, если результаты игры известны лишь приближенно.

В настоящей работе критерий из [2] обобщается для  $\varepsilon$ -ядер игр с нетрансферабельной полезностью, нечеткими выигрышами и четкими коалициями. Как и в работе [2], используемые математические средства достаточно простые. Основное значение имеет то, что таким образом получается рабочий критерий, применимый для решения конкретных задач. В работе [2] используется только понятие множества, исчерпывающего вниз, и не используется понятие множества, исчерпывающего вверх. В настоящей работе используются оба эти понятия, из-за чего доказательство становится значительно более прозрачным. В разд. 2 дается описание игры. В разд. 3 приводится критерий непустоты  $\varepsilon$ -ядра. В разд. 4 даются примеры и обсуждаются вычислительные процедуры. Показано, что для многих конкретных НТП-игр путем численного решения некоторой задачи математического программирования может быть дан ответ на вопрос, пусто или не пусто  $\varepsilon$ -ядро, а также найдены дележи, принадлежащие  $\varepsilon$ -ядру, если  $\varepsilon$ -ядро не пусто.

## 2 Описание игры

В дальнейшем, говоря о коалиции  $S \subseteq N$ , будем предполагать, что множество  $S$  не является пустым, запись  $S \subset N$  означает, что  $S \neq N$ . Через  $\mathbf{R}^S$  обозначим евклидово пространство, размерность которого равна числу элементов в множестве  $S$ . Под  $\mathbf{R}^N$  будем понимать обычное евклидово пространство  $\mathbf{R}^n$ ;  $\mathbf{R}^S$  является подпространством  $\mathbf{R}^N$ . Координаты векторов, входящих в  $\mathbf{R}^S$ , имеют те же номера, которые имеют игроки из коалиции  $S$ . Например, для  $S = \{2, 3, 7\}$  элементами пространства  $\mathbf{R}^S$  являются векторы  $(x_2, x_3, x_7)$ . Для вектора  $x \in \mathbf{R}^N$  через  $x_S$  обозначается проекция вектора  $x$  на  $\mathbf{R}^S$ .

В приложениях достаточно часто рассматриваются нечеткие множества с функциями принадлежности следующего вида. Пусть  $a$  и  $b$  — действительные числа,  $a \leq b$ . Функция принадлежности равна 1 на полупрямой  $(-\infty, a]$  и равна 0 на полупрямой  $(b, \infty)$ . При  $a < b$  функция принадлежности монотонно убывает на отрезке  $[a, b]$ . Если функция принадлежности линейна на отрезке  $[a, b]$  и убывает от 1 до 0, то нечеткое множество называется трапецидальным. В общем случае можно говорить о трапецидальном нечетком множестве с криволинейной правой границей. Напомним, что  $\eta$ -срезом нечеткого множества, где  $0 < \eta \leq 1$ , называется подмножество, состоящее из тех элементов универсального множества (в данном случае — множества действительных чисел), для которых значение функции принадлежности больше или равно  $\eta$ . Ясно, что для описанного трапецидального нечеткого множества  $\eta$ -срезами являются полупрямые  $\Pi_\eta$ . При  $\eta_1 < \eta_2$  выполняется условие  $\Pi_{\eta_2} \subseteq \Pi_{\eta_1}$ . Но тогда нечеткое множество может быть определено путем задания полупрямых  $\Pi_\eta$  или, по-другому, путем задания некоторой функции аргумента  $\eta$ ,  $\eta \in [0, 1]$ , описывающей правую границу трапеции. Такой подход широко распространен. Оказывается (см., например, [11]), что в ряде случаев от условия  $\Pi_{\eta_2} \subseteq \Pi_{\eta_1}$  при  $\eta_1 < \eta_2$  целесообразно отказаться.

Игра определена, если каждой коалиции  $S \subseteq N$  и каждому  $\eta \in [0, 1]$  поставлено в соответствие множество  $V_\eta(S) \subseteq \mathbf{R}^S$ . Какими свойствами эти множества должны обладать, уточняется ниже. При этом допустимое множество полезностей  $V(S)$  может рассматриваться как нечеткое множество с  $\eta$ -срезами  $V_\eta(S)$ . Тем самым допускается некоторая расплывчатость, неопределенность для допустимого множества полезностей коалиции. Можно рассматривать задачи, где множество  $V(S)$  известно лишь приближенно. Если  $V_\eta(S) = V_0(S)$  для любой коалиции  $S \subseteq N$  и для любого  $\eta \in (0, 1]$ , то игра называется игрой с четкими выигрышами.

Для векторов  $x, y \in \mathbf{R}^S$  записи  $x < y$  и  $x \leq y$  означают, что соответствующие неравенства выполняются покоординатно. Пусть  $M \subseteq \mathbf{R}^S$ .

**Определение 1.** Множество  $M$  называется исчерпывающим вниз, если из условий  $x \in M$  и  $y \leq x$  вытекает, что  $y \in M$ .

**Определение 2.** Множество  $M$  называется исчерпывающим вверх, если из условий  $x \in M$  и  $y \geq x$  вытекает, что  $y \in M$ .

Соотношение между этими понятиями дается следующей теоремой. Рассмотрим непустые замкнутые множества  $V$  и  $G$ , принадлежащие  $\mathbf{R}^S$ . Обозначим через  $V_0$  внутренность множества  $V$ . Будем считать, что выполняются следующие условия:

$$V_0 \cap G = \emptyset; \quad V_0 \cup G = \mathbf{R}^S.$$

**Теорема 1.** Если множество  $V$  является исчерпывающим вниз, то множество  $G$  является исчерпывающим вверх.

**Доказательство.** Предположим, что это не так. Тогда существуют  $x$  и  $y$  такие, что  $x \in G$ ,  $y \geq x$ ,  $y \in V_0$ . Поскольку множество  $V_0$  открытое, это означает, что можно найти точку  $z \in V_0$  такую, что  $z > x$ . Далее, можно найти открытую окрестность  $O$  точки  $x$  такую, что  $z > w$  для любой точки  $w \in O$ . Поскольку множество  $V$  является исчерпывающим вниз,  $O \subseteq V$ . Из этого следует, что  $O \subseteq V_0$ . В частности, это означает, что  $x \in V_0$ . Но это противоречит условию  $x \in G$ .

Теорема 1 доказана.

Будем предполагать, что каждое множество  $V_\eta(S)$  является замкнутым, исчерпывающим вниз и не совпадающим с  $\mathbf{R}^S$ .

Если каждое из множеств  $V_\eta(S)$  состоит из векторов, таких что

$$\sum_{i \in S} x_i \leq v_\eta(S),$$

где  $v_\eta(S)$  — действительное число, то говорят, что НТП-игра переходит в ТП-игру.

### 3 Основной результат

Обозначим через  $1_S$  единичный вектор пространства  $\mathbf{R}^S$ , т.е. вектор, все координаты которого равны 1. В дальнейшем, если не оговорено противное,  $\varepsilon \in \mathbf{R}$ ,  $\eta \in [0, 1]$ .

**Определение 3.** Вектор  $x \in \mathbf{R}^N$   $(\varepsilon, \eta)$ -блокируется коалицией  $S$ , если существует вектор  $x \in V_\eta(S)$  такой, что  $x > z_S + \varepsilon 1_S$ .

Обозначим через  $\partial V_\eta(N)$  границу множества  $V_\eta(N)$ .

**Определение 4.**  $(\varepsilon, \eta)$ -ядром называется множество векторов  $z \in \partial V_\eta(N)$  таких, что не существует коалиции  $S$ , которой вектор  $z$   $(\varepsilon, \eta)$ -блокируется.

Пусть  $\varepsilon_1 < \varepsilon_2$ . Если вектор  $z$   $(\varepsilon_2, \eta)$ -блокируется коалицией  $S$ , то данный вектор и  $(\varepsilon_1, \eta)$ -блокируется коалицией  $S$ . Действительно, существует  $x \in V_\eta(S)$  такой, что  $x > z_S + \varepsilon_2 1_S$ . Отсюда вытекает, что  $x > z_S + \varepsilon_1 1_S$ . Если вектор  $z$  принадлежит  $(\varepsilon_1, \eta)$ -ядру, то вектор  $z$  принадлежит и  $(\varepsilon_2, \eta)$ -ядру. Действительно, если вектор  $z$  не принадлежит  $(\varepsilon_2, \eta)$ -ядру, то вектор  $z$   $(\varepsilon_2, \eta)$ -блокируется некоторой коалицией  $S$ . Тогда вектор  $z$   $(\varepsilon_1, \eta)$ -блокируется коалицией  $S$ . Но это означает, что вектор  $z$  не может принадлежать  $(\varepsilon_1, \eta)$ -ядру.

**Определение 5.** Вектор  $z \in \mathbf{R}^N$  называется  $(\varepsilon, \eta)$ -допустимым для коалиции  $S$ , если он не  $(\varepsilon, \eta)$ -блокируется этой коалицией.

Обозначим через  $H_{\varepsilon, \eta}(S)$  — множество векторов  $z \in \mathbf{R}^N$ , которые являются  $(\varepsilon, \eta)$ -допустимыми для коалиции  $S$ . Очевидно, что  $H_{\varepsilon, \eta}(S)$  является цилиндрическим множеством с основанием  $G_{\varepsilon, \eta}(S)$ , где  $G_{\varepsilon, \eta}(S) \subseteq \mathbf{R}^S$  — непустое множество. Пусть

$$A_{\varepsilon, \eta} = \bigcap_{S \subset N} H_{\varepsilon, \eta}(S).$$

Очевидно, что  $(\varepsilon, \eta)$ -ядро — это  $A_{\varepsilon, \eta} \cap \partial V_\eta(N)$ .

**Теорема 2.** При любых  $\varepsilon \in \mathbf{R}$ ,  $\eta \in [0, 1]$  и  $S \subset N$  множество  $G_{\varepsilon, \eta}(S)$  замкнутое.

**Доказательство.** Предположим, что это не так. Тогда существуют последовательность векторов  $y_k \in \mathbf{R}^S$  и вектор  $y_0 \in \mathbf{R}^S$  такие, что  $y_k \rightarrow y_0$  при  $k \rightarrow \infty$ ,  $y_k \in G_{\varepsilon, \eta}(S)$  при любом натуральном  $k$ ,  $y_0 \notin G_{\varepsilon, \eta}(S)$ . Последнее условие означает, что существует  $x \in V_\eta(S)$  такой, что  $x > y_0 + \varepsilon 1_S$ . Но тогда при достаточно больших  $k$  должны выполняться неравенства  $x > y_k + \varepsilon 1_S$ , что является противоречием.

Теорема 2 доказана.

**Теорема 3.** При любых  $\varepsilon \in \mathbf{R}$ ,  $\eta \in [0, 1]$  и  $S \subset N$  множество  $G_{\varepsilon, \eta}(S)$  является исчерпывающим вверх.

**Доказательство.** Предположим, что это не так. Тогда существуют векторы  $y$  и  $z$  из  $\mathbf{R}^S$  такие, что  $z \in G_{\varepsilon, \eta}(S)$ ,  $y \geq z$  и  $y \notin G_{\varepsilon, \eta}(S)$ . Последнее условие означает, что существует  $x \in V_\eta(S)$  такой, что  $x > y + \varepsilon 1_S$ . Очевидно, отсюда следует, что  $x > z + \varepsilon 1_S$ . Но это является противоречием.

Теорема 3 доказана.

Основным результатом работы является следующая теорема.

**Теорема 4.** Множество  $A_{\varepsilon, \eta} \cap \partial V_\eta(N)$  непустое тогда и только тогда, когда существуют  $x \in \partial V_\eta(N)$  и  $y \in A_{\varepsilon, \eta}$  такие, что  $x \geq y$ .

Доказательство. Пусть выполняется первое условие. Тогда существует  $z$  такой, что  $z \in \partial V_\eta(N)$  и  $z \in A_{\varepsilon, \eta}$ . Очевидно, что  $z \geq z$ .

Пусть выполняется второе условие. В силу теоремы 3 множество  $A_{\varepsilon, \eta}$  является исчерпывающим вверх, поэтому из условия  $x \geq y$  вытекает, что  $x \in A_{\varepsilon, \eta}$ . Следовательно,  $x \in A_{\varepsilon, \eta} \cap \partial V_\eta(N)$ .

Теорема 4 доказана.

Для случая  $\varepsilon = 0$  и игр с четкими выигрышами теорема 4 доказана в [2]. Отметим, что в [2] накладываются некоторые условия на множества  $V(S)$ , которые фактически не нужны.

## 4 Примеры

**Пример 1.** Рассмотрим ТП-игру,  $n = 3$ . Пусть  $\eta$  фиксировано. Тогда

$$\partial V_\eta(N) = \{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = \nu_\eta(\{1, 2, 3\})\}.$$

При любом  $\varepsilon \in \mathbf{R}$  имеем

$$G_{\varepsilon, \eta}(\{1\}) = \{x_1 : x_1 \geq \nu_\eta(\{1\}) - \varepsilon\};$$

$$G_{\varepsilon, \eta}(\{2\}) = \{x_2 : x_2 \geq \nu_\eta(\{2\}) - \varepsilon\};$$

$$G_{\varepsilon, \eta}(\{3\}) = \{x_3 : x_3 \geq \nu_\eta(\{3\}) - \varepsilon\};$$

$$G_{\varepsilon, \eta}(\{1, 2\}) = \{(x_1, x_2) : x_1 + x_2 \geq \nu_\eta(\{1, 2\}) - 2\varepsilon\};$$

$$G_{\varepsilon, \eta}(\{1, 3\}) = \{(x_1, x_3) : x_1 + x_3 \geq \nu_\eta(\{1, 3\}) - 2\varepsilon\};$$

$$G_{\varepsilon, \eta}(\{2, 3\}) = \{(x_2, x_3) : x_2 + x_3 \geq \nu_\eta(\{2, 3\}) - 2\varepsilon\}.$$

Рассмотрим задачу линейного программирования

$$x_1 + x_2 + x_3 \rightarrow \min$$

при ограничениях

$$(x_1, x_2, x_3) \in \bigcap_{S \subset N} H_{\varepsilon, \eta}(S),$$

где  $H_{\varepsilon, \eta}(\{1\}) = \{(x_1, x_2, x_3) : x_1 \in G_{\varepsilon, \eta}(\{1\})\}$  и т.д. Если значение целевой функции в точке минимума не превосходит  $\nu_\eta(N)$ , то выполняется второе условие теоремы 4 и  $(\varepsilon, \eta)$ -ядро непустое. Если значение целевой функции в точке минимума больше, чем  $\nu_\eta(N)$ , то второе условие теоремы 4 не выполняется и  $(\varepsilon, \eta)$ -ядро пустое. Таким образом, для ответа на вопрос, пусто или не пусто  $(\varepsilon, \eta)$ -ядро, достаточно решить обычную задачу линейного программирования.

Отметим, что все результаты работы остаются верными (с очевидными изменениями), если считать, что каждое из множеств  $V_\eta(S)$  принадлежит неотрицательному ортанту  $\mathbf{R}_+^S$ . В следующем примере для наглядности ограничимся этим случаем.

**Пример 2.** Пусть  $\varepsilon = 0$ , значение  $\eta$  фиксировано. Предположим, что существуют функции  $f_S : \mathbf{R}_+^S \rightarrow \mathbf{R}$  такие, что для каждой коалиции  $S$

$$V_\eta(S) = \{x \in \mathbf{R}_+^S : f_S(x) \leq a_S\},$$

где  $a_S$  — некоторое действительное число. При этом функции  $f_S$  являются непрерывными и монотонно возрастающими по каждой координате при любых фиксированных значениях остальных координат. Тогда теорема 4 показывает, что для ответа на вопрос, пусто или не пусто ядро игры, достаточно решить задачу математического программирования

$$f_N(x) \rightarrow \min$$

при ограничениях

$$f_S(x_S) \geq a_S$$

для всех  $S \subset N$ . Пусть  $f^*$  — значение целевой функции в точке минимума. Тогда ядро игры непустое при  $f^* \leq a_N$  и пустое в противоположном случае. Трудностью при численном решении приведенной задачи математического программирования может оказаться то, что число ограничений быстро растет при увеличении  $n$ . В частности, могут рассматриваться функции  $f_S(x) = \sum_{i \in S} x_i$  при всех  $S \subseteq N$ , тогда это ТП-игра. Также методический интерес представляет случай  $f_S(x) = \sum_{i \in S} c_{iS} x_i^2$ , где все числа  $c_{iS}$  положительные. По-видимому, данный подход является единственным, который позволяет в зависимости от значений  $c_{iS}$  и  $a_S$  дать ответ на вопрос, пусто или не пусто ядро игры, а также найти дележи, принадлежащие ядру, если ядро не пусто.

## 5 Заключение

Множество  $A_{\varepsilon, \eta} \cap \partial V_\eta(N)$  можно рассматривать как  $\eta$ -срез некоторого нечеткого множества. Изучение свойств этих нечетких множеств представляет собой предмет будущих исследований.

## Литература

1. Мулен Э. Кооперативное принятие решений: Аксиомы и модели / Пер. с англ. — М.: Мир, 1991. 464 с. (Moulin H. Axioms of cooperative decision making. — Cambridge: Cambridge University Press, 1988. 332 p.)
2. Zhao J. New conditions for core existence in coalitional NTU games. — Ames, IA, USA: Department of Economics Iowa State University, 2001. 16 p.
3. Zhao J. The relative interior of the base polyhedron and the core // Econ. Theory, 2001. Vol. 18. P. 635–648.
4. Keiding H., Thorlund-Petersen L. The core of a cooperative game without side payments // J. Optimiz. Theory App., 1987. Vol. 54. P. 273–288.
5. Predtetchinski A., Herings P.J.-J. A necessary and sufficient condition for non-emptiness of the core of a non-



- transferable utility game // *J. Econ. Theory*, 2004. Vol. 116. P. 84–92.
6. Shapley L., S., Shubik M. Quasi-cores in a monetary economy with nonconvex preferences // *Econometrica*, 1966. Vol. 34. P. 805–827.
  7. Maschler M., Peleg B., Shapley L. S. Geometric properties of the kernel, nucleolus, and related solution concepts // *Math. Oper. Res.*, 1979. Vol. 4. P. 303–338.
  8. Mochaourab R., Jorswieck E. Coalitional games in MISO interference channels: Epsilon-core and coalition structure stable set // *IEEE T. Signal Proces.*, 2014. Vol. 62. P. 6507–6520.
  9. Liu J., Liu X. A necessary and sufficient condition for an NTU fuzzy game to have a non-empty fuzzy core // *J. Math. Econ.*, 2013. Vol. 49. P. 150–156.
  10. Larbani M. Non cooperative fuzzy games in normal form: A survey // *Fuzzy Set. Syst.*, 2009. Vol. 160. P. 3184–3210.
  11. Shvedov A. S. Instrumental variables estimation of fuzzy regression models // *J. Intell. Fuzzy Syst.*, 2019. Vol. 36. P. 5457–5462.

Поступила в редакцию 04.07.21

## A CONDITION FOR NON-EMPTINESS OF THE EPSILON-CORE OF A NONTRANSFERABLE UTILITY FUZZY GAME AND COMPUTATIONAL SCHEMES

A. S. Shvedov

National Research University Higher School of Economics, 20 Myasnitskaya Str., Moscow 101000, Russian Federation

**Abstract:** Zhao (2001) suggested a new condition for non-emptiness of the core of a cooperative game. At first, a necessary and sufficient condition was found by Zhao for a particular case of games with transferable utility. In this case, as was shown by Zhao, the condition can be easily used for construction of a computational scheme which decides if the core of a game is empty or non-empty and gives imputations belonging to the core if the core is non-empty. Then, the condition was generalized by Zhao for games with nontransferable utility. In this paper, the results are generalized in few directions. The problem of non-emptiness of the epsilon-core which is somewhat more general than the problem of non-emptiness of the core is studied. Games with fuzzy payoffs are considered. For some classes of games with nontransferable utility, possibility of construction of a computational scheme which is similar to the computational scheme for games with transferable utility is established.

**Keywords:** fuzzy set; cooperative game; epsilon-core; comprehensive set

**DOI:** 10.14357/19922264220301

### References

1. Moulin, H. 1988. *Axioms of cooperative decision making*. Cambridge: Cambridge University Press. 332 p.
2. Zhao, J. 2001. *New conditions for core existence in coalitional NTU games*. Ames, IA: Department of Economics Iowa State University. 16 p.
3. Zhao, J. 2001. The relative interior of the base polyhedron and the core. *Econ. Theory* 18:635–648.
4. Keiding, H., and L. Thorlund-Petersen. 1987. The core of a cooperative game without side payments. *J. Optimiz. Theory App.* 54:273–288.
5. Predtetchinski, A., and P. J.-J. Herings. 2004. A necessary and sufficient condition for non-emptiness of the core of a non-transferable utility game. *J. Econ. Theory* 116: 84–92.
6. Shapley, L. S., and M. Shubik. 1966. Quasi-cores in a monetary economy with nonconvex preferences. *Econometrica* 34:805–827.
7. Maschler, M., B. Peleg, and L. S. Shapley. 1979. Geometric properties of the kernel, nucleolus, and related solution concepts. *Math. Oper. Res.* 4:303–338.
8. Mochaourab, R., and E. Jorswieck. 2014. Coalitional games in MISO interference channels: epsilon-core and coalition structure stable set. *IEEE T. Signal Proces.* 62:6507–6520.
9. Liu, J., and X. Liu. 2013. A necessary and sufficient condition for an NTU fuzzy game to have a non-empty fuzzy core. *J. Math. Econ.* 49:150–156.
10. Larbani, M. 2009. Non cooperative fuzzy games in normal form: A survey. *Fuzzy Set. Syst.* 160:3184–3210.
11. Shvedov, A. S. 2019. Instrumental variables estimation of fuzzy regression models. *J. Intell. Fuzzy Syst.* 36:5457–5462.

Received July 4, 2021

### Contributor

**Shvedov Alexey S.** (b. 1956) — Doctor of Science in physics and mathematics, professor, Faculty of Economics, National Research University Higher School of Economics, National Research University Higher School of Economics, 20 Myasnitskaya Str., Moscow 101000, Russian Federation; ashvedov@hse.ru

# ПРИМЕНЕНИЕ САМООРГАНИЗУЮЩИХСЯ НЕЙРОННЫХ СЕТЕЙ К ПРОЦЕССУ ФОРМИРОВАНИЯ ИНДИВИДУАЛЬНОЙ ТРАЕКТОРИИ ОБУЧЕНИЯ\*

А. В. Босов<sup>1</sup>

**Аннотация:** Рассмотрена задача динамической классификации обучающихся в рамках поддержки процесса формирования индивидуальной траектории пользователя электронной обучающей системы. Модель обучения рассчитана на смешанную форму ведения образовательной деятельности с частичной самостоятельной работой и периодическими контрольными мероприятиями в форме тестов, частичной очной работой с выполнением контрольных работ и сдачей зачетов. Целью классификации ставится определение категории обучаемого по результатам очередного контрольного мероприятия. Семантика категорий предполагает возможность индивидуального выбора разного уровня сложности заданий на очередном шаге обучения. Направлением совершенствования существующих методик классификации определен отказ от накопления и использования статистики предыдущих (других) групп обучающихся. Отсутствии образцов правильной классификации обосновало применение самоорганизующихся нейронных сетей. Для решения использованы карты Кохонена, стандартный вариант которых адаптирован к имеющейся модели обучения и к задаче учета субъективной оценочной политики преподавателя. Описаны три варианта алгоритма самообучения. Выполнены экспериментальные исследования, их результаты проиллюстрированы.

**Ключевые слова:** электронное средство обучения; самоорганизующаяся нейронная сеть; карта Кохонена; задача классификации; индивидуальная траектория обучения

**DOI:** 10.14357/19922264220302

## 1 Введение

Типовым инструментом современных электронных обучающих систем, используемых как дистанционно, так и в дополнение к традиционным формам занятий, является средство построения индивидуальных траекторий обучаемых. Как правило, такое средство учитывает промежуточные результаты обучения и формирует рекомендации для настройки контента системы, например для определения уровня сложности контрольных заданий. Реализуемые этими средствами методы и алгоритмы составляют значимую часть теории тестирования (Item Response Theory). Чаще всего для описания результативности процесса обучения используют вероятностные модели, которые позволяют адаптировать процесс тестирования под уровень знаний пользователей [1–6]. Результатом адаптации является индивидуальная траектория обучения (тестирования), которая представляет собой привязанную ко времени последовательность решений задачи классификации обучаемых — отнесение их к различным категориям успешности

обучения. По результатам проведенной классификации разным категориям пользователей предлагаются разные траектории дальнейшего тестирования или обучения.

Поставленная перед обучающей системой задача классификации относится к типовым, поэтому для ее решения применяются самые распространенные средства. Так, классический байесовский классификатор [7] успешно реализован в системе дистанционного обучения, использующей рейтинг пользователя, который рассчитывается в процессе обучения как свертка показанных промежуточных результатов [8]. Недостаток этого варианта категоризации обучающихся и других средств формирования индивидуальной траектории, основанных на статистике, состоит в том, что для реализации соответствующих алгоритмов требуется знание некоторых вероятностных характеристик. В [8] это условные вероятности сохранения/изменения рейтинга после выполнения очередного контрольного мероприятия. Точно знать эти характеристики нельзя, поэтому их заменяют статистическими оценками, рассчитанными по имеющимся или накопленным

\* Исследование выполнено при финансовой поддержке Российского научного фонда (проект 22-28-00588). Работа проводилась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика» ФИЦ ИУ РАН, Москва).

<sup>1</sup> Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук, avbosov@ipiran.ru

данным. Соответственно, к недостаткам или трудностям применения байесовского классификатора при категоризации обучающихся можно отнести следующее:

- частотные оценки вероятностей требуют наличия большого объема достоверной статистики;
- любые изменения количества, порядка, характера контрольных мероприятий приводят к искажению накопленной статистики, невозможности ее использования;
- предполагается совпадение вероятностных характеристик разных групп обучающихся.

Цель данной статьи состоит в том, чтобы предложить способ классификации обучающихся по результатам использования электронного средства обучения, свободной от перечисленных недостатков. Для этого байесовский классификатор предлагается заменить не менее типовым инструментом — самообучающейся искусственной нейронной сетью. Поскольку решаемая задача классификации типична, то и вариант сети был выбран самый распространенный — карта самоорганизации Кохонена [9]. Наиболее известные финансовые приложения [10] этих карт к настоящему времени пополнились за счет самых разных сфер применения. Фактически к приложениям самообучающихся карт надо относить практически любой технический анализ, связанный с классификацией. Так, общепринята практика категоризации зачисленных студентов, сотрудников, прошедших обучение, действующих инженеров [11–14]. Вклад этих работ и многих других состоит не столько в развитии самой технологии самоорганизующихся карт и средств кластеризации, сколько в расширении сферы применения известных алгоритмов, подтверждения их эффективности в новых приложениях. Данная статья представляет результаты исследования такого же характера. Цель работы — применить существующий механизм классификации с помощью карт Кохонена в действующей системе дистанционного обучения CLASS.NET [15]. Для этого предложены несколько вариантов адаптации модели процесса обучения и алгоритма самоорганизации карты Кохонена, проведены практические эксперименты.

## 2 Модель процесса обучения и задача категоризации обучающихся

Модель процесса обучения предполагает смешанное дистанционно-очное обучение и включа-

ет  $T$  контрольных мероприятий. Типичная схема обучения для предлагаемой модели состоит в том, что имеющийся по изучаемому предмету курс разбит на несколько частей, каждая часть завершается очными мероприятиями (промежуточный зачет и/или контрольная работа), а в промежутках учащиеся выполняют по несколько заочных мероприятий (тестирование). После очередного контрольного мероприятия (шага обучения) для каждого студента определяется категория, например: «неуспевающий» (1); «низкий уровень» (2); «средний уровень» (3); «высокий уровень» (4). В рамках данной работы будут использованы указанные четыре категории студентов, которые обозначаются номером  $k \in K = \{1, 2, 3, 4\}$ . Ограничений общности это предположение не влечет, а для приведенных ниже иллюстративных расчетов надо зафиксировать число категорий и наделить их ясным содержанием.

Итак, на каждом шаге обучения должна решаться задача классификации: каждый студент должен быть отнесен к одной из категорий на основании продемонстрированных им результатов. Соответствующие моменты времени обозначаются  $t = 1, \dots, T$ . Типовое использование таких категорий в электронных обучающих системах состоит в определении уровня сложности следующего этапа/шага обучения и/или контрольного мероприятия. Таким образом обеспечивается соответствие контрольного контента уровню подготовки студента и формирование индивидуальной траектории обучения.

Будем предполагать, что для каждого  $t$  задано значение  $x_t \in \{т, з, к\}$ , определяющее тип контрольного мероприятия: т — тест; з — зачет; к — контрольная работа. Каждый из  $S$  обучающихся в момент  $t$  получает очередную отметку  $e_t^s$ ,  $s = 1, \dots, S$ , которая в зависимости от  $x_t$  характеризуется своей областью значений. Для определенности будем предполагать, что для  $x_t = т$  значения  $e_t^s \in \{0, 1, \dots, 100\}$ , для  $x_t = з$  значения  $e_t^s \in \{0, 1\}$ , для  $x_t = к$  значения  $e_t^s \in \{2, 3, 4, 5\}$ ; т.е. тестирование оценивается по 100 балльной шкале, зачет — традиционно зачет/незачет, контрольная работа — отметками неудовлетворительно, удовлетворительно, хорошо, отлично. Соответственно, задача классификации обучающихся в момент  $t$  состоит в определении категории  $k_t^s \in K$  для каждого  $s$ -го студента.

Вариант решения этой задачи, предложенный в [8], состоит в том, чтобы свернуть текущий вектор оценок  $(e_1^s, \dots, e_t^s)'$ ,  $x'$  — транспонированный вектор, в некоторую скалярную величину — текущий рейтинг, решение об отнесении студента с этим рейтингом к категории  $k$  обеспечивается байесовским правилом, в котором применяются



статистические характеристики (частотные оценки апостериорных вероятностей), вычисленные по результатам, продемонстрированным предыдущими группами обучающихся. Принципиальное предложение данной работы состоит в том, чтобы для классификации использовать только результаты, демонстрируемые одной небольшой группой обучающихся, и классифицировать студентов, сравнивая результаты обучения между собой. Формально такая задача состоит в определении отображения

$$\left\{ \left\{ e_{\tau}^s \right\}_{\tau=1}^t \right\}_{s=1}^S \rightarrow \left\{ k_t^s \right\}_{s=1}^S, \quad k_t^s \in K, \quad t = 1, \dots, T. \quad (1)$$

Если не учитывать масштабы и семантику входного набора данных  $\left\{ \left\{ e_{\tau}^s \right\}_{\tau=1}^t \right\}_{s=1}^S$ , т.е. величин  $\left\{ x_{\tau} \right\}_{\tau=1}^t$  типов контроля, то задача (1) становится типичной постановкой для построения одномерной самоорганизующейся карты Кохонена [9]. Необходимость учета порядка и типов контрольных мероприятий требует небольшой адаптации стандартного алгоритма самообучения. Кроме того, требуется еще учесть динамический характер модели обучения, что классификация выполняется вместе с очередными шагами обучения и контроля. Эта особенность также требует учета при построении карты.

### 3 Варианты реализации карт Кохонена

**Первый** из предлагаемых вариантов алгоритма самообучения отличается от классического только тем, что для определения расстояния между входными векторами и нейронами карты вместо евклидовой нормы  $|x|^2 = xx'$  используется функция нормировки  $\|x\|_Q^2 = x'Qx$  для симметричной неотрицательно определенной матрицы  $Q$ , т.е. единичной матрице  $Q = 1$  соответствует евклидова норма  $\|x\|_1^2 = |x|^2$ . Матрица  $Q$  будет использована с разными целями. В первом варианте ее задача — масштабировать отметки. Во всех вариантах эта матрица будет диагональной, в первом — на диагонали будут  $100^{-2}$ , если соответствующая величина  $x_t = \tau; 1$ , если  $x_t = 3; 5^{-2}$ , если  $x_t = \kappa$ . Поясним это на примере, воспользовавшись следующим распределением видов контрольных мероприятий:

$$\left. \begin{aligned} x_1, x_2, x_3, x_4 = \tau; \quad x_5 = 3; \quad x_6 = \kappa; \\ x_7, x_8, x_9 = \tau; \quad x_{10} = 3; \quad x_{11} = \kappa; \\ x_{12}, x_{13} = \tau; \quad x_{14} = 3; \quad x_{15} = \kappa; \quad x_{16}, x_{17} = \tau, \end{aligned} \right\} \quad (2)$$

т.е. в первой части курса четыре теста, зачет, контрольная; во второй части — три теста, зачет, кон-

трольная; в третьей части — два теста, зачет, контрольная, два теста. Этот порядок соответствует примерам, рассмотренным в [8]. Таким образом, для 17 указанных видов контрольных мероприятий матрица нормировки  $Q = Q^{1/2}Q^{1/2}$ :

$$Q^{1/2} = \text{diag} \left\{ \frac{1}{100}, \frac{1}{100}, \frac{1}{100}, \frac{1}{100}, 1, \frac{1}{5}, \frac{1}{100}, \frac{1}{100}, \frac{1}{100}, 1, \frac{1}{5}, \frac{1}{100}, \frac{1}{100}, \frac{1}{100}, 1, \frac{1}{5}, \frac{1}{100}, \frac{1}{100} \right\}. \quad (3)$$

Выбираем момент  $t$  и отвечающую ему матрицу нормировки  $Q_t = \text{diag} \{Q_{1,1}, \dots, Q_{t,t}\}$ . Далее следуем обычному алгоритму. Обозначим  $W$  матрицу весов карты размера  $4 \times t$ ,  $W_k$  —  $k$ -ю строку этой матрицы, т.е. вектор весов  $k$ -го нейрона,  $k = 1, 2, 3, 4$ . Выполняем последовательность итераций, нумеруя их  $i = 1, \dots, I$ . На  $i$ -й итерации последовательно для каждого  $s$ -го студента,  $s = 1, \dots, S$ , находится ближайший нейрон

$$k_{\min}^i(s) = \arg \min_{k=1,2,3,4} \|E_t^s - W_k(i)\|_{Q_t}^2,$$

где  $E_t^s = (e_1^s, \dots, e_t^s)'$  — вектор, составленный из отметок  $s$ -го студента к текущему моменту  $t$ ;  $W_k(i)$  —  $k$ -я строка матрицы  $W$ , отвечающая  $i$ -й итерации.

Далее для каждого  $k$ -го нейрона вычисляется значение функции соседства  $h(k, k_{\min}^i(s))$  и корректируется вектор весов:

$$\left. \begin{aligned} h(k, k_{\min}^i(s)) &= \exp \left\{ -\frac{|k - k_{\min}^i(s)|^2}{2\sigma_i^2} \right\}, \\ W_k(i) &= W_k(i) + \eta_i h(k, k_{\min}^i(s)) (E_t^s - W_k(i)), \\ s &= s + 1, \text{ если } s < S, \\ &\text{иначе } (s = S) \quad i = i + 1, \quad s = 1. \end{aligned} \right\} \quad (4)$$

Отметим, что в (4), в отличие от обычного алгоритма, в рамках одной итерации обрабатываются все входные воздействия (отметки всех студентов), причем с одинаковыми параметрами обучения  $\sigma_i$  и  $\eta_i$ . Такая организация алгоритма самообучения позволяет гарантированно учитывать отметки всех студентов с одинаковыми весами, уравнивая вклад в классификацию всех обучающихся. Переход к  $(i + 1)$ -й итерации выполняется, когда матрица весов скорректировалась всеми имеющимися  $S$  отметками.

Выбор параметров обучения выполнен согласно самым распространенным рекомендациям [16–18]:

$$\sigma_i = \sigma_0 \exp \left\{ -\frac{i}{\delta} \right\}, \quad \delta = \frac{1000}{\ln \sigma_0}, \quad \eta_i = \eta_0 \exp \left\{ -\frac{i}{\Delta} \right\}.$$

Более того, в расчетах даже не использовалась величина  $\delta$  и полагалось  $\sigma_i = \text{const} = 1$ . Осталь-

ные значения ( $\eta_0 = 0,1$  и  $\Delta = 1000$ ) — обычные рекомендации для  $I = 1000$  итераций.

В рамках этого же варианта алгоритма самообучения можно добавить возможность назначения весов значимости для разных типов контрольных мероприятий. Например, с учетом очного характера можно отметку за зачет учесть с двойным весом, а отметку за контрольную работу — с тройным. Для этого (3) заменяется на

$$Q^{1/2} = \text{diag} \left\{ \frac{1}{100}, \frac{1}{100}, \frac{1}{100}, \frac{1}{100}, 2, \frac{3}{5}, \frac{1}{100}, \frac{1}{100}, \frac{1}{100}, 2, \frac{3}{5}, \frac{1}{100}, \frac{1}{100} \right\}. \quad (5)$$

Характер влияния этой и последующих модификаций алгоритма иллюстрируется в следующем разделе на примерах.

Следующий, **второй**, вариант алгоритма или, точнее говоря, модели классификации устраняет два возможных недостатка первого варианта. Во-первых, нужно обеспечить корректную с точки зрения преподавателя категоризацию студентов в группе, в которой нет четко представленных всех четырех категорий обучаемых. Например, в группе без «неуспевающих» категория 1 должна остаться пустой, в то время как применение алгоритма (4), вполне возможно, все равно даст четыре категории, распределив студентов без учета оценочного контекста, вложенного в категории. В связи с этим можно увидеть второй недостаток — вкладываемый в категории преподавателем контекст «неуспевающий», «низкий уровень», «средний уровень» и «высокий уровень» может не гарантироваться алгоритмом (4). Другая формулировка этой же проблемы состоит в отсутствии учета алгоритмом (4) экспертного (преподавательского) представления о содержании понятий «неуспевающий», «низкий уровень», «средний уровень» и «высокий уровень».

Модифицируя (4) для устранения указанных недостатков, введем «экспертные» траектории  $(IE_T^s)^k$ ,  $k \in K$ , смысл которых состоит в том, чтобы представить «идеальный» экземпляр  $k$ -й категории:

$$\left. \begin{aligned} (IE_{17}^s)^1 &= (30, 30, 30, 30, 0, 2, 30, 30, 30, \\ &\quad 0, 2, 30, 30, 0, 2, 30, 30)^s; \\ (IE_{17}^s)^2 &= (50, 50, 50, 50, 0, 3, 50, 50, 50, \\ &\quad 0, 3, 50, 50, 0, 3, 50, 50)^s; \\ (IE_{17}^s)^3 &= (70, 70, 70, 70, 0, 4, 70, 70, 70, \\ &\quad 0, 4, 70, 70, 0, 4, 70, 70)^s; \\ (IE_{17}^s)^4 &= (90, 90, 90, 90, 0, 5, 90, 90, 90, \\ &\quad 0, 5, 90, 90, 0, 5, 90, 90)^s. \end{aligned} \right\} \quad (6)$$

Отметим, что эти «идеальные» траектории решают вопрос с начальными условиями, так как их можно использовать в (4) для начала итерирования, т.е. для величин  $W_k(0)$ . Во втором варианте траектории (6) используются в качестве входных векторов:

$$\left. \begin{aligned} W_k(i) &= W_k(i) + \eta_i h(k, k_{\min}^i(s)) (E_t^s - W_k(i)), \\ &\quad s = 1, \dots, S; \\ W_k(i) &= W_k(i) + \\ &\quad + \eta_i h(k, k_{\min}^i(s)) \left( (IE_t^s)^k - W_k(i) \right), \\ &\quad k = 1, \dots, K, s = 1, \dots, \left[ \frac{S}{k} \right]; \\ W_k(i+1) &= W_k(i), \text{ если } s = S, k = K. \end{aligned} \right\} \quad (7)$$

Таким образом, для рассматриваемого случая  $k = 1, 2, 3, 4$  каждая из четырех «идеальных» траекторий  $(IE_{17}^s)^k$  участвует в  $i$ -й итерации алгоритма, повторяясь в качестве входа в количестве  $25\%$ ,  $[S/k] \cdot 100\%$ , от числа имеющихся траекторий, т.е. числа студентов.

Наконец, последний, **третий**, вариант модели классификации учитывает фактор устаревания полученных студентом отметок или большей актуальности (приоритетности) последних, старших по времени результатов. Максимально просто добиться учета этого фактора можно, ограничившись числом учитываемых оценок, положив  $Q_t = \text{diag} \{0, \dots, 0, Q_{t,t-\varepsilon+1}, \dots, Q_{t,t}\}$ , т.е. учитывая только  $\varepsilon$  последних по времени отметок. Более гибкий способ состоит в использовании параметра устаревания  $\alpha$ ,  $0 < \alpha \leq 1$ , и матрицы нормировки

$$Q_t^{1/2} = \text{diag} \left\{ \alpha^{t-1} Q_{1,1}^{1/2}, \dots, \dots, \alpha^2 Q_{t-2,t-2}^{1/2}, \alpha Q_{t-1,t-1}^{1/2}, Q_{t,t}^{1/2} \right\}. \quad (8)$$

Такая нормировка означает динамическое изменение входных данных — умножение отметок до текущего контрольного мероприятия  $t$  на величину  $\alpha$ .

## 4 Результаты практических экспериментов

Для иллюстрации результатов работы предложенных классификаторов использовалась немногочисленная адаптированная модель обучения, реализованная в системе дистанционного обучения CLASS.NET [15] и тот же набор данных, что и в [8]. В расчетах были использованы отметки по схеме контроля (2) для  $S = 36$  студентов, характерные траектории для 5 студентов приведены в табл. 1.

**Таблица 1** Примеры траекторий результатов обучения

$s$	$e_1^s$	$e_2^s$	$e_3^s$	$e_4^s$	$e_5^s$	$e_6^s$	$e_7^s$	$e_8^s$	$e_9^s$	$e_{10}^s$	$e_{11}^s$	$e_{12}^s$	$e_{13}^s$	$e_{14}^s$	$e_{15}^s$	$e_{16}^s$	$e_{17}^s$
2	35	20	25	20	1	2	34	24	20	0	2	34	25	0	2	34	14
7	58	73	54	49	1	3	56	58	69	0	2	45	67	1	3	51	72
13	54	69	55	65	1	4	71	69	69	1	4	67	55	1	4	66	61
17	69	67	74	68	1	4	99	91	92	1	5	87	93	1	5	94	87
18	89	94	78	98	1	5	89	88	89	1	5	97	94	1	3	77	65

**Таблица 2** Результаты классификации — вариант 1, матрица нормировки (3)

$s$	$k_1^s$	$k_2^s$	$k_3^s$	$k_4^s$	$k_5^s$	$k_6^s$	$k_7^s$	$k_8^s$	$k_9^s$	$k_{10}^s$	$k_{11}^s$	$k_{12}^s$	$k_{13}^s$	$k_{14}^s$	$k_{15}^s$	$k_{16}^s$	$k_{17}^s$
2	1	1	1	1	2	2	2	1	1	1	1	1	1	1	1	1	1
7	2	3	2	2	3	3	2	2	2	2	2	2	2	2	2	2	2
13	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
17	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4
18	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

**Таблица 3** Результаты классификации — вариант 1, матрица нормировки (5)

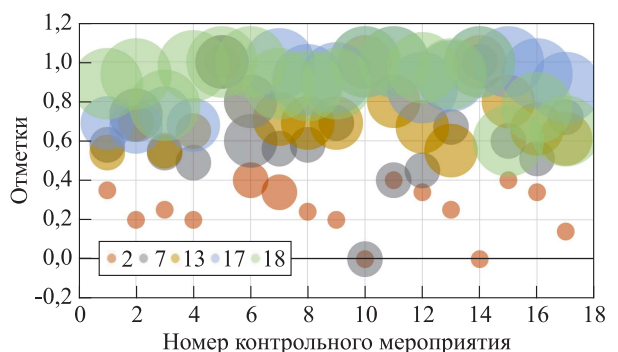
$s$	$k_1^s$	$k_2^s$	$k_3^s$	$k_4^s$	$k_5^s$	$k_6^s$	$k_7^s$	$k_8^s$	$k_9^s$	$k_{10}^s$	$k_{11}^s$	$k_{12}^s$	$k_{13}^s$	$k_{14}^s$	$k_{15}^s$	$k_{16}^s$	$k_{17}^s$
2	1	1	1	1	3	2	2	2	2	2	2	2	2	1	1	1	1
7	2	3	2	2	3	3	3	3	3	2	2	2	2	2	2	2	2
13	2	2	2	2	3	3	3	3	3	4	4	4	3	4	4	4	3
17	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4
18	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

Выбранные для таблицы траектории субъективно оценивались как типичные для классов «неуспевающий» ( $s = 2$ ), «низкий уровень» ( $s = 7$ ), «средний уровень» ( $s = 13$ ) и «высокий уровень» ( $s = 17$  и  $18$ ). Причем траектории  $s = 17$  и  $18$  иллюстрируют явно выраженные переходы между классами  $k = 4$  в класс  $k = 3$  и наоборот, они должны помочь иллюстрировать разницу предложенных классификаторов. В электронном представлении, в частности в программной реализации, данные (строки таблиц) дополнительно выделяются цветом, который также используется на рисунках для визуализации траекторий. Результаты классификации первым вариантом алгоритма (4) с матрицей нормировки (3) приведены в табл. 2, с матрицей нормировки (5) — в табл. 3.

В целом назначаемые классификатором категории вполне соответствуют ожидаемым. Причем более «лояльными» оказались результаты табл. 2. Это очевидная реакция на увеличение веса зачета (см.  $k_5^2$ ,  $k_5^7$  и  $k_{10}^{13}$ ). Также можно отметить «правильную» динамику траектории  $s = 17$ , а именно переход из класса  $k = 3$  в класс  $k = 4$ , и отсутствие обратного перехода для траектории  $s = 18$ . В последнем случае классификатор не отреагировал на ухудшение результатов, здесь сработал эффект «старых заслуг». Наконец, в табл. 3 еще выделяются изменения в траектории  $s = 13$ : классификатор явно

завышает результаты. Но это субъективная оценка. На самом деле по имеющимся данным классификатор не имеет возможности оценить верхнюю границу отметок, «помочь» ему в этом должен второй вариант алгоритма.

Визуализировать в удобной для преподавателя форме результаты классификации предлагается следующими двумя схемами. На рис. 1 иллюстрируется зависимость от времени (номера контрольного мероприятия) результатов и обучения, и классификации. Для каждого студента траектория оформлена серией кругов. Центр каждого круга расположен на высоте, соответствующей полученной отметке



**Рис. 1** Динамика результатов обучения — категории и отметки

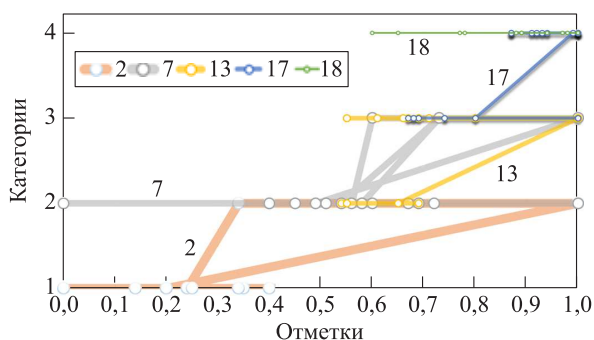


Рис. 2 Индивидуальные результаты обучения

(шкала слева), а размер — категории: круг самого маленького диаметра отвечает категории 1, самого большого — категории 4. В электронном представлении круг дополнительно окрашивается цветом, назначенным соответствующей траектории (как и в табл. 1–3). Такая форма визуализации хорошо характеризует ситуацию в целом. Если же важно более точно визуализировать совокупность отметок и категорий каждого студента, то лучше использовать представление типа фазового портрета,

т.е. линии, отображающие только отметки и категории. Этот вариант визуализации результата классификации показан на рис. 2. Категории на этом рисунке представляет шкала слева, отметки отображаются точками, которые соединены линиями, связывающими соседние по времени отметки  $e_t^s$  и  $e_{t+1}^s$ . Оба рисунка соответствуют результатам табл. 1.

Результаты расчетов для второго варианта алгоритма (7) приведены в табл. 4. Здесь для расчетов к имеющимся данным табл. 1 были добавлены 36 траекторий (6), по 9 каждого предложенного типа.

Этот вариант оправдывает ожидания, разделяя высокие категории  $k = 3$  и 4. Так, траектория  $s = 13$  относится к «правильной» категории  $k = 3$ . Такой же эффективный учет субъективного представления (6) демонстрируют и остальные траектории, не представленные в табл. 1. Видимый недостаток здесь остается у траектории  $s = 18$ , так как избавиться от эффекта «старых заслуг» не получается.

Последнюю группу расчетов представляет табл. 5 — результаты расчетов для третьего варианта алгоритма, т.е. с матрицей нормировки (8)

Таблица 4 Результаты классификации — вариант 2

$s$	$k_1^s$	$k_2^s$	$k_3^s$	$k_4^s$	$k_5^s$	$k_6^s$	$k_7^s$	$k_8^s$	$k_9^s$	$k_{10}^s$	$k_{11}^s$	$k_{12}^s$	$k_{13}^s$	$k_{14}^s$	$k_{15}^s$	$k_{16}^s$	$k_{17}^s$
2	1	1	1	1	2	2	2	2	2	1	1	1	1	1	1	1	1
7	2	3	2	2	3	3	3	2	2	2	2	2	2	2	2	2	2
13	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
17	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4
18	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

Таблица 5 Результаты классификации — вариант 3

$s$	$k_1^s$	$k_2^s$	$k_3^s$	$k_4^s$	$k_5^s$	$k_6^s$	$k_7^s$	$k_8^s$	$k_9^s$	$k_{10}^s$	$k_{11}^s$	$k_{12}^s$	$k_{13}^s$	$k_{14}^s$	$k_{15}^s$	$k_{16}^s$	$k_{17}^s$
$\alpha = 0,75$																	
2	1	1	1	1	3	2	2	1	1	1	1	1	1	1	1	1	1
7	2	3	2	2	3	3	2	2	2	1	1	1	1	3	3	3	3
13	2	3	2	3	3	3	4	3	3	3	3	3	3	3	4	3	3
17	3	3	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4
18	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	3
$\alpha = 0,5$																	
2	1	1	1	1	3	2	1	1	1	1	1	1	1	1	1	1	1
7	2	3	2	2	3	2	2	2	3	1	1	1	2	4	3	2	3
13	2	3	2	3	4	3	4	3	3	4	3	3	2	4	4	3	3
17	3	3	4	3	4	3	4	4	4	4	4	4	4	4	4	4	4
18	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3
$\alpha = 0,25$																	
2	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1
7	2	3	2	2	4	2	2	2	3	1	1	1	3	4	2	1	3
13	2	3	2	3	4	3	3	3	3	4	3	3	2	4	3	2	2
17	3	3	4	3	4	3	4	4	4	4	4	4	4	4	4	4	4
18	4	4	4	4	4	4	4	4	4	4	4	4	4	4	2	3	3



**Таблица 6** Результаты классификации — смешанный вариант 2–3,  $\alpha = 0,75$

$s$	$k_1^s$	$k_2^s$	$k_3^s$	$k_4^s$	$k_5^s$	$k_6^s$	$k_7^s$	$k_8^s$	$k_9^s$	$k_{10}^s$	$k_{11}^s$	$k_{12}^s$	$k_{13}^s$	$k_{14}^s$	$k_{15}^s$	$k_{16}^s$	$k_{17}^s$
2	1	1	1	1	3	3	2	1	1	1	1	1	1	1	1	1	1
7	2	3	2	2	3	3	3	2	3	1	1	1	1	3	3	2	3
13	2	3	2	3	3	3	3	3	3	4	3	3	3	3	3	3	3
17	3	3	3	3	4	3	4	4	4	4	4	4	4	4	4	4	4
18	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3

и значениями параметра устаревания  $\alpha = 0,75, 0,5$  и  $0,25$ .

Параметр устаревания весьма эффективно справился с недостатком предыдущего примера. Помимо траектории  $s = 18$ , для которой этот вариант алгоритма обеспечил понижение категории из-за ухудшения результатов на последних этапах обучения, алгоритм обнаружил аналогичную ситуацию и для траектории  $s = 13$ , к которой ранее в такой связи внимания не было. Относительно значений параметра представляется, что  $\alpha = 0,25$  не имеет смысла, так как классифицируется фактически последняя отметка. Значения  $\alpha = 0,5$  и  $0,75$  работают хорошо; на данном наборе примеров лучше выглядит  $\alpha = 0,75$ , поскольку в целом демонстрирует более «гладкий» результат.

## 5 Заключение

Полученные результаты можно подытожить рекомендацией использовать комбинацию второго и третьего вариантов алгоритма, полагая  $\alpha = 0,75$ , а возможность использования матрицы нормировки первого варианта оставить на усмотрение преподавателя. Такой расчет с имеющимися данными иллюстрируется в табл. 6, и его результаты выглядят наиболее предпочтительными для рекомендации к практическому применению.

## Литература

1. Rasch G. Probabilistic models for some intelligence and attainment tests. — Chicago, IL, USA: The University of Chicago Press, 1980. 199 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. No. 3. P. 195–210.
3. Кибзун А. И., Иноземцев А. О. Оценка уровней сложности тестов на основе метода максимального правдоподобия // Автоматика и телемеханика, 2014. № 4. С. 20–37.
4. Куравский Л. С., Мармалюк П. А., Юрьев Г. А., Думин П. Н., Панфилова А. С. Вероятностное моделирование процесса выполнения тестовых заданий на

основе модифицированной функции Раша // Вопросы психологии, 2015. № 4. С. 109–118.

5. Наумов А. В., Мхитарян Г. А. О задаче вероятностной оптимизации для ограниченного по времени тестирования // Автоматика и телемеханика, 2016. № 9. С. 124–135.
6. 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. No. 48. P. 2369–2380.
7. Callan R. The essence of neural networks. — Prentice Hall Europe, 1999. 232 p.
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.
9. Kohonen T. Self-organizing maps. — 3rd ed. — Berlin, Heidelberg: Springer, 2001. 501 p.
10. Deboeck G., Kohonen T. Visual explorations in finance. — London: Springer-Verlag, 1998. 258 p.
11. Ахтеров А. В., Лезина О. В., Шастина А. Е. Диагностика развития организационно-управленческих компетенций инженеров с помощью самоорганизующихся карт Кохонена // Автоматизация и управление в технических системах, 2013. № 42. С. 35–45.
12. Зарубина Н. К., Овчинкин О. В., Пыхтин А. И. Разведочный анализ результатов приема в вуз с применением нейронной сети Кохонена для планирования контингента студентов // Информационно-измерительные и управляющие системы, 2016. Т. 14. № 6. С. 65–69.
13. Сябренько А. П., Тынченко В. С., Бочарова О. А., Орешенко Т. Г. Анализ результатов аттестации сотрудников предприятий с применением карт Кохонена // Научно-технический вестник Поволжья, 2018. № 6. С. 159–162.
14. Тынченко В. С., Тынченко В. В., Букhtoyarov V. V., Kukartsev V. V., Kukartsev V. A., Ereemeev D. V. Application of Kohonen self-organizing maps to the analysis of enterprises' employees certification results // IOP Conf. Ser. — Mat. Sci., 2019. Vol. 537. Iss. 4. Art. 4042010. 5 p.
15. Наумов А. В., Джумурат А. С., Иноземцев А. О. Система дистанционного обучения математическим дисциплинам CLASS.NET // Вестник компьютерных и информационных технологий, 2014. № 10. С. 36–44.

16. Kohonen T. Self-organized formation of topologically correct feature maps // Biol. Cybern., 1982. Vol. 43. P. 59–69.
17. Ritter H., Martinetz T., Schulten K. Neural computation and self-organizing maps: An introduction. — Reading, MA, USA: Addison-Wiley, 1992. 306 p.
18. Kohonen T. Exploration of very large database by self-organizing maps // Conference (International) on Neural Networks Proceedings. — Piscataway, NJ, USA: IEEE, 1997. Vol. 1. P. PL1–PL6. doi: 10.1109/ICNN.1997.611622.

Поступила в редакцию 20.04.22

## APPLICATION OF SELF-ORGANIZING NEURAL NETWORKS TO THE PROCESS OF FORMING AN INDIVIDUAL LEARNING PATH

A. V. Bosov

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

**Abstract:** The problem of dynamic classification of students within the framework of supporting the process of forming an individual trajectory of the user of an electronic learning system is considered. The training model is designed for a mixed form of educational activities with partial independent work and periodic control events in the form of distance tests, partial full-time work with offline tests, and offset tests. The purpose of the classification is to determine the category of the student based on the results of the next control event. The semantics of the categories suggests the possibility of an individual task of different levels of task complexity at the next step of learning. The direction of improving the existing methods of classification is the rejection of the accumulation and use of statistics from previous (other) groups of students. The absence of samples of correct classification justified the use of self-organizing neural networks. For the solution, Kohonen’s maps were used, the standard version of which is adapted to the existing learning model and to the task of taking into account the subjective evaluation policy of the teacher. Three variants of the self-learning algorithm are described. Experimental research was carried out, its results are illustrated.

**Keywords:** electronic learning tool; self-organizing neural network; Kohonen’s map; classification; individual learning path

**DOI:** 10.14357/19922264220302

### Acknowledgments

The research was prepared with the support of the Russian Science Foundation according to the research project No. 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: The University of Chicago Press. 199 p.
2. Van der Linden, W.J., D.J. Scrams, D.L. Schnipke, et al. 1999. Using response-time constraints to control for differential speededness in computerized adaptive testing. *Appl. Psych. Meas.* 23(3):195–210.
3. 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.
4. Kuravsky, L.S., P.A. Marmalyuk, G.A. Yuryev, P.N. Dumin, and A.S. Panfilova. 2015. Veroyatnostnoye modelirovaniye protsessa vypolneniya testovykh zadaniy na osnove modifitsirovannoy funktsii Rasha [Probabilistic modeling of the process of carrying out test tasks on the basis of the modified function Rush]. *Voprosy psihologii* [Problems of Psychology] 4:109–118.
5. Naumov, A.V., and G.A. Mkhitarian. 2016. On the problem of probabilistic optimization of time-limited testing. *Automat. Rem. Contr.* 77(9):1612–1621.
6. 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.
7. Callan, R. 1999. *The essence of neural networks*. Prentice Hall Europe. 232 p.
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.
9. Kohonen, T. 2001. *Self-organizing maps*. 3rd ed. Berlin, Heidelberg: Springer. 501 p.
  10. Deboeck, G., and T. Kohonen. 1998. *Visual explorations in finance*. London: Springer-Verlag. 258 p.
  11. Akhterov, A., O. Lezina, and A. Shastina. 2013. Diagnostika razvitiya organizatsionno-upravlencheskikh kompetentsiy inzhenerov s pomoshch'yu samoorganizuyushchikhsya kart Kokhonena [Diagnostics of development of organizational and managerial competencies of the engineers using the self-organizing Kohonen maps]. *Avtomatizatsiya i upravlenie v tekhnicheskikh sistemakh* [Automation and Control in Technical Systems] 42:35–45.
  12. Zarubina, N. K., O. V. Ovchinkin, and A. I. Pykhtin. 2016. Razvedochnyy analiz rezul'tatov priema v vuz s primeneniem neyronnoy seti Kokhonena dlya planirovaniya kontingenta studentov [Exploratory data analysis of foster campaigns' results using Kohonen's neural network when planning the number of students]. *Informatsionno-izmeritel'nye i upravlyayushchie sistemy* [Information-Measuring and Control Systems] 14(6):65–69.
  13. Syabrenko, A. P., V. S. Tynchenko, O. A. Bocharova, and T. G. Oreshenko. 2018. Analiz rezul'tatov attestatsii sotrudnikov predpriyatii s primeneniem kart Kokhonena [Analysis of the enterprises employees attestation results based on the Kohonen maps]. *Nauchno-tekhnicheskiiy vestnik Povolzh'ya* [Scientific and Technical Volga Region Bulletin] 6:159–162.
  14. Tynchenko, V. S., V. V. Tynchenko, V. V. Bukhtoyarov, V. V. Kukartsev, V. A. Kukartsev, and D. V. Ereemeev. 2019. Application of Kohonen self-organizing maps to the analysis of enterprises' employees certification results. *IOP Conf. Ser. — Mat. Sci.* 537(4):4042010. 5 p.
  15. Naumov, A. V., A. S. Dzhumurat, and A. O. Inozemtsev. 2014. Sistema distantsionnogo obucheniya matematicheskimi distsiplinami CLASS.NET [Distance learning system for mathematical disciplines CLASS.NET]. *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Herald of Computer and Information Technologies] 10:36–44.
  16. Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 43:59–69.
  17. Ritter, H., T. Martinetz, and K. Schulten. 1992. *Neural computation and self-organizing maps: An introduction*. Reading, MA: Addison-Wiley. 306 p.
  18. Kohonen, T. 1997. Exploration of very large database by self-organizing maps. *Conference (International) on Neural Networks Proceedings*. Piscataway, NJ: IEEE. 1:PL1–PL6. doi: 10.1109/ICNN.1997.611622.

Received April 20, 2022

## Contributor

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

# НЕКОТОРЫЕ РЕЗУЛЬТАТЫ АНАЛИЗА ПРОЦЕССА ИЗМЕНЕНИЯ ЦЕНЫ БИВАЛЮТНОЙ КОРЗИНЫ НА ОСНОВЕ МЕТОДОВ СТАТИСТИКИ СЛУЧАЙНЫХ ПРОЦЕССОВ

П. В. Шнурков<sup>1</sup>, М. А. Мигуля<sup>2</sup>

**Аннотация:** Работа посвящена исследованию процесса изменения цены так называемой бивалютной корзины на валютном рынке Российской Федерации при наличии интервенций, проводимых Центральным банком. Основная цель работы заключается в проверке соответствия наблюдаемого процесса стохастической марковской модели с дискретным временем и дискретным множеством состояний. Теоретическую основу исследования составляют методы статистики марковских случайных процессов. В результате установлены условия, при которых реальный процесс может быть достаточно адекватно описан указанной марковской моделью.

**Ключевые слова:** цепь Маркова с дискретным временем и дискретным множеством состояний; статистика марковских случайных процессов; стохастические модели эволюции процессов на финансовых рынках; стохастические модели интервенций; цена бивалютной корзины

**DOI:** 10.14357/19922264220303

## 1 Введение

В работах [1, 2] была решена так называемая стохастическая задача о настройке, связанная с управлением случайным процессом, которое осуществляется в моменты его выходов на границу заданного подмножества множества состояний.

В дальнейшем идея решения задачи о настройке была использована при анализе проблемы управления интервенциями, проводимыми Центральным банком РФ на валютном рынке [3]. Однако для применения теоретических результатов решения задачи о настройке необходимо изучить характер основного процесса на интервалах времени между управлениями. В частности, необходимо подтвердить, что этот процесс является марковским. Соответствующее исследование проведено в настоящей работе.

В качестве методов исследования были использованы методы математической статистики марковских случайных процессов [4, 5]. По своей сути эти методы представляют собой проверку различных статистических гипотез, связанных со свойствами марковости и однородности наблюдаемых процессов. При этом реальные статистические данные об изменении цены бивалютной корзины за соответствующие периоды времени приведены на сайте Центрального банка РФ [6].

Особая сложность наблюдаемых процессов заключается в том, что множество их возможных значений не является фиксированным, а меняется во времени. Это связано как с объективным характером самих процессов, так и с тем, что на их эволюцию оказывали влияние действия Центрального банка РФ, который непосредственно устанавливал границы так называемых валютных коридоров и проводил периодические внешние воздействия на цену бивалютной корзины в форме валютных интервенций. В связи с этим в исследовании были рассмотрены несколько способов формирования множества состояний предлагаемых математических моделей.

В ходе исследования проверено большое число статистических гипотез, связанных со свойствами марковости и однородности. Эти гипотезы проверялись при различных параметрах (число наблюдений, длительность периода наблюдений, уровень значимости критерия или вероятность ошибки при принятии гипотезы). В своей совокупности проведенные проверки гипотез дали возможность установить ряд объективных закономерностей, связанных с поведением реальных процессов.

Кроме того, были построены статистические оценки матриц вероятностей перехода для рассматриваемых моделей и сделаны некоторые выводы о характере их изменения во времени.

<sup>1</sup> Национальный исследовательский университет «Высшая школа экономики», pshnurkov@hse.ru

<sup>2</sup> Национальный исследовательский университет «Высшая школа экономики», maxim.migulya@gmail.com



## 2 Дискретизация множества состояний процесса и общий подход к формированию марковской стохастической модели

Начнем изложение с разработки методики создания стохастической модели с дискретным временем и дискретным множеством состояний. Отметим, что параметр времени изначально дискретен, поскольку наблюдаемые значения фиксируются ежедневно, когда проводятся биржевые операции. Наблюдаемые значения цены соответствующего финансового актива (цены бивалютной корзины) принадлежат некоторому подмножеству множества действительных чисел. Будем сначала предполагать, что это подмножество задано. При этом предположении осуществим переход к стохастической модели с дискретным множеством состояний.

Обозначим через  $\{\tilde{\xi}_k\}_{k=0}^{\infty}$  наблюдаемую случайную последовательность, которая описывает эволюцию состояния системы во времени. Данный процесс будем считать исходным стохастическим объектом.

В рассматриваемой стохастической модели управляющая система, проводящая внешние воздействия или интервенции, устанавливает конкретные границы, в которых должны находиться состояния исходного процесса. Предположим, что множество допустимых значений этого процесса представляет собой конечный интервал во множестве неотрицательных вещественных чисел  $(x_0, x_1) \subset [0, \infty)$ . В дальнейшем будем называть это множество дискретизируемым. Таким образом, множество недопустимых значений данного процесса является объединением интервалов  $(0, x_0] \cup [x_1, \infty)$ .

Выберем достаточно большое целое положительное число  $N$  и обозначим

$$\Delta = \frac{x_1 - x_0}{N - 1}.$$

Разобьем интервал  $(x_0, x_1)$  на малые интервалы длины  $\Delta$ . Рассмотрим следующие множества вещественных чисел:

$$\begin{aligned} \tilde{X}_0 &= [0, x_0], \tilde{X}_1 = (x_0, x_0 + \Delta], \dots \\ \dots, \tilde{X}_s &= (x_0 + (s - 1)\Delta, x_0 + s\Delta], \\ & \quad s = 1, 2, \dots, N - 2; \end{aligned}$$

$$\begin{aligned} \tilde{X}_{N-1} &= (x_0 + (N - 2)\Delta, x_0 + (N - 1)\Delta = x_1], \\ \tilde{X}_N &= [x_1, \infty). \end{aligned}$$

Проведем преобразование введенных множеств по правилу

$$\begin{aligned} X_0 &= \tilde{X}_0, X_2 = \tilde{X}_1, X_{s+1} = \tilde{X}_s, \\ & \quad s = 1, 2, \dots, N - 1; \quad X_1 = \tilde{X}_N. \end{aligned}$$

Теперь определим новую случайную последовательность  $\{\xi_k\}_{k=0}^{\infty}$  при помощи соотношений:

$$\xi_k = \begin{cases} 0, & \text{если } \tilde{\xi}_k \in \tilde{X}_0; \\ s, & \text{если } \tilde{\xi}_k \in \tilde{X}_s, s = 2, 3, \dots, N; \\ 1, & \text{если } \xi_k \in X_1. \end{cases}$$

Из проведенного построения следует, что случайная последовательность  $\{\xi_k\}_{k=0}^{\infty}$  принимает значения в конечном множестве  $\{0, 2, \dots, N, 1\}$ , в котором состояния  $\{2, 3, \dots, N\}$  являются внутренними и допустимыми, а состояния  $\{0\}$  и  $\{1\}$  — граничными и недопустимыми. Предположим, что проблема управления этой случайной последовательностью рассматривается как стохастическая задача о настройке [1]. Для применения теоретических результатов решения задачи о настройке с дискретным временем [1, 2] необходимо при помощи статистических методов убедиться, что данная последовательность образует однородную цепь Маркова, в которой состояния  $\{0\}$  и  $\{1\}$  являются поглощающими.

Предположим, что в распоряжении исследователя имеется достаточно большое число наблюдений за реализациями случайной последовательности на интервалах времени между последовательными интервенциями. Зафиксируем один из таких интервалов. Обозначим через  $v_{ij}$  число переходов процесса за один шаг на указанном интервале из подмножества состояний  $\tilde{X}_i$  в подмножество  $\tilde{X}_j$ . Обозначим также через  $v_i$  общее число переходов за один шаг из состояния  $i$  во все состояния  $\tilde{X} = \cup_{i=0}^N \tilde{X}_i = [0, \infty)$ . Рассмотрим следующие статистические характеристики, определяемые по наблюдениям:

$$\hat{p}_{ij} = \frac{v_{ij}}{v_i}, \quad i \in \{2, 3, \dots, N\}, j \in \{0, 1, \dots, N\}.$$

При этом выполняется условие нормировки  $\sum_{j=0}^N \hat{p}_{ij} = 1, i \in \{2, 3, \dots, N\}$ . Полагая дополнительно

$$\begin{aligned} \hat{p}_{00} &= 1, \hat{p}_{0j} = 0, j \in \{1, 2, \dots, N\}; \\ \hat{p}_{11} &= 1, \hat{p}_{1j} = 0, j \in \{1, 2, \dots, N\}, \end{aligned}$$

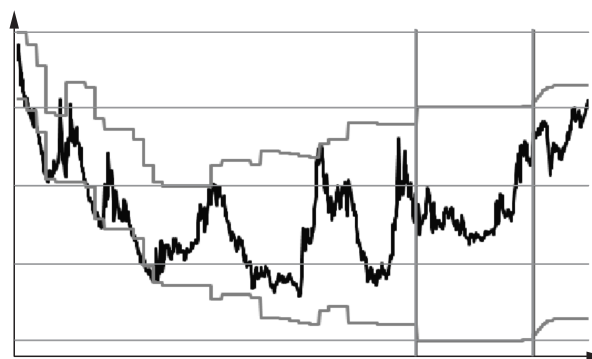
получим стохастическую матрицу  $\mathbf{P} = (\hat{p}_{ij})$ . Заметим, что структура этой матрицы соответствует структуре матрицы вероятностей перехода поглощающей цепи Маркова. Известно [4, 5], что

элементы этой матрицы представляют собой несмещенные и состоятельные оценки вероятностей перехода введенной случайной последовательности  $\{\xi_k\}_{k=0}^{\infty}$ .

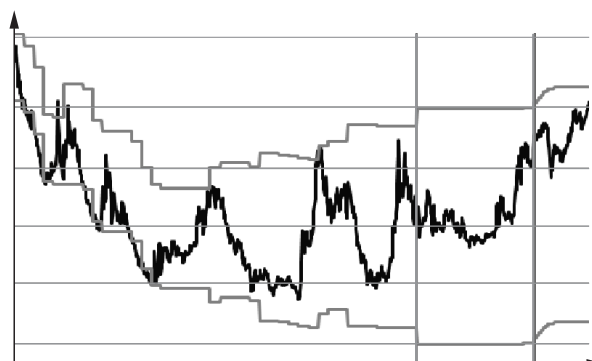
Заметим, что изложенная выше процедура дискретизации может быть с незначительными изменениями использована при переходе от исходного наблюдаемого процесса к модели общей цепи Маркова с дискретным (в частности, конечным) множеством состояний. При этом в качестве состояний новой цепи могут рассматриваться не только номера соответствующих подмножеств исходного дискретизируемого множества состояний, но и произвольные элементы, принадлежащие этим подмножествам.

Описанная выше процедура дискретизации связана с разбиением исходного заданного множества возможных значений наблюдаемого процесса. Однако, как уже отмечалось во введении, это множество может изменяться по ходу эволюции процесса. В связи с этим в данной работе предлагаются три варианта определения конечного множества состояний, которые могут рассматриваться как множество состояний модели после дискретизации, т.е. множество состояний марковского процесса.

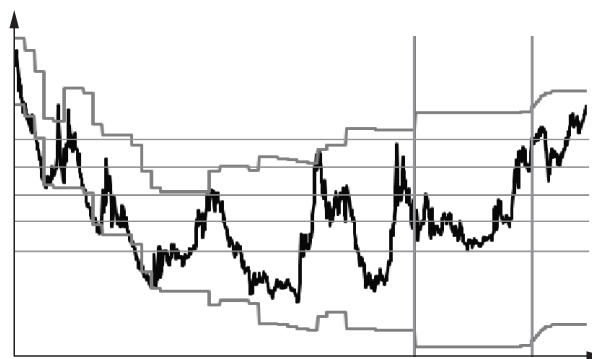
**Первый вариант.** В качестве исходного (дискретизируемого) множества состояний наблюдаемого процесса рассматривается некоторое достаточно «широкое» множество действительных чисел, которое включает в себя все возможные валютные коридоры, т.е. все возможные значения, которые может принимать этот наблюдаемый процесс. Интервал времени, на котором наблюдаются значения процесса и проверяются необходимые статистические гипотезы, может быть произвольным, однако при этом он не включает в себя интервалы, на которых процесс выходит за границы валютного коридора. В то же время сами границы валютного коридора на данном временном интервале могут изменяться. Тогда очевидно, что на каждом интервале времени, на котором установлены конкретные границы валютного коридора, наблюдаемый процесс будет принимать только значения из соответствующего подмножества состояний. Дискретные значения состояний, не соответствующие установленному валютному коридору, в определенной таким образом марковской модели будут недостижимы. Построенные оценки матриц вероятностей перехода за один шаг процесса будут отличаться друг от друга на тех интервалах времени, на которых установлены различные границы валютного коридора. Иллюстрация данного варианта дискретизации представлена на рис. 1, а.



(а)



(б)



(в)

**Рис. 1** Первый (а), второй (б) и третий (в) подходы к дискретизации

Несмотря на очевидные недостатки такого подхода к дискретизации, авторы считают целесообразным использовать его в данном исследовании, поскольку он позволяет сделать некоторые общие выводы о характере наблюдаемого процесса.

**Второй вариант.** Траектории наблюдаемого процесса рассматриваются только на интервалах времени, на которых границы валютного коридора остаются постоянными. Именно на этих интервалах времени проверяются статистические гипотезы, необходимые для того, чтобы установить наличие соответствующих свойств процесса. Исходное

(дискретизируемое) множество состояний наблюдаемого процесса совпадает с интервалом, граничными точками которого служат границы валютного коридора. При этом число элементов множества состояний после дискретизации постоянно для всех рассматриваемых интервалов времени, т. е. не зависит от конкретного валютного коридора. Иллюстрация данного способа дискретизации представлена на рис. 1, б.

**Третий вариант.** Как и в предыдущем варианте, траектории наблюдаемого процесса рассматриваются только на интервалах времени, на которых границы валютного коридора остаются постоянными. Необходимые статистические гипотезы проверяются на этих интервалах времени. Однако, в отличие от второго варианта, исходное (дискретизируемое) множество состояний наблюдаемого процесса совпадает с множеством значений, которые реально принимает наблюдаемый процесс. Это множество принадлежит валютному коридору. Естественно, как и для варианта 2, число элементов множества состояний после дискретизации постоянно для всех рассматриваемых интервалов времени, т. е. для всех соответствующих этим интервалам валютных коридоров. Иллюстрация данного варианта дискретизации представлена на рис. 1, в.

### 3 Общая схема статистического исследования свойств наблюдаемого процесса

Приведем описание общей структуры данного исследования, включая используемые статистические методы.

1. Проводится дискретизация стохастической модели, т. е. переход к модели с дискретным (конечным) множеством состояний. Процедура дискретизации подробно изложена в разд. 2.
2. Для каждого из трех вариантов дискретизации проводятся сбор и необходимая обработка имеющихся статистических данных. После этого по известному статистическому методу проверяется наличие общего марковского свойства у наблюдаемого процесса. Иными словами, устанавливается, обладает ли наблюдаемый процесс свойством марковости некоторого порядка  $s$ , где  $s \geq 1$  — некоторое целое число.
3. После подтверждения общего марковского свойства для каждого из трех вариантов дискретизации проверяется гипотеза о том, что наблюдаемый процесс обладает марковским

свойством первого порядка, т. е. образует классическую цепь Маркова с дискретным временем и дискретным (конечным) множеством состояний.

4. После подтверждения марковского свойства первого порядка для каждого из трех вариантов дискретизации проверяется гипотеза о том, что наблюдаемый процесс обладает свойством однородности, т. е. образует однородную цепь Маркова с дискретным временем и конечным множеством состояний.

Уточним, что перечисленные выше свойства проверяются последовательно. Каждое последующее свойство проверяется при тех же условиях, при которых было подтверждено предыдущее.

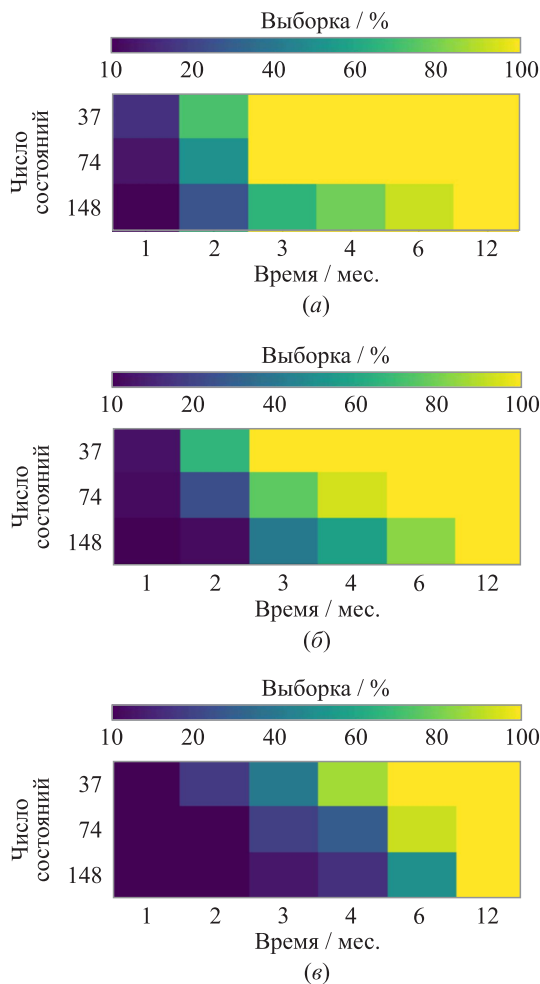
### 4 Проверка статистических гипотез о независимости и марковости некоторого фиксированного порядка

В данном разделе будут изложены результаты проверки общего марковского свойства наблюдаемой случайной последовательности.

Формально проверяется основная статистическая гипотеза, заключающаяся в том, что наблюдаемая случайная последовательность состоит из независимых случайных величин, против конкурирующей гипотезы, которая заключается в том, что значения наблюдаемой последовательности зависят и образуют общую цепь Маркова порядка  $s$ , где  $s \geq 1$  — некоторое целое число. Теоретические основы такой проверки, а также метод ее проведения изложены в работах [4, § 5.5; 5].

В рамках исследования по данной задаче различения гипотез были проведены более 10 000 статистических тестов. В результате можно сделать несколько общих эвристических выводов.

1. При уменьшении числа дискретных разбиений, возрастает число выборок, обладающих марковским свойством.
2. При увеличении временного интервала возрастает число выборок, обладающих марковским свойством.
3. Для первого способа дискретизации достаточно выбрать промежуток в 3 мес. или более с разбиением на 74 состояния или менее для достижения результата в виде 100% выборок, обладающих марковскими свойствами.



**Рис. 2** Выводы о марковости процесса при первом (а), втором (б) и третьем (в) способах дискретизации

- Для второго способа дискретизации достаточно выбрать промежуток в 4 мес. или более с разбиением на 74 состояния или менее для достижения результата в виде 100% выборок, обладающих марковскими свойствами.
- Для третьего способа дискретизации достаточно выбрать промежуток в 6 мес. или более с разбиением на 74 состояния или менее для достижения результата в виде 100% выборок, обладающих марковскими свойствами.

Сформулированные выводы проиллюстрированы диаграммами на рис. 2. Заметим, что в этих диаграммах по горизонтали при помощи использования различных цветов и их насыщенности отображается доля выборок, в которых соответствующая проверяемая гипотеза принимается, а также длительность интервалов времени, на которых формируются проверяемые выборки. Аналогичная структура характерна и для диаграмм, приводимых на рис. 3 и 4.

## 5 Проверка статистических гипотез о классической марковской зависимости и марковской зависимости более высокого порядка

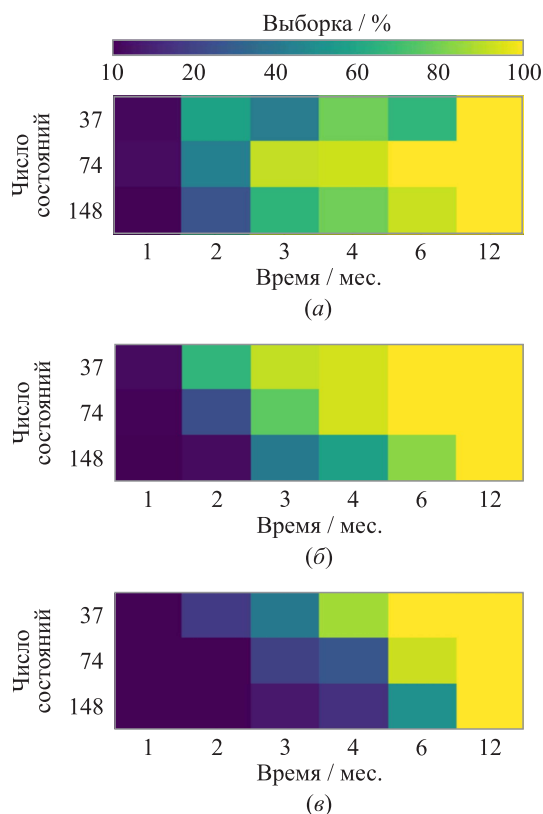
В данном разделе изложены результаты проверки наличия классического марковского свойства первого порядка у наблюдаемой случайной последовательности. Формально проверяется основная статистическая гипотеза, заключающаяся в том, что наблюдаемая случайная последовательность является цепью Маркова первого порядка, против конкурирующей гипотезы, которая заключается в том, что значения наблюдаемой последовательности образуют общую цепь Маркова порядка  $s$ , где  $s$  — некоторое целое число, большее единицы. Заметим, что в настоящем исследовании конкурирующая гипотеза была проверена только для варианта  $s = 2$ . Теоретические основы такой проверки, а также метод ее проведения изложены в работах [4, § 5.5; 5].

По данной проблеме были проведены более 5000 статистических тестов. На основе анализа их результатов можно сделать несколько эмпирических выводов.

- При увеличении числа дискретных разбиений возрастает число выборок с порядком, равным единице.
- При увеличении временного интервала возрастает число выборок, обладающих единичным порядком.
- Для первого способа дискретизации достаточно выбрать промежуток в 6 мес. или более с разбиением на 74 или более состояния для достижения результата в виде 100% выборок, обладающих единичным порядком.
- Для второго способа дискретизации достаточно выбрать промежуток в 4 мес. или более с разбиением на 74 или более состояния для достижения результата в виде 100% выборок, обладающих единичным порядком.
- Для третьего способа дискретизации достаточно выбрать промежуток в 2 мес. или более с разбиением на 74 или более состояния для достижения результата в виде 100% выборок, обладающих единичным порядком.

Сформулированные выводы проиллюстрированы диаграммами на рис. 3.



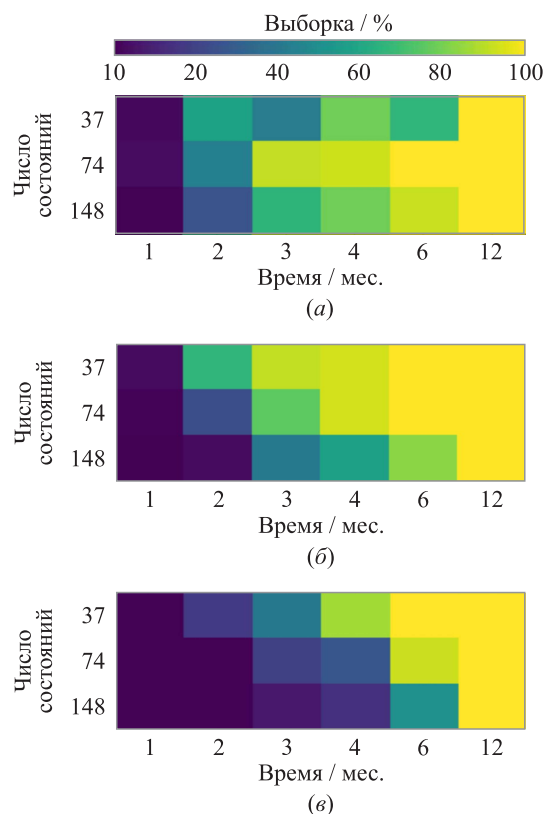


**Рис. 3** Выводы о единичном порядке марковской цепи при первом (а), втором (б) и третьем (в) способах дискретизации

## 6 Проверка статистических гипотез об однородности и неоднородности наблюдаемого процесса

Завершающим этапом проводимого исследования стала проверка свойства однородности. Поскольку предполагается, что предварительно уже было установлено марковское свойство первого порядка у наблюдаемой случайной последовательности, то проверка гипотезы об однородности сводится к следующему. Необходимо установить, что случайные последовательности, образующие цепи Маркова на разных интервалах времени, имеют одинаковые распределения вероятностей переходов. Такая задача также была решена в статистической теории марковских цепей [4, § 5.5, п. 5; 5]. Для проверки гипотезы однородности используется специальная статистика хи-квадрат с соответствующим предельным распределением.

В результате проведенных статистических испытаний установлено, что все временные выборки, обладающие свойством марковости первого поряд-



**Рис. 4** Выводы об однородности марковской цепи при первом (а), втором (б) и третьем (в) способах дискретизации

ка, в 100% случаев обладают также и свойством однородности. Результаты исследования проиллюстрированы диаграммами, представленными на рис. 4.

## 7 Общие выводы о вероятностных свойствах процесса изменения цены бивалютной корзины

Теперь можно сформулировать выводы, связанные с исследованием в целом, которые обобщают установленные опытным путем статистические закономерности.

1. При уменьшении числа элементов дискретных разбиений возрастает число выборок, обладающих марковским свойством, порядком, равным единице, и свойством однородности. Заметим, что это свойство не является устойчивым для небольших временных интервалов и соответствующих малых выборок.

Данную закономерность можно объяснить усложнением математической модели при уве-

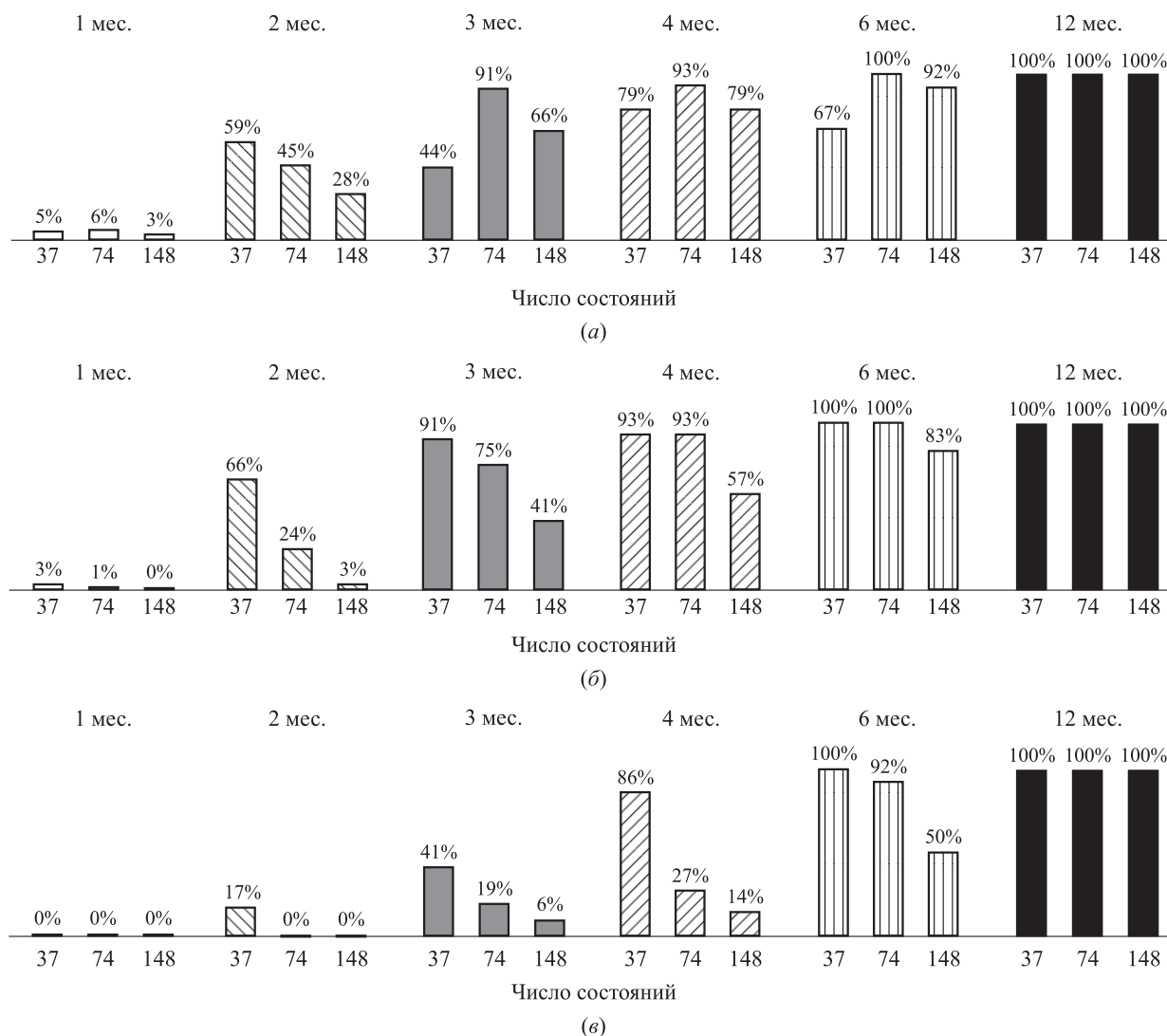


Рис. 5 Общие эмпирические выводы исследования при первом (а), втором (б) и третьем (в) способах дискретизации

личении числа элементов разбиений и приближении множества состояний к исходному (непрерывному).

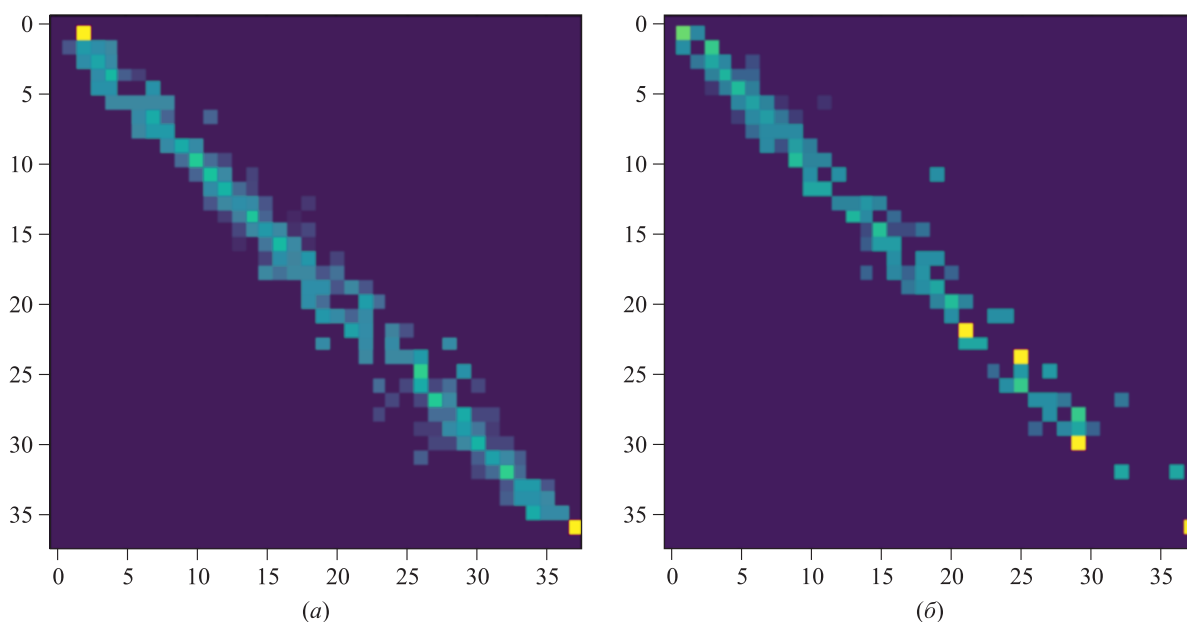
- При увеличении временного интервала возрастает число выборок, обладающих всеми тремя свойствами.

Данную закономерность можно объяснить зависимостью результатов от объема данных. На сравнительно малых интервалах времени при небольших объемах наблюдений выводы о характере процесса неустойчивы. При увеличении рассматриваемых интервалов времени и соответствующем росте объемов наблюдений используемые статистические методы позволяют обнаружить некоторые устойчивые закономерности, характерные для исследуемого процесса.

Сформулированные выше основные выводы проиллюстрированы диаграммами на рис. 5.

## 8 Построение статистических оценок матриц вероятностей перехода

Если в результате проведенного исследования установлено, что можно принять гипотезу о классическом марковском характере наблюдаемого стохастического процесса (случайный процесс описывается классической цепью Маркова первого порядка), а также гипотезу об однородности процесса, то целесообразно построить статистическую оценку матрицы вероятностей перехода данного процесса, которая, как известно, служит основной вероятностной характеристикой цепи Маркова



**Рис. 6** Матрицы оценок вероятностей перехода для марковской цепи с 37 состояниями, построенные на различных временных интервалах: (а) 01.01.2012–01.12.2013; (б) 01.11.2010–25.09.2011

и позволяет полностью описывать его эволюцию вероятностными методами. Для оценок элементов этой матрицы используются относительные частоты переходов процесса за одну единицу времени (или за один шаг процесса) из фиксированного состояния  $i$  во все остальные состояния, которые может принимать процесс.

Методика построения таких наборов относительных частот описана в разд. 2. Там же приведена классическая формула для оценки вероятности перехода.

Такую матрицу оценок можно построить на любом интервале времени, на котором приняты гипотезы о марковости и однородности наблюдаемого процесса. Это создаст возможность сравнить оценки матриц вероятностей перехода на разных интервалах времени и охарактеризовать динамику изменения классической марковской модели.

При построении оценок естественно использовать второй и третий подходы к дискретизации, при которых множество состояний построенной дискретной марковской цепи не изменяется. Еще раз подчеркнем, что при принятых способах построения оценок границы валютных коридоров на рассматриваемых интервалах времени могут быть различными.

Так как матрицы вероятностей перехода имеют большие размерности, используем их визуализацию в виде рисунков, при создании которых принято следующее соглашение: чем ярче цвет, тем больше вероятность перехода.

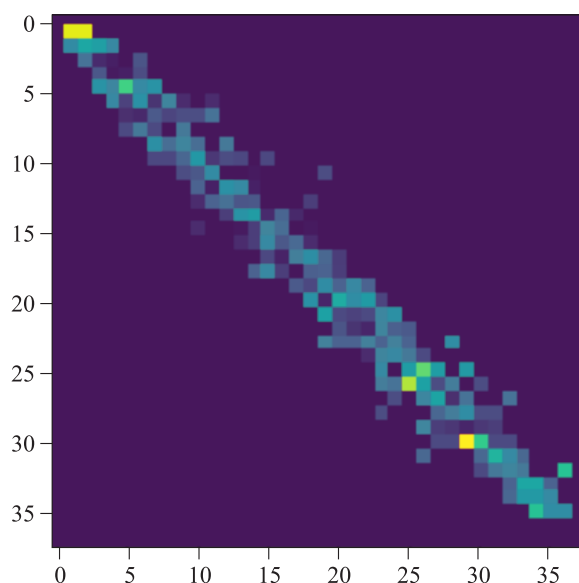
Для краткости ограничимся рассмотрением только второго подхода к дискретизации. При этом будем использовать данные наблюдений из следующих временных интервалов: 01.01.2012–01.12.2013 и 01.11.2010–25.09.2011.

Вначале рассмотрим первый временной интервал. Матрица оценок вероятностей перехода для дискретной марковской цепи, имеющей 37 состояний, визуально представлена на рис. 6, а.

Далее перейдем ко второму временному интервалу. Матрица оценок вероятностей перехода для дискретной марковской цепи, также имеющей 37 состояний, визуально представлена на рис. 6, б.

Для более наглядного сравнения построим матрицу, определяющую разность по абсолютной величине между всеми элементами матриц вероятностей перехода на двух рассмотренных временных интервалах. Соответствующая матрица визуально представлена на рис. 7.

Качественные выводы из сравнения матриц оценок вероятностей перехода заключаются в следующем. В приведенном примере число элементов этих матриц, которые сильно отличаются друг от друга (выражаемое числом ярких светлых точек на визуализации матрицы разностей, см. рис. 7), сравнительно невелико. В то же время проблема изменения матрицы оценок вероятностей перехода марковской цепи для разных интервалов времени, на которых построены соответствующие оценки, нуждается в дальнейшем, более глубоком исследовании.



**Рис. 7** Матрица, характеризующая разности элементов матриц оценок вероятностей перехода на разных временных интервалах

довании. Матрицы вероятностей перехода можно использовать для решения многих задач, связанных с исследованием марковских случайных процессов, и, в частности, для исследования проблем прогнозирования. В связи с этим исследование влияния изменчивости матрицы оценок вероятностей перехода на вероятностные характеристики созданной марковской модели представляет собой важную и актуальную проблему.

## 9 Заключение

Настоящую работу необходимо рассматривать как первую часть большого научного исследования реальных стохастических процессов, происходящих на валютном рынке РФ. В частности, при

продолжении исследований можно дополнительно использовать результаты и методы классической математической статистики, не связанные непосредственно со статистическими методами, разработанными для анализа марковских процессов. Комплексное использование различных методов математической статистики может оказаться весьма плодотворным.

В то же время используемые в работе статистические методы можно применять для анализа не только экономических процессов с интервенциями, но и многих других стохастических процессов, происходящих на товарных, фондовых и валютных рынках.

## Литература

1. *Shnurkov P. V.* Optimal control problem in a stochastic model with periodic hits on the boundary of a given subset of the state set (tuning problem). — Cornell University, 2017. arXiv: 1709.03442 [math.OC]. 16 p. <https://arxiv.org/abs/1709.03442v1>.
2. *Shnurkov P. V., Novikov D. A.* Analysis of the problem of intervention control in the economy on the basis of solving the problem of tuning. — Cornell University, 2018. arXiv: 1811.10993 [q-fin.GN]. 15 p. <https://arxiv.org/abs/1811.10993>.
3. *Шнурков П. В., Новиков Д. А.* О концепции стохастической модели с управлением в моменты выхода процесса на границу заданного подмножества множества состояний // Информатика и её применение, 2020. Т. 14. Вып. 3. С. 104–111.
4. *Ивченко Г. И., Медведев Ю. И.* Введение в математическую статистику. — М.: ЛКИ, 2010. 600 с.
5. *Ивченко Г. И., Медведев Ю. И.* Дискретные распределения. Вероятностно-статистический справочник: Многомерные распределения. — М.: ЛЕНАНД, 2016. 336 с.
6. Стоимость бивалютной корзины // Банк России. <https://www.cbr.ru/archive/db/bicurbase>.

Поступила в редакцию 13.07.22

---



---

# SOME RESULTS OF THE ANALYSIS OF THE PROCESS OF CHANGING THE PRICE OF A DUAL CURRENCY BASKET BASED ON RANDOM PROCESS STATISTICS METHODS

P. V. Shnurkov and M. A. Migulya

National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation

**Abstract:** The work is devoted to the study of the process of changing the price of the so-called dual-currency basket in the foreign exchange market of the Russian Federation in the presence of interventions carried out by the Central Bank. The main goal of the work is to verify the compliance of the observed process with a stochastic



Markov model with discrete time and a discrete set of states. The theoretical basis of the study is the methods of statistics of Markov random processes. As a result, conditions are established under which the real process can be adequately described by the indicated Markov model.

**Keywords:** Markov chain with discrete time and discrete set of states; statistics of Markov random processes; stochastic models of the evolution of processes in financial markets; stochastic models of interventions; price of a dual-currency basket

**DOI:** 10.14357/19922264220303

## References

1. Shnurkov, P. V. 2017. Optimal control problem in a stochastic model with periodic hits on the boundary of a given subset of the state set (tuning problem). Cornell University. arXiv.org. 16 p. Available at: <https://arxiv.org/abs/1709.03442v1> (accessed July 24, 2022).
2. Shnurkov, P. V., and D. A. Novikov. 2018. Analysis of the problem of intervention control in the economy on the basis of solving the problem of tuning. Cornell University. arXiv.org. 15 p. Available at: <https://arxiv.org/abs/1811.10993> (accessed July 24, 2022).
3. Shnurkov, P. V., Novikov D. A. 2020. O kontseptsii stokhasticheskoy modeli s upravlniem v momenty vykhoda protsessa na granitsu zadannogo podmnozhestva mnozhestva sostoyaniy [On the concept of a stochastic model with control at the moments of the process at the border of a presented subset of multiple states]. *Informatika i ee Primeneniya — Inform. Appl.* 14(3):104–111.
4. Ivchenko, G. I., and Y. I. Medvedev. 2010. *Vvedenie v matematicheskuyu statistiku* [Introduction to mathematical statistics]. Moscow: LKI. 600 p.
5. Ivchenko, G. I., and Y. I. Medvedev. 2016. *Diskretnye raspredeleniya. Veroyatnostno-statisticheskii spravochnik: Mnogomernye raspredeleniya* [Discrete distributions. Probabilistic-statistical handbook: Multivariate distributions]. Moscow: Lenand. 336 p.
6. Stoimost' bivalyutnoy korziny [Dual currency basket value]. Bank of Russia. Available at: <https://www.cbr.ru/archive/db/bicurbase/> (accessed July 24, 2022).

Received July 13, 2022

## Contributors

**Shnurkov Peter V.** (b. 1953) — Candidate of Science (PhD) in physics and mathematics, associate professor, National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation; [pshnurkov@hse.ru](mailto:pshnurkov@hse.ru)

**Migulya Maxim A.** (b. 2001) — bachelor, National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation; [maxim.migulya@gmail.com](mailto:maxim.migulya@gmail.com)

# ПОВЫШЕНИЕ ДОХОДНОСТИ ТОРГОВЛИ НА FOREX С ПОМОЩЬЮ LSTM-ИДЕНТИФИКАЦИИ СВЕЧНЫХ ПАТТЕРНОВ И ИНДИКАТОРА ТИКОВЫХ ОБЪЕМОВ\*

А. К. Горшенин<sup>1</sup>, Е. И. Гусейнова<sup>2</sup>

**Аннотация:** Статья посвящена исследованию эффективности использования рекуррентных нейронных сетей LSTM (Long–Short Term Memory — долгая–краткосрочная память) для свечных данных и индикатора технического анализа для большого числа наиболее распространенных валютных пар (всего — 27) за длительный период времени с целью построения автоматических торговых стратегий. Продемонстрировано, что средняя итоговая и годовая доходности за 8 лет при проведении модельных торгов составили 286% и 15,4% соответственно, что более чем в 50 раз превышает значения для классической торговой стратегии Buy & Hold за тот же временной период. Кроме того, в работе предложен новый индикатор технического анализа на основе тиковых объемов, который используется как самостоятельно в качестве альтернативной торговой стратегии (итоговая и годовая доходности LSTM-моделей превосходят ее в среднем в 7,2 и 2,3 раза), так и в качестве дополнительного признака для повышения доходности нейросетевой стратегии за счет использования ансамблирования. Установлено, что для 37% анализируемых валютных пар использование именно ансамбля нейронных сетей позволяет дополнительно повысить итоговую доходность в среднем на 17,2%.

**Ключевые слова:** LSTM; ансамблевое обучение; свечные данные; технический индикатор; FOREX; валютные пары

**DOI:** 10.14357/19922264220304

## 1 Введение

С развитием информационных технологий и их внедрением в работу бирж различных финансовых рынков у трейдеров появилась возможность использовать программы для осуществления торговых операций, построенные на основе их собственных методов и правил торговли.

Ведение торговли с помощью программ-роботов получило название автоматической торговой системы. Такие системы способствовали существенному повышению эффективности работы трейдеров. В отличие от механической торговой системы, автоматические торговые системы способны совершать торговые операции по купле и продаже финансовых активов без непосредственного участия человека за счет встроенного алгоритма, который отвечает за автоматическую выработку сигнала на открытие/закрытие позиции, а также доставку ордера (заявки) на торговую платформу брокера. В основе автоматических торговых систем чаще всего лежит стратегия на основе комбинации различных индикаторов и паттернов технического анализа.

В последние годы все большее влияние на инструменты подобной автоматизированной торговли, как и повсеместно в мировом финансовом секторе [1], стали оказывать алгоритмы на основе машинного обучения и нейронных сетей. В частности, одним из направлений применения таких подходов стало построение торговых стратегий с более высокой доходностью по сравнению с классическими инструментами. Зачастую для этого используется прогнозирование цен различных финансовых инструментов с помощью алгоритмов машинного обучения, причем как на основе исходных ценовых данных, так и расширением признакового пространства различными характеристиками, включая различные индикаторы технического анализа (см., например, статью [2]).

В литературе наиболее распространены исследования для фондового рынка, однако также рассматриваются криптовалюты [3] и высокочастотная торговля отдельными валютными парами, прежде всего евро–доллар [4].

Целью данной статьи ставится исследование эффективности применения рекуррентных нейросе-

\* Исследование выполнено при поддержке Российского научного фонда, проект 22-11-00212. Для обучения нейронных сетей использовалась инфраструктура Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

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

<sup>2</sup> Московский государственный университет имени М. В. Ломоносова, ei.guseynova@yandex.ru

тевых архитектур LSTM сразу для 27 разнообразных валютных пар. В качестве базового уровня доходности используются показатели простой тестовой стратегии Buy & Hold [5], также называемой позиционной торговлей, при которой инвестор приобретает некоторые активы для долгосрочного хранения в ожидании повышения цен на них в пределах не менее нескольких лет. Кроме того, как было упомянуто выше, весьма популярны стратегии, использующие индикаторы технического анализа, — и в статье предложен новый инструмент, ориентированный на тиковые объемы. Наконец, представляет интерес построение стратегий с одновременным использованием индикаторов и нейронных сетей, которое в данной работе реализуется с помощью ансамблевого подхода к обучению LSTM-архитектур.

Статья организована следующим образом. В разд. 2 приведен обзор известных решений на основе машинного обучения для различных финансовых инструментов. Раздел 3 содержит описание анализируемых данных и используемый подход для обработки свечных данных. В разд. 4 представлена динамика доходности для каждой из пар, получаемая в рамках использования LSTM-стратегии. Раздел 5 посвящен описанию нового технического индикатора и сравнению его доходности с нейросетевым подходом. В разд. 6 продемонстрирована возможность дополнительного повышения доходности для 10 валютных пар в случае использования ансамблевого обучения для свечных данных и этого индикатора. В заключительном разделе подводятся краткие итоги и обсуждаются возможные направления дальнейших исследований.

## 2 Машинное обучение в задачах финансового прогнозирования

Одним из популярных подходов к прогнозированию цен на фондовом рынке с использованием нейронных сетей стал графический анализ. Так, в статье [6] применяется сверточная архитектура для обработки исходных финансовых временных рядов — 15 различных технических индикаторов, преобразованных в двумерные изображения, каждое из которых затем помечается как Buy, Sell или Hold в зависимости от предполагаемых точек входа в рынок и выхода из него. Тестирование на данных промышленного индекса Доу Джонса и биржевых инвестиционных фондов показало высокую эффективность данного подхода, в частности по сравнению со стратегией Buy & Hold.

В статье [7] предложена торговая система, не ограничивающаяся известными техническими пат-

тернами. Она позволяет сравнивать текущее рыночное состояние с более ранними похожими паттернами с целью получения торговых сигналов. Для ее тестирования использовались 560 акций Нью-Йоркской фондовой биржи, при этом авторы не проводили автоматическую и динамическую оптимизацию параметров торговой стратегии — и тем не менее предложенная система приводила к 92,5% прибыльных сделок.

Весьма популярны методы на основе анализа свечных данных. В статье [8] предложена ансамблевая нейронная сеть, включающая и сверточную архитектуру CNN (Convolutional Neural Network), которая позволяет автоматически идентифицировать восемь типов свечных паттернов со средней точностью 90,7% в реальных данных, превосходя LSTM-сеть. Еще один пример анализа с использованием свечных данных для Тайваньской фондовой биржи и индекса фондового рынка Nikkei 225 Токійской фондовой биржи приведен в статье [9].

Другая ансамблевая архитектура, сочетающая CNN и LSTM-сети, предложена в статье [10]. Она использует последовательность исторических данных и опережающие индикаторы (опционы и фьючерсы) для извлечения дополнительных признаков с помощью CNN, а затем использует их в качестве входных данных для LSTM. Для тестирования авторы использовали характеристики десяти американских и тайваньских акций.

Эффективным оказалось сочетание алгоритмов машинного обучения и нейронных сетей даже достаточно простых архитектур. Так, в статье [5] при построении модели прогнозирования учитывались не только ценовые данные по торговым инструментам, но и данные объема торговли, а для получения оптимального набора параметров торговой стратегии использована комбинация регрессионного метода опорных векторов и многослойного перцептрона. Полученный алгоритм продемонстрировал высокие результаты на шести инструментах фондового рынка в период с 2001 по 2015 гг.

В статье [11] поиск оптимальных параметров технического индикатора осуществляется с помощью генетических алгоритмов. Затем его значения передаются в глубокий многослойный перцептрон для получения одного из трех торговых сигналов (покупка, продажа, удержание текущей позиции). Результаты тестирования на исторических данных цен акций промышленного индекса Доу Джонса в период с 2007 по 2016 гг. показывают, что оптимизация параметров технического индикатора повышает эффективность торговли акциями и дает сопоставимые или лучшие результаты по сравнению с Buy & Hold и другими торговыми стратегиями.

В работе [12] показано, что прогнозирование точки разворота цены акций может быть весьма эффективным при использовании только LSTM-сетей, для которых на основе комбинаций свечных индикаторов и технических индикаторов строятся наборы дополнительных признаков. Для 10 китайских и 10 американских акций результаты превзошли и метод опорных векторов, и многослойный перцептрон, и CNN-архитектуру.

В статье [13] предложен метод предварительной обработки исходных ценовых свечей и значений технических индикаторов для генерации торговых сигналов, используемых для обучения нейронной сети LSTM. Такая замена исходных ценовых данных позволила повысить точность предсказаний для пяти типов торговых стратегий, однако, как и в работе [7], оптимизация их параметров не проводилась.

В статье [2] на примере фондового рынка Китая с 2000 по 2017 г. получена точность прогноза более 60% для некоторых моделей с использованием анализа свечных графиков и получения признаков с помощью ансамблевых методов машинного обучения. Отмечено, что дополнительные технические индикаторы могут в разной степени повысить точность прогноза.

Работа с валютной парой евро–доллар рассмотрена в статье [4], в которой предлагаются инструменты прогнозирования краткосрочного тренда на валютном рынке FOREX с использованием глубокого обучения и алгоритмов обучения с подкреплением для высокочастотной торговли.

Таким образом, весьма актуально исследование сразу большого числа валютных пар за длительный период на основе свечных данных с использованием нейронных сетей и инструментов технического анализа, которому и посвящена данная статья.

### 3 Анализируемые данные и создание пространства признаков

Тиковые данные, содержащие метки времени и цены сделки, по наиболее торгуемым 27 валютным парам загружены с помощью программного интерфейса на языке Python с торговой платформы «Metatrader»<sup>1</sup>.

Начало доступного периода для каждой валютной пары приведено в таблице. В качестве конца периода для всех пар использована дата 15 апреля 2022 г., 23:54:00. Таким образом, для большинства валютных пар доступный интервал составляет около 10 лет.

Анализируемые валютные пары и даты начала периода для данных

Валютная пара	Начало периода
AUDCAD	2011-12-19 00:00:00
AUDCHF	2011-12-19 00:00:00
AUDJPY	2011-12-19 21:00:00
AUDNZD	2011-12-20 21:00:00
AUDUSD	2011-12-19 21:00:00
CADCHF	2011-12-21 21:00:00
CADJPY	2013-03-19 00:00:00
CHFJPY	2011-12-26 21:00:00
GBPAUD	2013-03-19 00:00:00
GBPCHF	2012-01-02 00:00:00
GBPJPY	2012-01-09 00:00:00
GBPNZD	2014-07-30 13:47:00
GBPUSD	2011-12-19 21:00:00
EURAUD	2012-01-13 21:00:00
EURCAD	2011-12-28 21:00:00
EURCHF	2011-12-29 21:00:00
EURGBP	2011-12-28 21:00:00
EURJPY	2011-12-29 21:00:00
EURNZD	2012-01-12 21:00:00
EURUSD	2011-12-19 21:00:00
NZDCAD	2012-01-17 20:58:00
NZDCHF	2012-01-17 20:58:00
NZDJPY	2012-01-17 20:58:00
NZDUSD	2011-12-26 21:00:00
USDCAD	2011-12-28 21:00:00
USDCHF	2011-12-26 21:00:00
USDCNH	2015-11-17 07:38:00

В связи со значительным объемом занимаемой памяти параллельно с выгрузкой данные конвертировались в датафрейм, в котором одно наблюдение соответствовало свече на минутном таймфрейме (торговом периоде — интервале времени, используемом для группировки котировок). Данные содержат информацию о цене открытия и закрытия свечи, ее максимуме и минимуме, а также число тиковых колебаний цен Ask (покупка) и Bid (продажа) (рис. 1).

Метод обработки свечей состоит в том, чтобы из количественных данных цен открытия, закрытия, максимума и минимума образовать одномерный массив категориальных объектов. Такое упрощение структуры данных сокращает объем выборки, ускоряя обучение, и позволяет нейросети находить свечные паттерны, выявляя последовательности в них, используемые в дальнейшем для формирования сигналов в рамках торговой стратегии.

В зависимости от длины тела свечи и ее хвостов были выделены по десять подклассов для интервалов 1 и 2 (рис. 2, а). Сочетание двух подклассов

<sup>1</sup><https://www.metaquotes.net>.

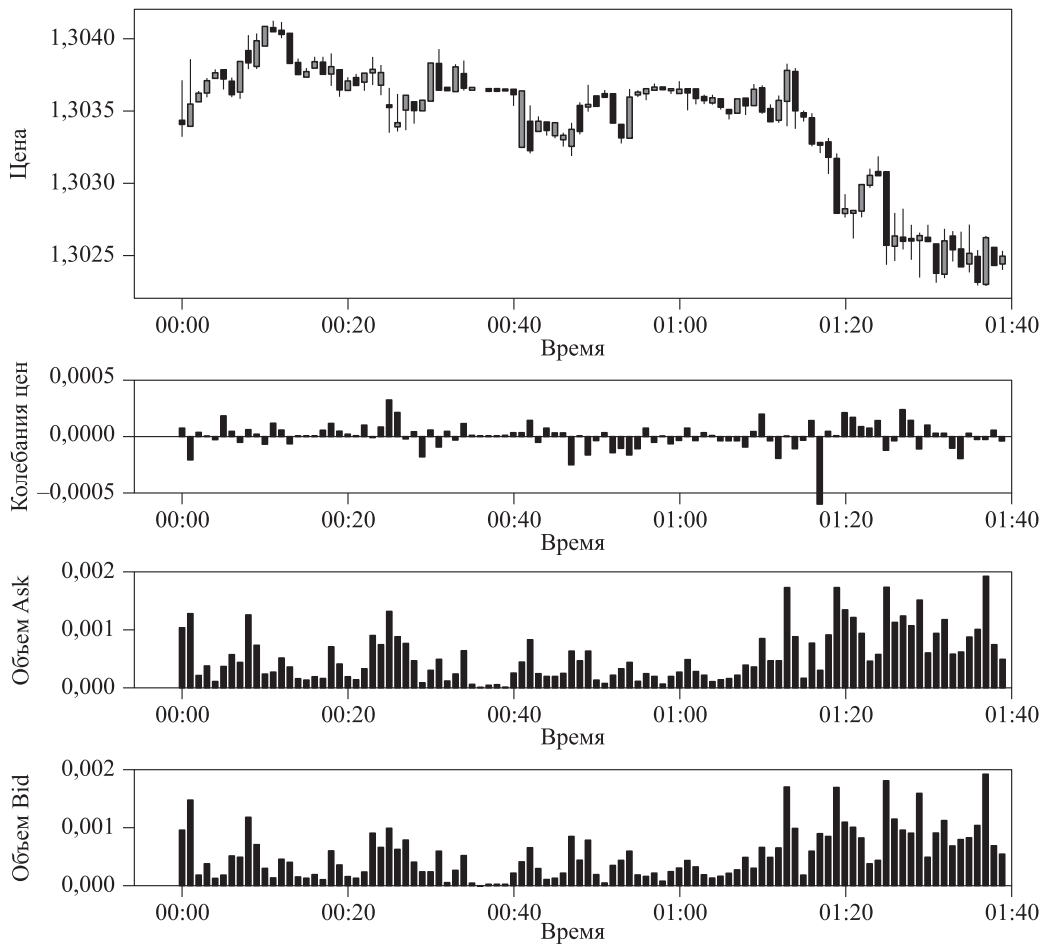


Рис. 1 Цены открытия, закрытия, минимума, максимума, тиковые объемы Ask и Bid

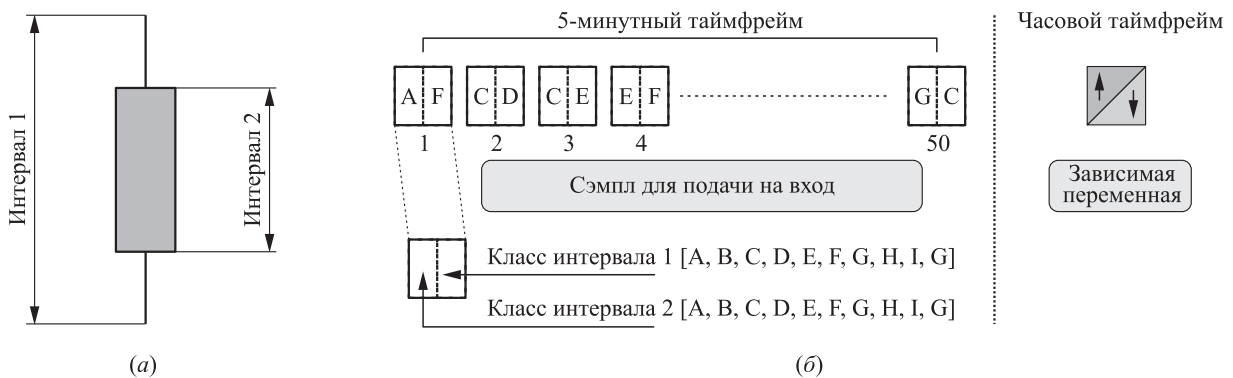


Рис. 2 Схема преобразования числовых данных в категориальные: (а) объекты для определения класса свечи; (б) преобразование в шаблоны из классов свечей и зависимой переменной направления рынка

определяло класс свечи. Все свечи на пятиминутном таймфрейме были переведены в категориальные объекты (рис. 2, б). Задача прогнозирования сводится к определению направления следующего категориального объекта в последовательности, или, другими словами, к определению семанти-

ки, т. е. к бинарной классификации направления следующей свечи. Для этого длинная цепочка категориальных данных была разбита на блоки, состоящие из 50 объектов. Каждому блоку соответствовало значение зависимой переменной, характеризующей семантику — направление рынка на



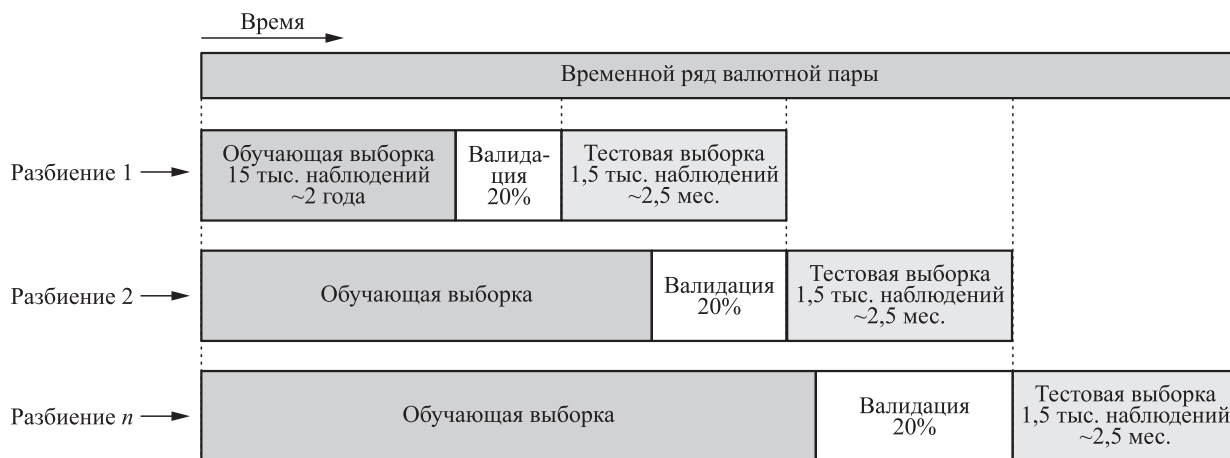


Рис. 3 Схема разбиения данных на обучающую, тестовую и валидационную выборки

следующей свече. Стоит отметить, что направление рынка определялось на более старшем таймфрейме, а именно на часовом, в то время как объекты блока — на пятиминутном.

Поскольку с ростом обучающей выборки сеть «знакомится» с большим числом паттернов, то предпочтительнее обучать сеть на как можно большем временном периоде. Одновременно с этим, чтобы можно было судить о состоятельности предложенного метода обучения, тестирование также должно проводиться на достаточно продолжительном временном отрезке. Учитывая эти два фактора, было решено предварительно обучить сеть на данных, соответствующих двум годам, а затем дообучать модель на новых данных с периодичностью в два с половиной месяца. Схема разбиения данных на обучающие, валидационные и тестовые выборки изображена на рис. 3.

Для каждой валютной пары было сделано около 30 разбиений: при первом разбиении из 15 тыс. наблюдений была создана валидационная и тестовая выборка, а из следующих 1500 наблюдений — тестовая; при втором — тестовая выборка предыдущего разбиения добавлялась к обучающей и валидационной, а следующие 1500 наблюдений составляли новую тестовую выборку и т. д.

#### 4 Торговая стратегия для свечных данных с использованием LSTM-сети

Суть предлагаемой стратегии заключается в следовании сигналам нейронной сети (фактически, прогнозам) о покупке/продаже на основе свечных паттернов. Удержание позиции происходит до генерации следующего сигнала через один модельный

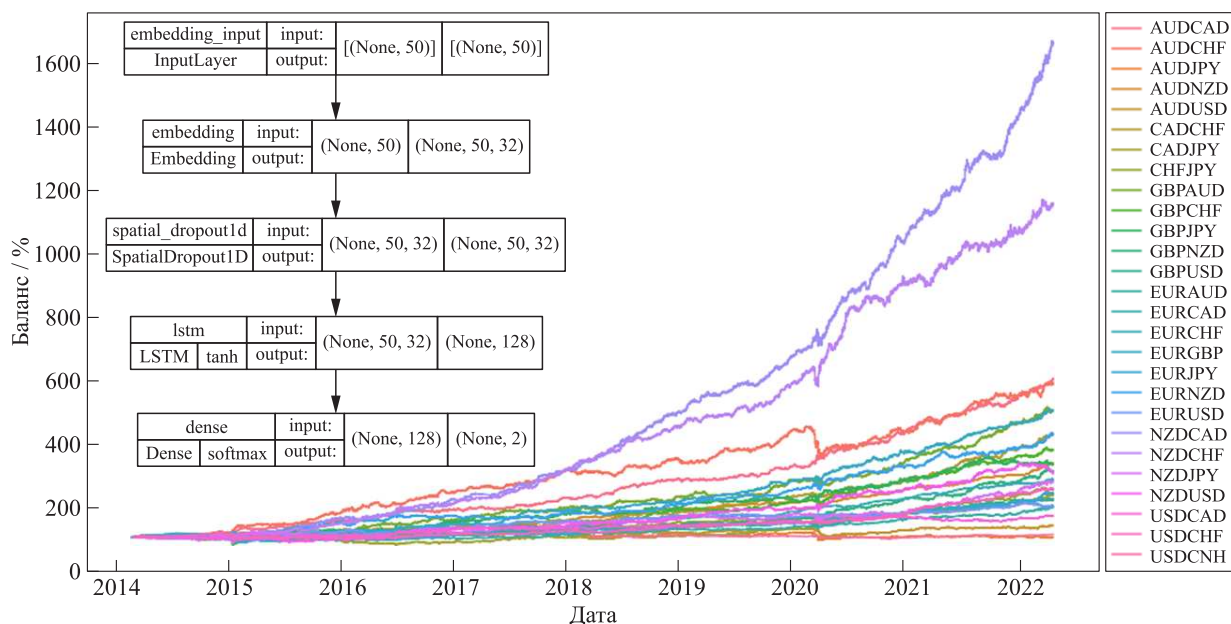
торговый час. В ходе исследования был проведен ряд экспериментов для определения оптимальной конфигурации сети, в результате которых была выбрана структура, схематично представленная на рис. 4 (слева внутри графика).

Слой embedding — векторное представление категориальных объектов в одном наблюдении. В его основе лежит векторизация каждой категории, всего 100, так как класс свечи определяется десятью вариантами интервала 1 и десятью вариантами интервала 2. Веса инициализируются случайным образом, а затем они корректируются с помощью алгоритма обратного распространения ошибки.

Слой dropout с коэффициентом 0,5 традиционно используется для предотвращения переобучения сети. LSTM-слой в архитектуре позволяет сохранять и передавать информацию от одного шага сети к другому, учитывая последовательность, в какой данные подаются, что является ключевым в распознавании паттернов. Слой dense — стандартный полносвязный.

В качестве активационной функции выходного слоя используется softmax — обобщение логистической функции для многомерного случая. В качестве оптимизатора используется Adam, в качестве функции потерь — бинарная кросс-энтропия, точность оценивается в смысле качества распознавания объектов.

При обучении экспериментальным путем были выбраны следующие значения основных параметров обучения: общее число тренировочных объектов, представленных в одном пакете, — 50; общее число эпох обучения — 300. При этом, если значение функции потерь не уменьшалось в течение 7 эпох, лучшая модель сохранялась и обучение прекращалось. Для программной реализации был выбран язык Python, в качестве основной платфор-



**Рис. 4** Динамика баланса в ходе тестирования LSTM-стратегии на тестовом периоде для 27 валютных пар и архитектура используемой нейронной сети (на графике слева внутри)

мы для обучения нейронных сетей — TensorFlow с программным интерфейсом Keras. Эффективность стратегий оценивалась доходностью соответствующих финансовых инструментов.

Результаты тестирования для 27 валютных пар представляют собой динамику изменения баланса в процентах (см. рис. 4), а также итоговую и годовую доходность в процентах, которая будет обсуждаться ниже. Период тестирования для 23 валютных пар составляет 8 лет, для четырех он сокращен до 6–7 лет в силу доступности меньшего объема данных. Визуализация динамики баланса на рис. 4 демонстрирует отсутствие существенных просадок ниже уровня 100% на протяжении всего периода, что свидетельствует о стабильности предложенной стратегии. Наибольшая доходность у валютной пары NZDCAD — 1570% (41,2% годовых), а наименьшая — у AUDJPY: —0,57% (—0,08% годовых). Средняя годовая доходность по всем валютным парам составляет 15,4%. Сравнение с другими стратегиями будет приведено в разд. 5 и 6.

## 5 Индикатор тиковых объемов

Рассмотренная в предыдущем разделе стратегия, основанная на нейронных сетях, не относится к традиционному техническому анализу, хотя в ее основе лежит распознавание последовательностей свечей. В данном разделе сравним результаты LSTM-стратегии и подхода на основе технического индикатора. Можно отметить, что само по себе

значение индикатора не является сигналом к покупке или продаже. Пороговые значения, после которых цену можно считать подходящей для совершения сделки, определяются трейдерами, а значит, сигналы одного и того же индикатора можно интерпретировать по-разному. При построении торговой стратегии помимо графических паттернов цены важную роль играют уровни поддержки и сопротивления. Многие трейдеры строят свои системы исключительно на принципах использования ценовых уровней, зон поддержки и сопротивления (в них сосредоточены заявки крупных участников рынка на покупку и продажу), не уделяя внимания паттернам в чистом виде [14].

Предлагаемый в данной статье индикатор связан с поиском и обнаружением ликвидности. Он построен на анализе поступающих на биржу заявок с целью выявления крупных ордеров (заявок) или агентов финансового рынка, которые выставляют большие объемы активов на покупку или продажу. Всеобщие данные объемов недоступны, но по характеру прошлых изменений цен Ask и Bid можно установить, какие ценовые уровни заставляли цену менять свое направление и хотя бы на время останавливали тренд. Индикатор создан для выявления повышенных уровней объема у свечей. Стоит обратить внимание, что если объем Ask и объем Bid одновременно велики, то это значит, что силы покупателей и продавцов не преобладают друг над другом, а значит, смена тренда маловероятна. По этой причине в индикаторе используются не

абсолютные значения ценовых объемов, а их разница — именно повышение разницы объемов Bid и Ask в данном индикаторе служит сигналом о смене тренда. Формула индикатора тиковых объемов Bid и Ask имеет следующий вид:

$$\text{fracInd}_t = \frac{\text{CountAsk}_t - \text{CountBid}_t}{\text{CountAsk}_t + \text{CountBid}_t} \times \frac{(1 - (|\text{Close}_t - \text{Open}_t|) / (\text{High}_t - \text{Low}_t))}{|\text{Close}_t - \text{Open}_t|},$$

где

$\text{fracInd}_t$  — значение индикатора в момент времени  $t$ ;

$\text{CountAsk}_t$  — объем тиков Ask в момент времени  $t$ ;

$\text{CountBid}_t$  — объем тиков Bid в момент времени  $t$ ;

$\text{Close}_t$  — цена закрытия свечи в момент времени  $t$ ;

$\text{Open}_t$  — цена открытия свечи в момент времени  $t$ ;

$\text{High}_t$  — максимальное значение цены в момент времени  $t$ ;  $\text{Low}_t$  — минимальное значение цены в момент времени  $t$ .

Чем больше разница тиков Ask и Bid, чем больше хвосты у свечи и чем меньше тело свечи, тем выше значение индикатора  $\text{fracInd}_t$ . Для генерации сигналов  $\text{fracInd}_t$  был нормализован в соответствии с формулой:

$$\text{stoch}_t = \frac{\text{fracInd}_t - \min(\text{fracInd}_n)}{\max(\text{fracInd}_k) - \min(\text{fracInd}_n)} \cdot 100,$$

где

$\text{stoch}_t$  — значение нормализованного индикатора в момент времени  $t$ ;

$\min(\text{fracInd}_n)$  — минимальное значение  $\text{fracInd}_t$  за  $n$  периодов;

$\max(\text{fracInd}_k)$  — максимальное значение  $\text{fracInd}_t$  за  $k$  периодов.

В зависимости от изменения параметров  $n$  и  $k$  интерпретация индикатора может меняться. Чем выше их значения, тем большее окно ценовых значений рассматривается для выявления зон с повышенными объемами ликвидности. Таким образом, для старших таймфреймов предпочтительнее использовать более высокие значения, чем для низких.

Для тестирования была создана простейшая торговая стратегия: при пиковых значениях индикатора  $\text{stoch}_t$  в случае восходящего тренда позиция менялась на короткую, а в случае нисходящего — на длинную. Индикатор использовался для определения ценовых разворотов и смены тренда. С помощью созданной функции бэктестирования описанная стратегия была проверена на всех валютных парах на протяжении пяти лет с 15 апреля 2017 г. по 15 апреля 2022 г. Тестирование на более длительных периодах без проведения дополнительной оптимизации неэффективно, так как торговля с индикаторами предусматривает периодическое обновление параметров стратегии.

На рис. 5 представлен фрагмент сигналов, вырабатываемых предложенной стратегией: черточки поверх свечей обозначают сделки: красные — убыточные, зеленые — прибыльные. Легко заметить, что рекомендуемые цены открытия/закрытия позиции совпадают с пиковыми значениями локальных трендов. Это означает, что сигналы индикатора можно использовать как точки входа и выхода из позиции. Разработанный индикатор хорошо предсказывает развороты рынка даже с учетом неизменных параметров на протяжении всего тестируемого

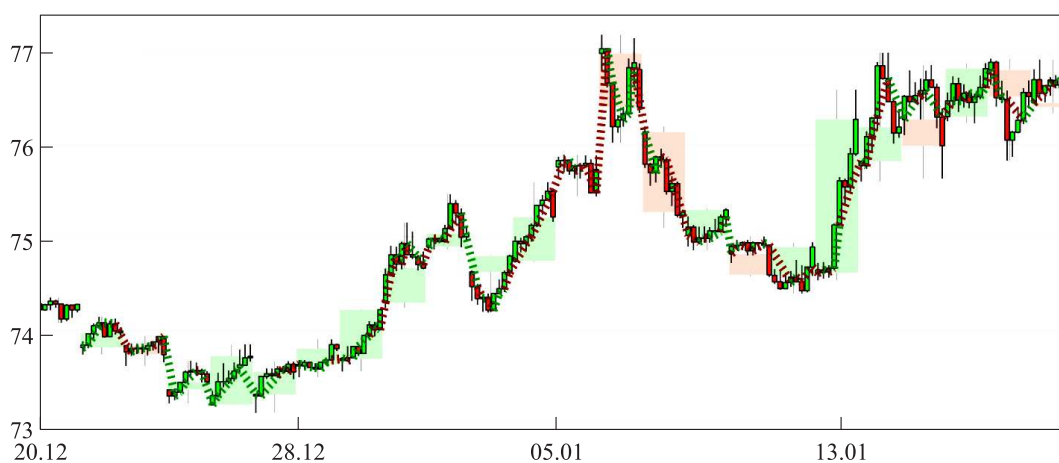


Рис. 5 Визуализация сигналов стратегии на основе индикатора тиковых объемов (данные 122 сделок)

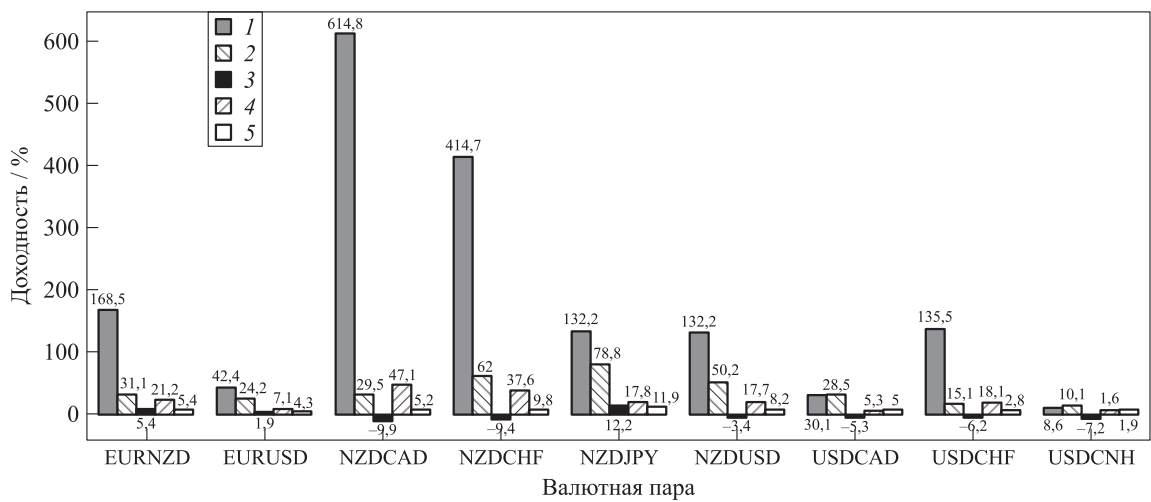
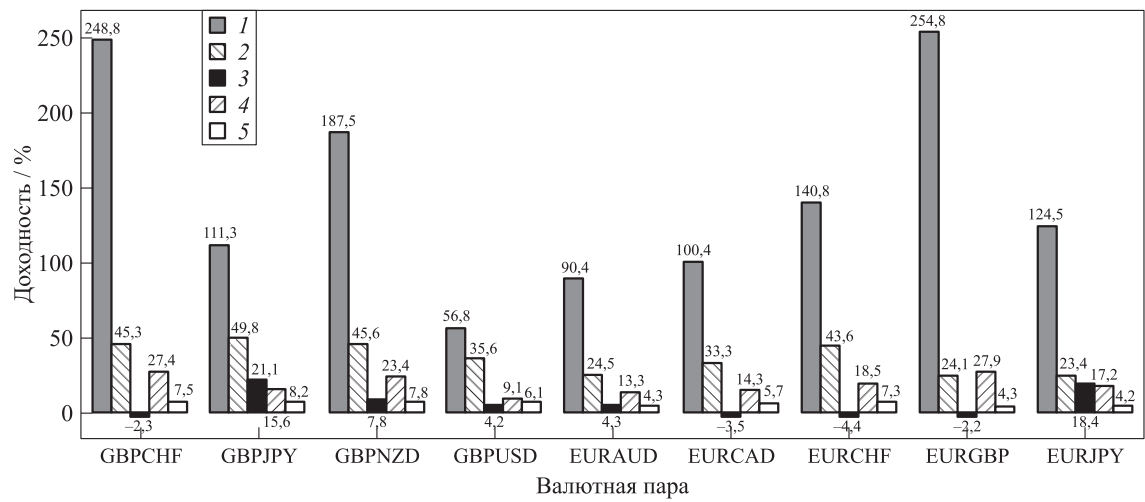
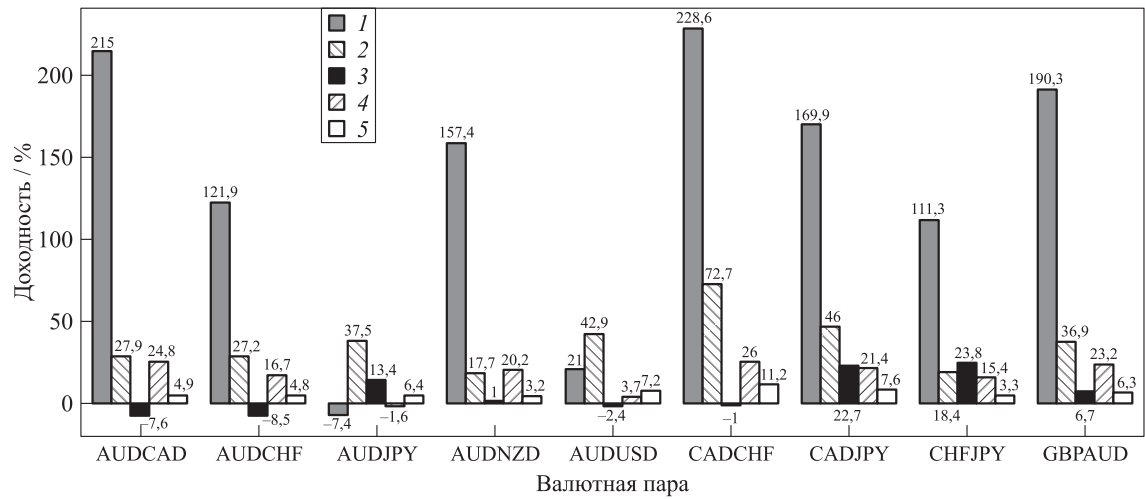
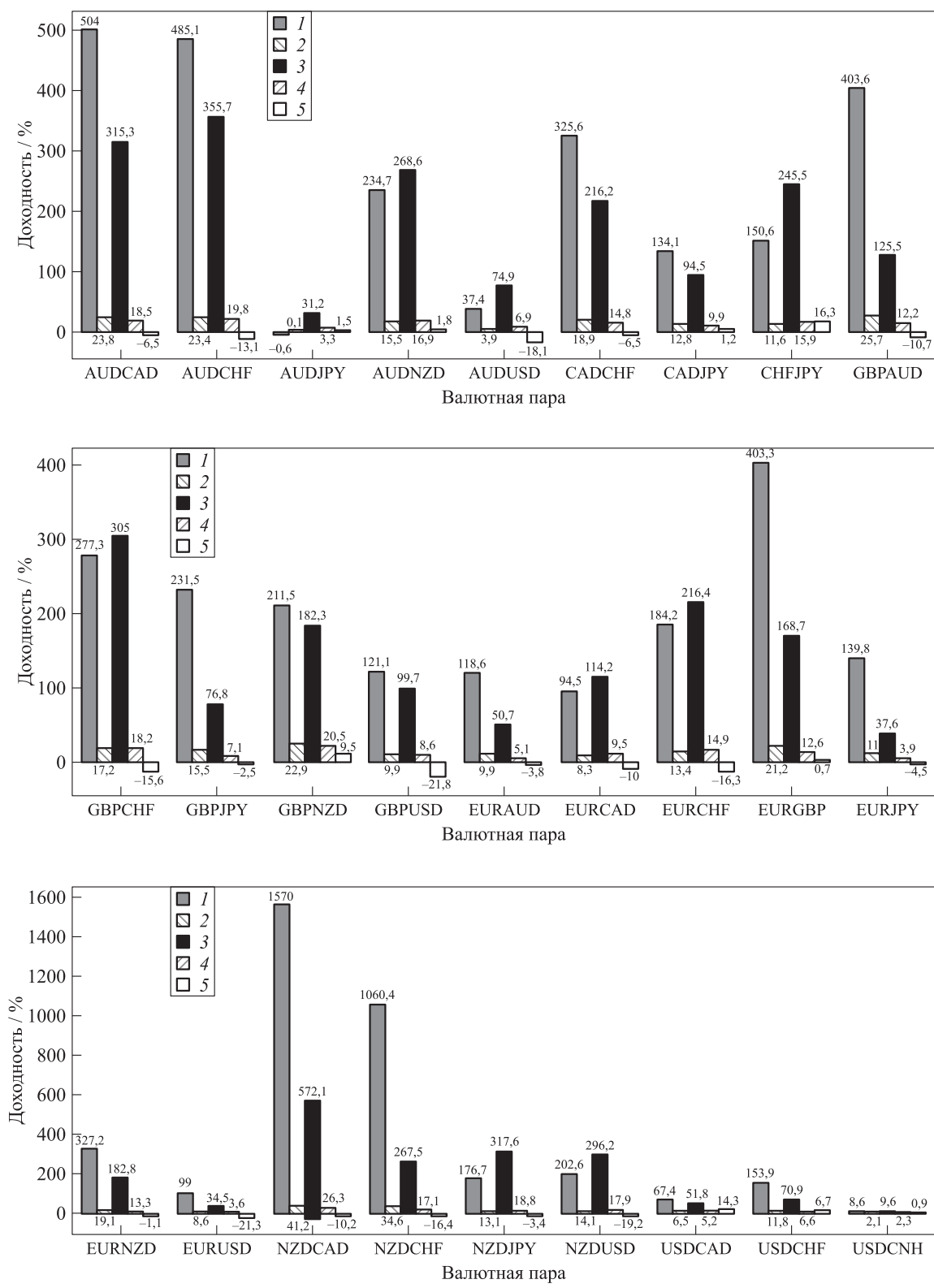


Рис. 6 Результаты сравнения стратегий на основе LSTM (1 — за 5 лет и 4 — годовая), индикатора тиковых объемов (2 — за 5 лет и 5 — годовая) и Buy & Hold за 5 лет (3)



**Рис. 7** Сравнение доходностей LSTM-стратегии (1 — доходность за 8 лет; 2 — годовая доходность), стратегии на основе ансамблевой LSTM-архитектуры (3 — доходность за 8 лет; 4 — годовая доходность) и базовой стратегии Buy & Hold (5) за восьмилетний период для 27 валютных пар



периода. В среднем доходность по всем валютным парам за 5 лет составила 36,36%.

Чтобы результаты стратегии с нейронной сетью были сопоставимы с результатами тестирования индикатора, нейронные сети были протестированы на том же временном участке: они продемонстрированы на рис. 6. Кроме того, добавлено сравнение с классической стратегией Buy & Hold.

Средняя годовая доходность LSTM-стратегии — 18,2%. Это несколько выше средней годовой доходности за восьмилетний период, приведенной в разд. 4. Причина этого может быть связана с лучшим обучением нейронной сети на более поздних тестируемых периодах. За тот же период времени средняя годовая доходность стратегии на основе индикатора значительно ниже — 6,1%. При этом стоит отметить, что по валютным парам AUDJPY, AUDUSD и USDCNH годовая доходность индикатора превосходит годовую доходность LSTM-стратегии, причем для AUDJPY обе «технические» стратегии показали положительную годовую доходность.

## 6 Ансамблевая архитектура

Для улучшения результатов, описанных в предыдущих разделах, возможно построение стратегии, которая учитывает сигналы, генерируемые и нейронной сетью, и индикатором.

Для этого архитектура базовой нейронной сети (см. рис. 4) была изменена на ансамбль из двух LSTM-сетей: она имеет два входных потока, на вход одного из которых подаются прежние величины, соответствующие LSTM-стратегии, а на вход второго — значения индикатора, приведенные к аналогичному виду категориальных объектов путем отнесения значения к одному из десяти интервалов. Обе части нейронной сети включают в себя слои embedding, dropout и LSTM — эти фрагменты в точности соответствуют LSTM-стратегии. Далее выходные потоки попадают в слой concatenate, который объединяет их и единым потоком передает в выходной полносвязный слой с функцией активации softmax. Выходной слой генерирует итоговые вероятности, на основе которых создается прогноз движения цены.

Модифицированная нейронная сеть была протестирована на всех 27 ранее рассмотренных валютных парах (рис. 7) за период в 6–8 лет (по этой причине на графиках не приводятся значения индикатора). Для этого искусственные нейронные сети были предварительно обучены на данных каждой валютной пары, а затем с использованием полученных сигналов покупки и продажи было осуществлено бэк-тестирование модифицированной стратегии

на исторических данных. Обучение проходило по тому же принципу, что и для исходной LSTM-стратегии.

Можно заметить, что на 10 валютных парах стратегия с применением ансамбля LSTM-сетей превзошла первую LSTM-стратегию, продемонстрировав более высокую доходность. Для этих валютных пар среднее значение итоговой доходности возросло на 17,19%, а годовая доходность увеличилась на 0,409%. При этом средняя годовая доходность, рассчитанная для всех валютных пар, составила 12,2%, что ниже аналогичного показателя простой LSTM-стратегии. Однако можно обратить внимание на ранее убыточную пару AUDJPY, для которой получена итоговая доходность 31,2% и годовая 3,3%, что существенно превосходит значения для базовой стратегии Buy & Hold. Таким образом, в целом ряде случаев (37% тестируемых валютных пар) использование ансамблевого подхода позволяет значимым образом повысить доходность базовой LSTM-стратегии. Это обстоятельство необходимо учитывать при разработке реальных автоматизированных торговых инструментов, в рамках которых обычно не используется единый алгоритм (метод) сразу для всех объектов из портфеля.

## 7 Заключение

В работе продемонстрирована высокая эффективность использования нейросетевых подходов к построению автоматических торговых стратегий на основе валютных пар. Показано, что предложенные методы позволяют многократно превзойти базовые методы, в том числе основанные на классическом техническом анализе. При этом можно отметить, что многие валютные пары обладают высокой корреляцией — как положительной, так и отрицательной (рис. 8).

Можно предположить, что для большей информативности признаков, созданных на основе значений индикатора, в дальнейших исследованиях стоит использовать уникальное разбиение на классы для разных валютных пар, а не унифицированное, как это реализовано в данной работе. Потенциально подбор порогов для разбиения на классы может повысить доходность стратегии с применением ансамбля LSTM-сетей, что сделает модификацию LSTM-стратегии перспективной областью дальнейшего исследования, включая использование композиционных и ансамблевых подходов.

Принципиальное улучшение может быть достигнуто за счет привлечения сложных математических методов и моделей, в частности использования параметров семейств вероятностных распределений для расширения признакового пространства

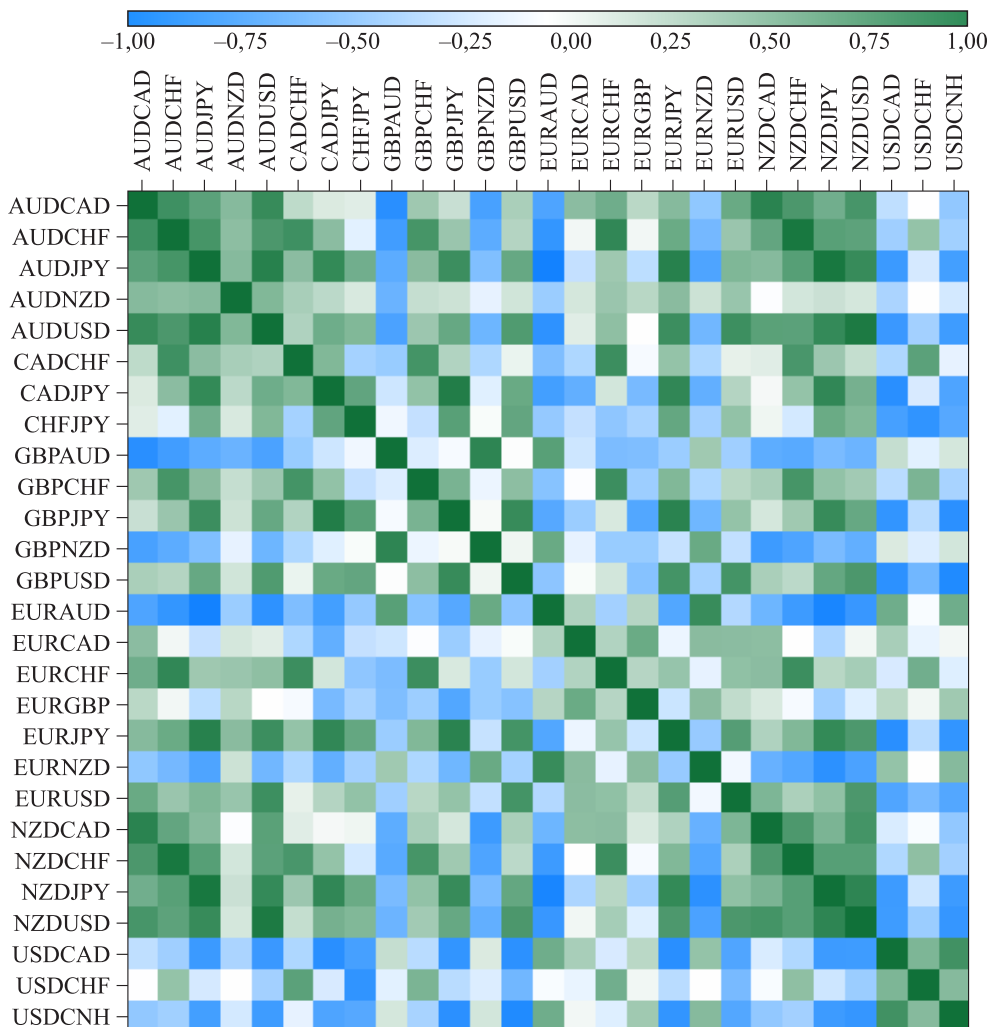


Рис. 8 Корреляционная матрица валютных пар

для обучения нейронных сетей, как предложено в статье [15] на примере различных моделей добавления дополнительных признаков — четырех первых моментов конечных нормальных смесей.

Определенные успехи в этом направлении для ряда валютных пар уже продемонстрированы [16], что свидетельствует в пользу перспективности подобных исследований.

## Литература

1. *Bholat D., Susskind D.* The assessment: Artificial intelligence and financial services // *Oxford Rev. Econ. Pol.*, 2021. Vol. 37. Iss. 3. P. 417–434.
2. *Lin Y., Liu S., Yang H., Wu H.* Stock trend prediction using candlestick charting and ensemble machine learning techniques with a novelty feature engineering scheme // *IEEE Access*, 2021. Vol. 9. P. 101433–101446.
3. *Lahmiri S., Bekiros S.* Intelligent forecasting with machine learning trading systems in chaotic intraday bitcoin market // *Chaos Soliton. Fract.*, 2020. Vol. 133. Art. No. 109641. 7 p.
4. *Rundo F.* Deep LSTM with reinforcement learning layer for financial trend prediction in FX high frequency trading systems // *Appl. Sci.* — Basel, 2019. Vol. 9. Iss. 20. Art. No. 4460. 18 p.
5. *Dinh T.-A., Kwon Y.-K.* An empirical study on importance of modeling parameters and trading volume-based features in daily stock trading using neural networks // *Informatics*, 2018. Vol. 5. Iss. 3. Art. No. 36. 12 p.
6. *Sezer O. B., Ozbayoglu A. M.* Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach // *Appl. Soft Comput.*, 2018. Vol. 70. P. 525–538.
7. *Tsinaslanidis P., Guijarro F.* What makes trading strategies based on chart recognition profitable? // *Expert Syst.*, 2020. Vol. 38. Iss. 5. Art. No. e12596. 17 p.

8. *Chen J. H., Tsai Y. C.* Encoding candlesticks as images for pattern classification using convolutional neural networks // *Financial Innovation*, 2020. Vol. 6. Art. No. 26. 19 p.
9. *Hung C.-C., Chen Y.-J.* DPP: Deep predictor for price movement from candlestick charts // *PLoS ONE*, 2021. Vol. 16. Iss. 6. Art. No. e0252404. 22 p.
10. *Wu J. M.-T., Li Z., Herencsar N., Vo B., Lin J. C.-W.* A graph-based CNN-LSTM stock price prediction algorithm with leading indicators // *Multimedia Syst.*, 2021. 20 p. doi: 10.1007/s00530-021-00758-w.
11. *Sezer O. B., Ozbayoglu A. M., Dogdu E.* A deep neural-network based stock trading system based on evolutionary optimized technical analysis parameters // *Procedia Comput. Sci.*, 2017. Vol. 114. P. 473–480.
12. *JuHyok U., PengYu L., ChungSong K., UnSok R., KyongSok P.* A new LSTM based reversal point prediction method using upward/downward reversal point feature sets // *Chaos Soliton. Fract.*, 2020. Vol. 32. Art. No. 109559. 15 p.
13. *Tsantekidis A., Tefas A.* Transferring trading strategy knowledge to deep learning models // *Knowl. Inf. Syst.*, 2021. Vol. 63. P. 87–104.
14. *Edwards R. D., Magee J., Bassetti W. H. C.* Technical analysis of stock trends. — 11th ed. — Boca Raton, FL, USA: CRC Press, 2018. 685 p.
15. *Gorshenin A. K., Kuzmin V. Yu.* Statistical feature construction for forecasting accuracy increase and its applications in neural network based analysis // *Mathematics*, 2022. Vol. 10. Iss. 4. Art. No. 589. 21 p.
16. *Виляев А. Л., Горшенин А. К.* О моделировании торговых стратегий для валютных пар с использованием глубоких нейронных сетей и метода скользящего разделения смесей // *Интеллектуальные системы. Теория и приложения*, 2021. Т. 25. Вып. 4. С. 92–95.

*Поступила в редакцию 13.07.22*

## INCREASING FOREX TRADING PROFITABILITY WITH LSTM CANDLESTICK PATTERN RECOGNITION AND TICK VOLUME INDICATOR

A. K. Gorshenin<sup>1</sup> and E. I. Guseynova<sup>2</sup>

<sup>1</sup>Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

<sup>2</sup>M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

**Abstract:** The paper introduces the research of the effectiveness of using LSTM (Long–Short Term Memory) for candlestick data and a technical analysis indicator for a large number of the most common currency pairs (27 in total) over a long period in order to build automatic trading strategies. The achieved average total and annual return for 8 years of a model trading were 286% and 15.4%, respectively. It is more than 50 times higher than the values for the classic Buy & Hold trading strategy for the same period. In addition, the paper introduces a new technical indicator based on tick volumes which is an alternative trading strategy (the total and annual returns of LSTM models exceed it by an average of 7.2 and 2.3 times) as well as an additional feature to increase the profitability of the neural network strategy through the use of ensemble learning. For 37% of the analyzed currency pairs, the use of an ensemble of LSTMs allows one to increase further the total return by an average of 17.2%.

**Keywords:** LSTM; ensemble learning; candlestick; technical indicator; FOREX; currency pairs

**DOI:** 10.14357/19922264220304

### Acknowledgments

The research was supported by the Russian Science Foundation (grant No. 22-11-00212). 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. Bholat, D., and D. Susskind. 2021. The assessment: Artificial intelligence and financial services. *Oxford Rev. Econ. Pol.* 37(3):417–434.
2. Lin, Y., S. Liu, H. Yang, and H. Wu. 2021. Stock trend prediction using candlestick charting and ensemble machine learning techniques with a novelty feature engineering scheme. *IEEE Access* 9:101433–101446.
3. Lahmiri, S., and S. Bekiros. 2020. Intelligent forecasting with machine learning trading systems in chaotic intraday Bitcoin market. *Chaos Soliton. Fract.* 133:109641. 7 p.

4. Rundo, F. 2019. Deep LSTM with reinforcement learning layer for financial trend prediction in FX high frequency trading systems. *Appl. Sci. — Basel* 9(20):4460. 18 p.
5. Dinh, T.-A., and Y.-K. Kwon. 2018. An empirical study on importance of modeling parameters and trading volume-based features in daily stock trading using neural networks. *Informatics* 5(3):36. 12 p.
6. Sezer, O. B., and A. M. Ozbayoglu. 2018. Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Appl. Soft Comput.* 70:525–538.
7. Tsinaslanidis, P., and F. Guijarro. 2020. What makes trading strategies based on chart recognition profitable? *Expert Syst.* 38(5):e12596. 17 p.
8. Chen, J. H., and Y. C. Tsai. 2020. Encoding candlesticks as images for pattern classification using convolutional neural networks. *Financial Innovation* 6:26. 19 p.
9. Hung, C.-C., and Y.-J. Chen. 2021. DPP: Deep predictor for price movement from candlestick charts. *PLoS ONE* 16(6):e0252404. 22 p.
10. Wu, J. M.-T., Z. Li, N. Herencsar, B. Vo, and J. C.-W. Lin. 2021. A graph-based CNN-LSTM stock price prediction algorithm with leading indicators. *Multimedia Syst.* 20 p. doi: 10.1007/s00530-021-00758-w.
11. Sezer, O. B., A. M. Ozbayoglu, and E. Dogdu. 2017. A deep neural-network based stock trading system based on evolutionary optimized technical analysis parameters. *Procedia Comput. Sci.* 114:473–480.
12. JuHyok, U., L. PengYu, K. ChungSong, R. UnSok, and P. KyongSok. 2020. A new LSTM based reversal point prediction method using upward/downward reversal point feature sets. *Chaos Soliton. Fract.* 32:109559. 15 p.
13. Tsantekidis, A., and A. Tefas. 2021. Transferring trading strategy knowledge to deep learning models. *Knowl. Inf. Syst.* 63:87–104.
14. Edwards, R. D., J. Magee, and W. H. C. Bassetti. 2018. *Technical analysis of stock trends*. 11th ed. Boca Raton, FL: CRC Press. 685 p.
15. Gorshenin, A. K., and V. Yu. Kuzmin. 2022. Statistical feature construction for forecasting accuracy increase and its applications in neural network based analysis. *Mathematics* 10(4):589. 21 p.
16. Vilyaev, A. L., and A. K. Gorshenin. 2021. O modelirovanii torgovykh strategiy dlya valyutnykh par s ispol'zovaniem glubokikh neyronnykh setey i metoda skol'zyashchego razdeleniya smesey [On modeling trading strategies for currency pairs using deep neural networks and method of moving separation of mixtures]. *Intellektual'nye sistemy. Teoriya i prilozheniya* [Intelligent Systems. Theory and Applications] 25(4):92–95.

Received July 13, 2022

## Contributors

**Gorshenin Andrey K.** (b. 1986) — Doctor of Science in physics and mathematics, associate professor, head of department, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; agorshenin@freesc.ru

**Guseynova Ekaterina I.** (b. 1999) — Master of Science, Faculty of Economics, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; ei.guseynova@yandex.ru

# О ДОСТАТОЧНЫХ УСЛОВИЯХ ЭКСТРЕМУМА В МНОГОМЕРНЫХ ВАРИАЦИОННЫХ ЗАДАЧАХ

Н. С. Васильев<sup>1</sup>

**Аннотация:** Формализуя системные представления о теориях, вариационные принципы дают общий подход к моделированию широкого круга явлений из многих областей знания. Эти модели описывают искомые процессы как обеспечивающие стационарность некоторого «универсального» функционала. Соответствующие решения находят с помощью необходимых условий оптимальности, а достаточные условия позволяют ее доказывать. В естествознании это служит обоснованию принципов минимума энергии и действия Гамильтона. В статье получены достаточные условия минимальности экстремалей в многомерных вариационных задачах. К их числу относятся задачи, функционал которых содержит старшие производные нескольких искомым функций многих переменных. Доказательство теорем основано на использовании инвариантных поверхностных интегралов. Это позволило обобщить классические результаты.

**Ключевые слова:** экстремаль; экстремальная гиперповерхность; поле нормалей; дивергенция и поток векторного поля; дифференциальная форма; внешнее дифференцирование; инвариантность интеграла; множитель Лагранжа

**DOI:** 10.14357/19922264220305

## 1 Введение

Рассмотрены многомерные вариационные задачи, в которых требуется найти слабый, в пространстве  $C^1(S)$   $S \subset R^n$  или сильный, в пространстве  $C(S)$ ,  $S \subset R^n$ , локальный минимум интегрального функционала  $J : C^1(S) \rightarrow R$  вида

$$J(z) = \int_S f(x, z, z_x) dx \rightarrow \min \quad (1)$$

при заданном краевом условии

$$z|_{\gamma S} = u. \quad (2)$$

В (1) использовано краткое обозначение

$$z_x = (z_{x_1}, \dots, z_{x_n})$$

для градиента функции

$$z : S \rightarrow R.$$

Предполагается, что подынтегральная функция является гладкой  $f \in C^2[S \times W]$ . Будем выбирать окрестности рассматриваемых экстремалей так, чтобы они не покидали множество  $S \times W \subset R^{2n+1}$ ,  $(z, z_x) \in W$ . Кроме того, считаем, что замкнутая односвязная область  $S$  имеет кусочно-гладкую границу  $\gamma S$ . Оптимизация (1), (2) проводится на классе допустимых функций  $z \in C^1(S)$ ,

графики которых представляют собой гладкие гиперповерхности  $\Pi_z \subset R^{n+1}$  с краем  $(x, u) \subset R^{n+1}$ ,  $x \in \gamma S$ .

Одномерным аналогом вариационной задачи (1), (2),  $n = 1$ , служит простейшая вариационная задача, решаемая на отрезке  $[a, b]$  с фиксированными краевыми ограничениями

$$z(a) = z|_a; \quad z(b) = z|_b.$$

Для нее известны необходимые условия оптимальности решения  $z = z^*$ , называемого экстремалью. Во-первых, это условие Лежандра: квадратичная форма, определяемая матрицей Гессе  $(f_{z_{x_i} z_{x_j}})|_{z=z^*}$ , является неотрицательно определенной. Во-вторых, сама экстремаль  $z^*$  находится как решение уравнения Эйлера [1, 2].

В теоремах Якоби, Гильберта и Вейерштрасса сформулированы достаточные условия локальной оптимальности  $z^*$ , которые опираются на наличие поля экстремалей [1, 2]. Условие Якоби обеспечивает существование этого поля в некоторой области плоскости  $R^2$ , включающей график функции  $z^*$ . Это требование отсутствия сопряженной точки на полуинтервале  $[a, b]$ .

В многомерной вариационной задаче (1), (2) необходимые условия оптимальности выражаются в форме краевой задачи для уравнения Эйлера–Остроградского, решаемой при ограничении (2) (см. [1, 2]).

<sup>1</sup>Московский государственный технический университет имени Н. Э. Баумана, nik8519@yandex.ru



В статье получены достаточные условия локального минимума в многомерной вариационной задаче (1), (2). Обоснование опирается на свойство инвариантности поверхностных интегралов, определенных в разд. 2 и 3. Это обобщение криволинейного инвариантного интеграла Гильберта [1, 2]. В разд. 3 и 4 свойство инвариантности применено к вариационным задачам, функционал которых зависит от нескольких искомых функций многих переменных и их старших производных.

## 2 Достаточные условия минимума

Рассмотрим семейство вариационных задач с функционалом (1) и различными краевыми условиями

$$z|_{\gamma S} = v, \quad v|_{\Gamma} = u|_{\Gamma}. \quad (3)$$

Здесь лишь на части границы  $\Gamma \subset \gamma S$  исходной области  $S$  значения искомой функции совпадают с функцией  $u$ , заданной в граничном условии (2). Это приводит к тому, что имеется множество решений задач (1), (3), определяемое выбором функции  $v$ . Все функции семейства  $\{z\}$  суть решения вариационных задач с функционалом (1) и граничными условиями  $z|_{\gamma S} = \tilde{v}$ . Здесь  $\tilde{v}$  — гладкое продолжение заданной функции  $u|_{\Gamma}$  с части границы  $\Gamma$ ,  $\Gamma \subset \gamma S$ , на всю границу  $\gamma S$  области  $S$ . Все указанные функции удовлетворяют необходимому условию минимальности функционала (1). Их можно найти как решения краевой задачи для уравнения Эйлера—Остроградского [1, 2]:

$$\frac{\partial f}{\partial z} - \sum_i \frac{d}{dx_i} \frac{\partial f}{\partial z_{x_i}} = 0, \quad z|_{\gamma S} = \tilde{v}. \quad (4)$$

В семействе  $\{z\}$  содержится искомое решение  $z^*$ , удовлетворяющее исходному ограничению (2) при  $\tilde{v} = u$ . Требуется проверить его оптимальность.

Пусть существует область  $G \subset R^{n+1}$ , содержащая поверхность  $\Pi_{z^*}$ , через каждую точку которой проходит единственная экстремальная поверхность  $\Pi_z$ , которая является графиком функции  $z \in \{z\}$  (см. (4)). В этом случае будем говорить, что в области  $G$  имеется *поле экстремальных поверхностей*, включающее  $\Pi_{z^*}$ . Благодаря этому в области  $G$  существует поле нормалей к поверхностям  $\Pi_z$ :

$$n_z(x, z) = (-z_x, 1), \quad n_z \perp \Pi_z. \quad (5)$$

В каждой точке  $(x, z) \in G$  существует наклон поля — вектор  $p = p(x, z) = z_x$  (см. (5)), характеризующий

касательное пространство к поверхности  $\Pi_z$ . С его помощью в области  $G$  определим векторное поле, строящееся по функционалу (1):

$$V(x, z) = \left( -f_{z_x}(x, z, p(x, z)), f(x, z, p(x, z)) - \sum_i p^i(x, z) f_{z_{x_i}}(x, z, p(x, z)) \right). \quad (6)$$

В поле  $V \equiv (V_1, V_2)$  имеются векторная  $V_1(x, z) \in R^n$  и скалярная  $V_2(x, z) \in R$  составляющие

$$V_1 = -f_{z_x}(x, z, p); \quad V_2 = f(x, z, p) - (p, f_{z_x}(x, z, p)).$$

Здесь сумма, входящая в определение (6), записана кратко в виде скалярного произведения  $(p, f_{z_x})$  указанных векторов. Вычислим дивергенцию поля  $V$  — его инвариант вида [3]:

$$\operatorname{div} V = \sum_i \frac{\partial}{\partial x_i} V_1^i + \frac{\partial}{\partial z} V_2.$$

**Лемма 2.1.** *Справедливо равенство  $\operatorname{div} V(x, z) \equiv 0$ ,  $(x, z) \in G$ .*

*Доказательство.* Через всякую точку  $(x, z) \in G$  проходит некоторая экстремаль  $z = \bar{z}(x)$ , для которой, согласно (5), имеют место соотношения:

$$z_{x_i} = p^i(x, z), \quad z_{x_i x_i} = p_{x_i}^i + p^i p^i.$$

Всякая экстремаль удовлетворяет уравнению (4), а значит, выполнены равенства:

$$\begin{aligned} 0 &= \sum_i \frac{d}{dx_i} V_1^i|_{z=\bar{z}(x)} + f_z = \\ &= - \sum_i (f_{z_{x_i} x_i} + f_{z_{x_i} z} z_{x_i} + f_{z_{x_i} z_{x_i}} z_{x_i x_i}) + f_z = \\ &= \sum_i \frac{\partial}{\partial x_i} V_1^i + f_z - \sum_i (z_{z_{x_i} z} p^i + f_{z_{x_i} z_{x_i}} p^i p^i) = \\ &= \sum_i \frac{\partial}{\partial x_i} V_1^i + \frac{\partial}{\partial z} V_2 = \operatorname{div} V(x, z). \end{aligned}$$

Это и требовалось доказать.

Для краткости внешнее произведение  $dx_1 \wedge \dots \wedge dx_n$  1-форм  $dx_i$  обозначим  $dx$  [3]. С учетом равенств

$$n_{z_i} = \frac{\partial z}{\partial x_i}, \quad i = 1, \dots, n,$$

проинтегрируем дифференциальную  $n$ -форму  $\omega = (V, n) d\varphi$  по поверхности  $\Phi$  [3]. В результате

получим величину  $I(\Phi) = \int_{\Phi} \omega$  — поток поля  $V$  через поверхность  $\Phi$ :

$$I(\Phi) = \int_{\Phi} \left( f(x, z, p(x, z)) - \sum_i p^i(x, z) f_{z_{x_i}}(x, z, p(x, z)) \right) dx + \sum_k (-1)^{k+1} f_{z_{x_k}}(x, z, p(x, z)) dx_1 \wedge \dots \wedge dx_{k-1} \wedge dz \wedge dx_{k+1} \wedge \dots \wedge dx_n. \quad (7)$$

**Лемма 2.2.** *Поверхностный интеграл второго рода (7) инвариантен на множестве всех поверхностей  $\Phi = \Pi$  с одинаковой границей  $\gamma\Pi = \gamma$ .*

*Доказательство.* Склеим две поверхности  $\Pi_1$  и  $\Pi_2$ , являющиеся графиками функций  $z_1$  и  $z_2$ , по общей границе  $\gamma$ . Полученная замкнутая поверхность  $\Phi = \Pi_{z_1} \cup \Pi_{z_2}$ ,  $z_1 \geq z_2$ , ограничивает некоторое тело  $T$ . «Внешние» нормали к  $\Phi = \gamma T$  в точке  $(x, z)$  задаются векторами  $\pm n_z(x, z)$  вида (5), отличающимися только выбором знака в зависимости от участка  $\Pi_{z_1}$  или  $\Pi_{z_2}$  поверхности  $\Phi$ , на котором лежит эта точка. Внешнее дифференцирование [3] формы  $\omega = (V, n) d\varphi$  дает  $d\omega = \operatorname{div} V dt$ . Поэтому

$$\oint_{\gamma T^+} \omega = \int_{T^+} \operatorname{div} V dt.$$

Согласно лемме 2.1, выполнено равенство

$$I(\Pi_1) - I(\Pi_2) = \oint_{\Phi} \omega = \int_{T^+} \operatorname{div} V dx = 0,$$

означающее инвариантность интеграла:  $I(\Pi_1) = I(\Pi_2)$ . Лемма 2.2 доказана.

По доказанному, интеграл  $I(\Pi)$ ,  $\gamma\Pi = \gamma$ , естественно обозначать  $I_{\gamma}$ , указывая лишь край поверхности. Особый интерес представляют собой поверхности  $\Pi$ , задаваемые графиками гладких функций  $z$  из семейства  $\{z\}$ . Они однозначно проецируются на фигуру  $S$ . Тогда инвариантный интеграл (7) допускает запись вида

$$I_{\gamma} = I(\Pi) = \int_S (f(x, z, p(x, z)) + (z_x - p(x, z), f_{z_x}(x, z, p(x, z)))) dx. \quad (8)$$

**Теорема 2.1.** *Пусть в области  $G$  существует поле экстремальных поверхностей с нормальными  $(-p(x, z), 1)$ , включающее экстремаль  $z^*$  задачи (1), (2). Если выполнено усиленное условие Лежандра  $\forall (x, z) \in G \Rightarrow (f_{z_{x_i} z_{x_j}})|_{z_x=p(x, z)} > 0$ , то  $z^*$  является слабым локальным минимумом функционала (1).*

*Если дополнительно для любых  $(x, z) \in G$  подынтегральная функция  $f$  выпукла по переменной  $z_x$ , то  $z^*$  — сильный локальный минимум в  $G$ .*

*Доказательство.* Так как  $z_x^* = p(x, z^*)$ , то  $J(z^*) = I(\Pi_{z^*})$  (см. (7)). По лемме 2.2,  $I(\Pi_{z^*}) = I(\Pi)$ ,  $\gamma\Pi = \gamma S$ , для любой функции  $z$ ,  $z|_{\gamma S} = u$ , имеющей график  $\Pi \subset G$ . Тогда приращение функционала (1) может быть преобразовано к виду

$$J(z) - J(z^*) = \int_S f(x, z, z_x) dx - I(\Pi_{z^*}) = \int_S f(x, z, z_x) dx - I(\Pi).$$

Учитывая (8), перепишем правую часть этого равенства и получим

$$J(z) - J(z^*) = \int_S (f(x, z, z_x) - f(x, z, p(x, z)) - (z_x - p(x, z), f_{z_x}(x, z, p(x, z)))) dx \geq 0. \quad (9)$$

Выполнение условия Лежандра обеспечивает неотрицательность подынтегральной функции в выражении (9). Это неравенство справедливо для любой функции  $f(z, x, z_x)$ , локально выпуклой по переменной  $z_x$ . Таким образом, слабая локальная оптимальность точки  $z^*$  доказана. В случае выпуклости  $f$  по этой переменной неравенство (9) имеет место при подстановке любых гладких функций  $z$ , лишь бы их графики лежали в области  $G$ . Тогда неравенство  $J(z) \geq J(z^*)$  справедливо и в сильной окрестности точки  $z^*$ . Это завершает доказательство теоремы 2.1.

**Пример 2.1.** Пусть имеется вариационная задача, в которой

$$J(z) = \int_S \left( \left( \frac{\partial z}{\partial x} \right)^2 + \left( \frac{\partial z}{\partial y} \right)^2 - 2z^2 \right) dx dy \rightarrow \min_{z: z|_{\gamma S} = u};$$

$$S = \{(x, y) : 0 \leq x, y \leq 1\};$$

$$u = \begin{cases} \sin(y - 1), & x = 0; \\ -\sin 1 \cos(x), & y = 0; \\ x, & y = 1; \\ y, & x = 1. \end{cases}$$

Согласно уравнению Эйлера—Остроградского (4), необходимо, чтобы все экстремальные поверхности были решениями следующей первой краевой задачи для уравнения Пуассона ( $\Delta$  — оператор Лапласа):

$$\Delta z = -2z, \quad z|_{\Gamma} = x|_{\Gamma}.$$

В качестве  $\Gamma$  выберем часть границы квадрата, выделяемой уравнением  $y = 1$ . Покажем, что экстремаль  $z^* = xy + \sin(y - 1) \cos(x)$  удовлетворяет всем условиям теоремы 2.1. Действительно, однопараметрическое семейство функций

$$z = Cxy + (1 - C)x + \sin(y - 1) \cos(x)$$

образует поле экстремальных поверхностей в области  $G = S \times R$ , включающее  $z^*$ . По теореме 2.1, отображение  $z^*$  является даже глобальным сильным минимумом.

### 3 Многомерные задачи с несколькими искомыми функциями

Рассмотрим вариационную задачу относительно двух неизвестных функций  $z, y$ :

$$J(z, y) = \int_S f(x, z, z_x, y, y_x) dx \rightarrow \min_{\substack{z|_{\gamma, S} = u_1 \\ y|_{\gamma, S} = u_2}}. \quad (10)$$

Предположим, что условия теоремы 2.1 выполняются для каждого из функционалов  $J(z, y^*)$ ,  $J(z^*, y)$ , имеющих одну неизвестную функцию  $z$  или  $y$ , причем их экстремальными являются  $z^*$  и  $y^*$  соответственно. Тогда точку  $(z^*, y^*)$  назовем *раздельным* слабым локальным минимумом в задаче (10). Раздельное условие Лежандра означает то, что оно выполнено по каждой из переменных  $z_x, y_x$  для обоих интегрантов из функционалов  $J(z, y^*)$ ,  $J(z^*, y)$ . Итак, доказано

**Следствие 3.1.** Пусть выполнено раздельное усиленное условие Лежандра и в некоторых областях, содержащих экстремали  $z^*$  и  $y^*$ , существуют поля экстремальных поверхностей  $p(x, z)$ ,  $(x, z) \in G_1$ , и  $q(x, y)$ ,  $(x, y) \in G_2$ . Тогда  $z^*, y^*$  — раздельный слабый локальный минимум в (10).

Вместо поиска минимума в задаче (10) зачастую требуется найти минимакс этого функционала по переменным  $z$  и  $y$ . Например, это так, если рассматривается вариационная задача на условный экстремум, имеющая дополнительные фазовые ограничения, которые снимаются с помощью множителей Лагранжа  $y$  [4]. Тем самым исходная задача сводится к задаче на минимакс для функции Лагранжа.

**Следствие 3.2.** Если существуют поля экстремальных поверхностей  $p(x, z)$  и  $q(x, y)$  в некоторых областях  $G_i, i = 1, 2$ , и для любых  $(x, z) \in G_1$  и  $(x, y) \in G_2$  подынтегральная функция  $f$  выпукла по

переменной  $z_x$  и вогнута по переменной  $y_x$ , то пара  $z^*, y^*$  — седловая точка функционала (10).

Для получения более сильных результатов рассмотрим фазовое пространство  $G = G_1 \times G_2$ . В частности, это означает, что введены дополнительные переменные  $x'_{n+1} = x_1, \dots, x'_{2n} = x_n$ , отвечающие исходному вектору  $x$ . По предположению, имеется поле наклонов экстремальных поверхностей ( $p = p(x, z)$ ,  $q = q(x', y)$ ), снова позволяющее определить векторное поле, исходя из функционала вариационной задачи (10):

$$V \equiv (V_1^1, V_1^2; V_2^1, V_2^2), \\ V_1 = (V_1^1, V_1^2), V_2 = (V_2^1, V_2^2),$$

где

$$V_1^1 = -f_{z_x}(x, z, y, p, q); \\ V_1^2 = f(x, z, y, p, q) - (p, f_{z_x}(x, z, y, p, q)); \\ V_2^1 = -f_{y_x}(x, z, y, p, q); \\ V_2^2 = f(x, z, y, p, q) - (q, f_{y_x}(x, z, y, p, q)).$$

В области  $G$  находится поверхность  $S = z \times y$ , через которую «проходит» поток поля  $V$ . Так как нормаль к этой поверхности равна  $n = (-z_x, 1, -y_x, 1)$ , то этот поток равен

$$I(S) = \int_S (n, V) ds = \\ = \int_S (2f(x, z, y, p, q) + (z_x - p, f_{z_x}(x, z, y, p, q)) + \\ + (y_x - q, f_{y_x}(x, z, y, p, q))) ds.$$

Рассмотрим две гиперповерхности  $S_1$  и  $S_2$ , определяемые парами допустимых функций  $(z_1, y_1)$  и  $(z_2, y_2)$ . Как и при доказательстве леммы 2.2, склеим поверхности  $S_1$  и  $S_2$  по общей границе, определив некоторое тело  $T$ , имеющее границу  $S_1 \cup S_2$ .

**Лемма 3.1.** Интеграл  $I(S)$  инвариантен.

**Доказательство.** Как и в лемме 2.2, воспользуемся равенством интегралов от дифференциальных форм  $\omega = (V, n) d\varphi$  и  $d\omega = \operatorname{div} V dt$ , взятых соответственно по поверхности  $S_1 \cup S_2$  и по телу  $T$ . По лемме 2.1,  $\operatorname{div} V_1 = \operatorname{div} V_2 = 0$ . Так как  $\operatorname{div} V = \operatorname{div} V_1 + \operatorname{div} V_2$ , то  $\operatorname{div} V = 0$ , поэтому интеграл по замкнутой поверхности  $S_1 \cup S_2$  равен нулю, что доказывает его инвариантность:  $I(S_1) = I(S_2)$ .

**Теорема 3.1.** Пусть в области  $G$  существует поле экстремальных поверхностей с коэффициентами  $(p(x, z), q(x', y))$ , включающее экстремаль  $z^*, y^*$  задачи (10). Если выполнено усиленное условие Лежандра по переменным  $w = (z_x, y_x)$ , то эта экстремаль

является слабым локальным минимумом функционала (10).

Доказательство. Рассмотрим величину

$$\begin{aligned} \Pi(S) &= \\ &= \int_S (f(x, z, y, p, q) + (z_x - p, f_{z_x}(x, z, y, p, q)) + \\ &\quad + (y_x - q, f_{y_x}(x, z, y, p, q))) ds, \end{aligned}$$

являющуюся аналогом выражения (8). Не ограничивая общности, можно считать, что  $f(x, z, y, p, q) \leq 0$  в области  $G$ . Тогда найденный инвариант  $I(S)$  (см. лемму 3.1) связан с интегралом  $\Pi(S)$  соотношением  $\Pi(S) \geq I(S)$ . Отсюда приращение функционала задачи (10) оценивается с помощью следующего неравенства:

$$\begin{aligned} J(z, y) - J(z^*, y^*) &= J(z, y) - I(S) \geq \\ &\geq J(z, y) - \Pi(S). \end{aligned} \quad (11)$$

Как показано при доказательстве теоремы 2.1, применение усиленного условия Лежандра к интегранту функционала из правой части неравенства (11) обосновывает его неотрицательность (см. также (9)). В итоге  $J(z, y) \geq J(z^*, y^*)$ , что и требовалось обосновать.

#### 4 Вариационные задачи, содержащие старшие производные

Пусть функционал вариационной задачи имеет вид:

$$J(z, y) = \int_S f(x, z, z_x, z_{xx}) dx \rightarrow \min_{\substack{z|_{\gamma S} = u_1; \\ z_x|_{\gamma S} = u_2}}. \quad (12)$$

Введем вспомогательные переменные  $y = z_x$  и вместо (12) рассмотрим задачу условной оптимизации

$$\left. \begin{aligned} J(z, y) &= \int_S f(x, z, y, y_x) dx \rightarrow \min_{\substack{z|_{\gamma S} = u_1; \\ z_x|_{\gamma S} = u_2}}; \\ y &= z_x. \end{aligned} \right\} \quad (13)$$

Для снятия фазового ограничения в (13) воспользуемся правилом множителей Лагранжа [4]. Получим вариационную задачу на минимум, эквивалентную исходной задаче (12):

$$\left. \begin{aligned} J(z, y, \lambda) &= \int_S [f(x, z, y, y_x) + \lambda(z_x - y)] dx \rightarrow \\ &\rightarrow \min_{z, y} \sup_{\lambda}; \\ z|_{\gamma S} &= u_1; \quad y|_{\gamma S} = u_2. \end{aligned} \right\} \quad (14)$$

В ней требуется найти седловую точку  $(z^*, y^*; \lambda^*)$ .

Наряду с искомым решением задачи рассмотрим экстремали критерия (14), удовлетворяющие варьированным граничным условиям  $z|_{\Gamma} = u_1|_{\Gamma}$ ,  $z_x|_{\Gamma} = u_2|_{\Gamma}$ ,  $\Gamma \subset \gamma S$ . Тогда, применив следствие 3.2 и теорему 3.1, получим обобщение следствия 3.1.

**Теорема 4.1.** Пусть существуют поля экстремалей  $p(x, z)$ ,  $q(x, y)$  и  $r(x, \lambda)$  функционала (14), включающие экстремаль  $(z^*, y^*; \lambda^*)$ . Если подынтегральная функция в (14) выпукла по переменной  $w = (z_x, y_x)$ , то эта экстремаль является седловой точкой, а пара  $(z^*, y^*)$  — локальным минимумом в задаче (12).

**Пример 4.1.** Рассмотрим следующую вариационную задачу:

$$\left\{ \begin{aligned} J(z) &= \int_0^{\pi} \left[ -z^2 - \frac{z_x^2}{2} + \frac{z_{xx}^2}{2} \right] dx \rightarrow \min; \\ z(0) &= z_x(0) = 0; \quad z(\pi) = 1; \quad z_x(\pi) = 0. \end{aligned} \right.$$

Требуется исследовать на локальный минимум экстремаль  $z = z^*$ ,

$$z = C_1 e^x + C_2 e^{-x} \cos(x) + C_3 e^{-x} \sin(x), \quad (15)$$

в которой вектор констант  $C = \overline{C}$  равен

$$\left. \begin{aligned} \overline{C}_1 &= \frac{e^{\pi}}{2(1 + e^{2\pi})}; \\ \overline{C}_2 &= -\frac{2 + e^{2\pi}}{2} \overline{C}_1; \\ \overline{C}_3 &= -\frac{\overline{C}_1}{2}. \end{aligned} \right\} \quad (16)$$

В самом деле, необходимое условие седловой точки функционала (14) выражается с помощью системы уравнений Эйлера

$$\left. \begin{aligned} y_{xx} + \lambda + y &= 0, \\ \lambda_x + 2z &= 0, \\ z_x - y &= 0. \end{aligned} \right\} \quad (17)$$

Ищутся решения, удовлетворяющие однородным граничным условиям  $z(0) = y(0) = 0$ . Тогда поля  $p(x, z)$  и  $q(x, y)$ , определяемые этими экстремальными поверхностями, задаются с помощью однопараметрического семейства функций (15),  $C_3 = -2C_1$ ,  $C_2 = -C_1$ , являющихся решениями краевых задач для системы дифференциальных уравнений (17).

В свою очередь, поле  $r(x, \lambda)$  вычисляется с помощью семейства функций  $\lambda_C = \lambda_C(x)$ , получаемых из первого уравнения системы (17) подстановкой функции  $y = z_x$  (см. (15)). Таким образом,

выполнены все условия теоремы 4.1 и следствия 3.2. Поэтому экстремаль (15), (16),  $C = \overline{C}$ , является слабым локальным минимумом рассмотренного функционала.

## 5 Заключение

Вариационное исчисление не утратило своей актуальности. Например, в современных исследованиях ищут обобщенные решения дифференциальных нелинейных уравнений не интегрированием, а с применением вспомогательной вариационной задачи. Достаточные условия служат проведению более тонкого анализа свойств найденной экстремальной поверхности с целью установления ее оптимальности. Существование поля экстремалей обеспечено следующим обобщением

условия Якоби. Требуется, чтобы возмущение краевых ограничений исходной вариационной задачи приводило не только к однозначности возникновения новых экстремалей, но и исключало их попарные пересечения.

## Литература

1. Эльсгольц Л. Э. Дифференциальные уравнения и вариационное исчисление. — М.: Наука, 1969. 424 с.
2. Янг Л. Лекции по вариационному исчислению и оптимальному управлению / Пер. с англ. — М.: Мир, 1974. 488 с. (Young L. C. Lectures on the calculus of variations and optimal control theory. — Saunders, 1969. 331 p.)
3. Ильин В. А., Позняк Э. Г. Основы математического анализа. — М.: Наука, 2000. Ч. II. 448 с.
4. Васильев Ф. П. Численные методы решения экстремальных задач. — М.: Наука, 1980. 520 с.

Поступила в редакцию 13.01.22

---



---

# ON EXTREMUM SUFFICIENT CONDITIONS IN MULTIDIMENSIONAL VARIATION CALCULUS PROBLEMS

N. S. Vasilyev

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

**Abstract:** Variation principles give formalization and general approach to construct and study the models from different fields of knowledge. They provide system presentations about theories origins. In the models, sought-for solution is a stationary point of a criterion. Its search on the basis of necessary conditions should not accomplish problem investigation. Sufficient conditions are needed to assert its optimality. In natural sciences, such results substantiate principles of energy or Hamilton's action minimization. In the paper, invariant surface integrals discovery gave possibility to prove minimum availability in multidimensional variation calculus problems. The functional in the problems may depend on several unknown functions of many variables and their high-order derivatives. Classical theorems are generalized.

**Keywords:** extremal; extreme hypersurface; field of normals; divergence and flow of a vector field; differential form; external differentiation; integral invariance; Lagrange multiplier

**DOI:** 10.14357/19922264220305

## References

1. El'sgol'ts, L. E. 1969. *Differentsial'nye uravneniya i variatsionnoe ischislenie* [Differential equations and the calculus of variations]. Moscow: Nauka. 424 p.
2. Young, L. C. 1969. *Lectures on the calculus of variations and optimal control theory*. Saunders. 331 p.
3. Il'in, V. A., and E. G. Poznyak. 2000. *Osnovy matematicheskogo analiza* [Basics of mathematical analysis]. Moscow: Nauka. Part II. 448 p.
4. Vasil'ev, F. P. 1980. *Chislennyye metody resheniya ekstremal'nykh zadach* [Numerical methods for solving extremal problems]. Moscow: Nauka. 520 p.

Received January 13, 2022

## Contributor

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



# ПОСЛЕДОВАТЕЛЬНЫЙ АНАЛИЗ И МЕТРИЧЕСКИЕ ОЦЕНКИ ПРЕДЕЛЬНЫХ РАСПРЕДЕЛЕНИЙ МЕЖУЗЛОВЫХ ПОТОКОВ В МНОГОПОЛЬЗОВАТЕЛЬСКОЙ СЕТИ

Ю. Е. Малащенко<sup>1</sup>

**Аннотация:** Для оценки функциональных возможностей многопользовательской сети связи анализируется множество векторов межузловых потоков при предельных распределениях ресурсов сети. В рамках многопродуктовой модели пропускные способности ребер рассматриваются как компоненты вектора ресурсов различных типов, которые требуются для передачи потоков различных видов. При проведении вычислительных экспериментов на каждой итерации вычисляются нормы векторов совместно допустимых межузловых потоков, при передаче которых полностью используется пропускная способность всех ребер сети. Полученные последовательности метрических оценок позволяют анализировать особенности множества достижимости и эффективность использования ресурсов сети при равномерном распределении пропускной способности между корреспондентами.

**Ключевые слова:** многопродуктовая потоковая сетевая модель; множество достижимых межузловых потоков; предельные распределения пропускной способности

**DOI:** 10.14357/19922264220306

## 1 Введение

Данная работа продолжает исследования функциональных характеристик сетевых систем связи [1]. В настоящее время математические модели передачи многопродуктового потока применяются для поиска распределений потоков и ресурсов в многопользовательских телекоммуникационных сетях [2–4]. Разрабатываются методы анализа с учетом вектора требований всех *равноправных* и взаимозаменяемых корреспондентов [5]. С позиций методологии исследования операций изучаются справедливые распределения потоков и ресурсов [6].

Соответствующие *недискриминирующие* правила управления потоками являются решениями задач на максимум и/или получаются в результате использования процедур последовательной лексикографически упорядоченной оптимизации [7].

В настоящей работе пути соединения корреспондентов прокладываются через соответствующие минимальные разрезы. Указанный метод можно рассматривать как возможный вариант решения задачи о построении SPLIT-маршрутов [8, 9]. В рамках вычислительных экспериментов на многопродуктовой модели анализируются распределения межузловых потоков и пропускной способности сети. Для оценки функциональных возможностей многопользовательской сети используется вектор совместно допустимых меж-

узловых потоков. Под ресурсом, выделяемым некоторой паре узлов-корреспондентов, понимается суммарное значение требуемых пропускных способностей на всех ребрах, расположенных на всех маршрутах при прохождении межузлового потока данного вида. Сумма соответствующих реберных потоков трактуется как полная нагрузка на сеть, возникающая при передаче заданного межузлового потока. Рассматриваются распределения пропускной способности и межузловых потоков при предельной загрузке сети. При проведении вычислительных экспериментов на каждой итерации вычисляется норма вектора совместно допустимых межузловых потоков. Для оценки величины требуемых ресурсов при соединении корреспондентов по различным путям для каждой пары узлов определяется максимальный однопродуктовый поток. Маршруты передачи всех допустимых межузловых потоков проходят по ребрам соответствующих минимальных разрезов. Вычислительные эксперименты проводились для получения последовательности метрических оценок векторов межузловых потоков, принадлежащих множеству достижимости многопользовательской сети.

## 2 Математическая модель

В качестве математической модели для описания многопользовательской сетевой системы ис-

<sup>1</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии mala-yur@yandex.ru

пользуется следующая формальная запись условий и ограничений, которые должны выполняться при одновременной передаче потоков различных видов между всеми парами улов-корреспондентов:

Сеть  $G(\mathbf{d})$  задается множествами  $\langle V, R, U, P \rangle$ :

- узлов (вершин) сети

$$V = \{v_1, v_2, \dots, v_n, \dots, v_N\};$$

- неориентированных ребер

$$R = \{r_1, r_2, \dots, r_k, \dots, r_E\}.$$

Ребро  $r_k$  соединяет концевые вершины  $v_{n_k}$  и  $v_{j_k}$ . Ребру  $r_k$  ставятся в соответствие две ориентированные дуги  $\{u_k, u_{k+E}\}$  из множества ориентированных дуг  $U = \{u_1, u_2, \dots, u_k, \dots, u_{2E}\}$ . Дуги  $\{u_k, u_{k+E}\}$  определяют прямое и обратное направление передачи потока по ребру  $r_k$  между концевыми вершинами  $\{v_{n_k}, v_{j_k}\}$ .

В многопользовательской сети  $G(\mathbf{d})$  рассматривается  $M = N(N - 1)$  независимых, невзаимозаменяемых и равноправных потоков различных видов, которые передаются между узлами-корреспондентами из множества

$$P = \{p_1, p_2, \dots, p_M\}.$$

По определению, каждой паре узлов-корреспондентов  $p_m$  соответствуют:

- вершина-источник с номером  $s_m$ , через которую входной поток  $m$ -го вида  $z_m$  поступает в сеть;
- вершина-приемник с номером  $t_m$ , из которой поток  $m$ -го вида  $z_m$  покидает сеть.

В множестве  $P$  выделяется подмножество  $P(R^+)$  пар узлов-корреспондентов, расположенных в концевых вершинах ребра  $r_k$ ,  $k = \overline{1, E}$ . Вводятся следующие обозначения: пусть ребро  $r_k$  с номером  $k$  соединяет вершины с номерами  $n$  и  $j$  такими, что  $n < j$ . Для соответствующей пары узлов-корреспондентов  $p_k$ , расположенных в узлах  $\{v_n, v_j\}$ , узел  $v_n$  считается источником, а узел  $v_j$  — приемником потока  $z_k$   $k$ -го вида, который передается из узла с номером  $n$  в узел с номером  $j$  для пары  $p_k$  с номером  $k$ . Для пары  $p_{k+E} \iff \{v_j, v_n\}$  узел  $v_j$  считается источником  $s_{k+E}$ , а узел  $v_n$  — приемником  $t_{k+E}$  для пары с номером  $p_{k+E}$ . Формируется  $R^+ = \{1, 2, 3, \dots, E, E + 1, \dots, 2E\}$  — список номеров смежных пар.

Пары  $p_k$  из подмножества  $P(R^+)$  называются смежными узлами-корреспондентами. Все остальные несмежные пары узлов-корреспондентов относятся к множеству  $P(R^-)$ :

$$P = P(R^+) \cup P(R^-); \quad P(R^+) \cap P(R^-) = \emptyset.$$

Введем обозначения:

$z_m$  — величина междузвонного потока  $m$ -го вида, который поступает в сеть из узла с номером  $s_m$  и покидает из узла с номером  $t_m$ ;

$S(v_n)$  — множество номеров исходящих дуг, по которым поток покидает узел  $v_n$ ;

$T(v_n)$  — множество номеров входящих дуг, по которым поток поступает в узел  $v_n$ .

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

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

$$n = \overline{1, N}, m = \overline{1, M}, x_{mi} \geq 0, z_m \geq 0.$$

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

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

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

В рамках данной модели пропускная способность ребер сети — вектор  $\mathbf{d}$  — трактуется как «ресурсное ограничение», а сумма дуговых потоков рассматривается как показатель использования «ресурсов» сети при передаче междузвонных потоков различных видов.

Для всех  $z_m$  и  $x_{mi}$ , удовлетворяющих условиям (1) и (2), вычисляются суммарные потоки:

$$y_m = \sum_{i=1}^{2E} x_{mi}, \quad m = \overline{1, M}. \quad (3)$$

Суммарный реберный поток  $y_m$  характеризует «нагрузку» на сеть при передаче междузвонного потока

величины  $z_m$  из узла-источника  $s_m$  в узел-приемник  $t_m$ . Величина  $y_m$  показывает, какой суммарный ресурс – пропускная способность сети – требуется для передачи межузлового потока  $z_m$ , а отношение  $w_m = y_m/z_m$ ,  $m = \overline{1, M}$ , показывает, какие ресурсы необходимы для передачи единичного потока  $m$ -го вида между узлами  $s_m$  и  $t_m$ .

Ограничения (1)–(3) задают подмножество допустимых значений компонент вектора межузловых потоков  $\mathbf{z} = (z_1, z_2, \dots, z_m, \dots, z_M)$ :

$$Z(\mathbf{d}) = \{\mathbf{z} \geq 0 \mid (\mathbf{z}, \mathbf{x}, \mathbf{y}) \text{ удовлетворяют (1)–(3)}\},$$

а все допустимые распределения ресурсов принадлежат подмножеству

$$Y(\mathbf{d}) = \{\mathbf{y} \geq 0 \mid (\mathbf{z}, \mathbf{x}, \mathbf{y}) \text{ удовлетворяют (1)–(3)}\}.$$

### 3 Метрические оценки предельных распределений

Для оценки функциональных возможностей системы рассматриваются допустимые распределения межузловых потоков при предельных нагрузках ребер сети.

В рамках данного модельного описания монопольным режимом называется способ управления, при котором все ресурсы сети используются для передачи потока одной выделенной пары узлов-корреспондентов  $p_a \in P(R^-)$ , а для всех остальных потоки полагаются равными нулю.

Предельно допустимый поток, который можно передать между фиксированной парой узлов-корреспондентов  $p_a$  в монопольном режиме, является решением стандартной, в данном случае однопродуктовой, задачи о максимальном потоке.

**Задача 1.** Найти  $z_a^0 = \max_{\langle z, x \rangle \in Z(d)} z_a$  при условии  $z_i = 0$ ,  $i = \overline{1, M}$ ,  $i \neq a$ .

При решении задачи 1 для пары  $p_a$  вычисляются: межузловой поток  $z_a^0$ ; дуговые потоки  $\{x_{ak}^0; x_{a(k+E)}^0\}$ ,  $k = \overline{1, E}$ ; суммарное значение реберного потока  $y_a^0 = \sum_{i=1}^{2E} x_{ai}^0$ .

Поток величины  $z_a^0$  является *максимальным потоком*, передаваемым в *монопольном режиме* для пары узлов-корреспондентов  $p_a$ ,  $p_a \in P(R^-)$ , в сети  $G(d)$ .

Задача 1 решается последовательно для всех  $p_m \in P(R^-)$ , вычисляются значения  $z_m^0(t)$ .

При проведении вычислительных экспериментов использовалась итерационная процедура для оценки функциональных возможностей системы

при передаче межузловых потоков по нескольким маршрутам. На предварительном этапе шага  $t$  в сети  $G(t)$  при заданных значениях пропускной способности ребер  $d_k(t)$  для каждой пары узлов-корреспондентов  $p_m \in P(R^-)$  определяется максимально допустимый однопродуктовый поток  $z_m^0(t)$ , соответствующие дуговые потоки  $(x_{mk}^0(t), x_{m(k+E)}^0(t))$ ,  $p_m \in P(R^-)$ , и коэффициенты нормировки  $\xi_m^0(t) = 1/z_m^0(t)$  для всех  $p_m \in P(R^-)$ , таких что  $z_m^0(t) > 0$ ,  $y_m^0(t) > 0$ . Коэффициенты  $\xi_m^0(t)$  используются для поиска текущих совместно допустимых квот на передачу потоков одновременно между всеми парами  $p_m \in P(R^-)$ .

**Задача 2.** Найти  $\alpha^*(t) = \max_{\alpha} \alpha$  при условиях

$$\alpha \sum_{m \in R^-} \xi_m^0 (x_{mk}^0(t) + x_{m(k+E)}^0(t)) \leq d_k(t), \quad k = \overline{1, E}.$$

На основании  $\alpha^*(t)$  вычисляются совместно допустимые дуговые потоки:

$$\begin{aligned} x_{mk}^*(t) &= \alpha^*(t) \xi_m^0(t) x_{mk}^0(t), \\ x_{m(k+E)}^*(t) &= \alpha^*(t) \xi_m^0(t) x_{m(k+E)}^0(t), \\ m &= \overline{1, M}, \quad k = \overline{1, E}, \end{aligned}$$

и остаточная пропускная способность ребер в сети  $G(t+1)$ :

$$\begin{aligned} d_k(t+1) &= d_k(t) - \sum_{m \in R^-} (x_{mk}^*(t) + x_{m(k+E)}^*(t)), \\ k &= \overline{1, E}, \quad p_m \in P(R^-). \end{aligned}$$

Формируется вектор допустимых межузловых потоков:

$$\begin{aligned} z_k^+(t) &= d_k(t+1), \quad p_k \in P(R^+), \quad k = \overline{1, E}; \\ z_m^-(t) &= \sum_{\tau=1}^t \alpha^*(\tau) \xi_m^0(\tau) z_m^0(\tau), \quad p_m \in P(R^-). \end{aligned}$$

По построению, на шаге  $t$  при передаче вектора межузлового потока  $\mathbf{z}(t) = \{\mathbf{z}^+(t), \mathbf{z}^-(t)\}$  достигается предельная нагрузка, и пропускная способность всех ребер сети используется полностью.

Точка с координатами  $\mathbf{z}(t)$  принадлежит множеству  $Z(d)$ .

Расстояние точки от начала координат определяется как норма соответствующего вектора:

$$\begin{aligned} \rho^+(t) &= \|\mathbf{z}^+(t)\| = \left[ \sum_{k=1} (\mathbf{z}^+(t))^2 \right]^{1/2}; \\ \rho^-(t) &= \|\mathbf{z}^-(t)\| = \left[ \sum_{p_m \in P(R^-)} (\mathbf{z}_m^-(t))^2 \right]^{1/2}. \end{aligned}$$

Если при выполнении шага  $(t + 1)$  окажется, что  $z_m^0(t + 1) = 0$  для всех  $p_m \in P(R^-)$ , то произойдет останов и сформируются массивы финальных данных:

$$z_m^-(T) = \sum_{\tau=1}^t \alpha^*(\tau) \xi_m^0(\tau) z_m^0(\tau), \quad p_m \in P(R^-),$$

$$z_k^+(T) = d_k(t + 1), \quad p_k \in P(R^+), \quad k = \overline{1, E}.$$

Вышеописанная вычислительная процедура далее обозначается как MFPL-процедура (от англ. *max-flow-peak-load*).

При проведении второй серии вычислительных экспериментов MFPL-процедура использовалась для оценки функциональных характеристик системы при *уравнительном* поэтапном распределении пропускной способности между всеми парами-корреспондентами.

При реализации MFPL-процедуры выполнение каждого шага разбивается на несколько этапов. На предварительном этапе шага  $t$  в сети  $G(t)$  при заданных значениях пропускной способности ребер  $d_k(t)$  для каждой пары узлов-корреспондентов  $p_m \in P(R^-)$  определяется максимально допустимый однопродуктовый поток  $z_m^0(t)$ , соответствующие дуговые потоки  $(x_{mk}^0(t), x_{m(k+E)}^0(t))$ ,  $p_m \in P(R^-)$ , и суммарная реберная нагрузка

$$y_m^0(t) = \sum_{k=1}^E (x_{mk}^0(t), x_{m(k+E)}^0(t)), \quad p_m \in P(R^-).$$

Для каждой пары  $p_m \in P(R^-)$  вычисляются коэффициенты нормировки  $\theta_m^0(t) = 1/y_m^0(t)$  для всех  $p_m \in P(R^-)$ , таких что  $z_m^0(t) > 0$ ,  $y_m^0(t) > 0$ . Коэффициенты  $\theta_m^0(t)$  используются для поиска совместно допустимых дуговых потоков для всех  $p_m \in P(R^-)$ .

**Задача 3.** Найти  $\beta^*(t) = \max_{\beta} \beta$  при условиях

$$\beta \sum_{p_m \in P(R^-)} \theta_m^0(x_{mk}^0(t) + x_{m(k+E)}^0(t)) \leq d_k(t), \quad k = \overline{1, E}.$$

С помощью  $\beta^*(t)$  (решения задачи 3) вычисляются текущие допустимые значения дуговых потоков:

$$x_{mk}^*(t) = \beta^*(t) \theta_m^0(t) x_{mk}^0(t),$$

$$x_{m(k+E)}^*(t) = \beta^*(t) \theta_m^0(t) x_{m(k+E)}^0(t), \quad k = \overline{1, E},$$

и реберных нагрузок при одновременной передаче межузловых потоков:

$$y_m^*(t) = \sum_{i=1}^E [x_{mi}^*(t) + x_{m(i+E)}^*(t)] =$$

$$= \frac{\beta^*(t)}{y_m^0(t)} \sum_{i=1}^E [x_{mi}^0(t) + x_{m(i+E)}^0(t)] = \beta^*(t),$$

$$p_m \in P(R^-).$$

Таким образом на каждом шаге определенная часть имеющегося ресурса (пропускной способности) делится строго поровну между всеми корреспондентами  $p_m \in P(R^-)$ , для которых существует путь передачи в  $G(t)$ .

Формируется вектор допустимых межузловых потоков:

$$z_k^{++}(t) = d_k(t + 1) =$$

$$= d_k(t) - \sum_{p_m \in P(R^-)} (x_{mk}^*(t) + x_{m(k+E)}^*(t)),$$

$$k = \overline{1, E}, \quad p_k \in P(R^+);$$

$$z_m^{(=)}(t) \triangleq \sum_{\tau=1}^t \beta^*(\tau) \theta_m^0(\tau) z_m^0(\tau), \quad p_m \in P(R^-).$$

Определяются расстояния:

$$\rho^{++}(t) = \|z^{++}(t)\| \triangleq \left[ \sum_{k=1}^E (d_k(t + 1))^2 \right]^{1/2};$$

$$\rho^{(=)}(t) = \|z^{(=)}(t)\| = \left[ \sum_{p_m \in P(R^-)} (z_m^{(=)}(t))^2 \right]^{1/2}.$$

Если на предварительном этапе на шаге  $(t + 1)$  окажется, что в сети  $G(t + 1)$  для всех  $p_m \in P(R^-)$  все значения  $z_m^0(t + 1) = 0$ , то произойдет останов и сформируются финальные массивы:

$$z_k^{(++)}(T) = d_k(t + 1), \quad p_k \in P(R^+), \quad k = \overline{1, E};$$

$$z_m^{(=)}(T) = \sum_{\tau=1}^{t+1} \beta^*(\tau) \theta_m^0(\tau) z_m^0(\tau), \quad p_m \in P(R^-).$$

## 4 Вычислительный эксперимент

Результаты вычислительных экспериментов, описанные ниже, служат продолжением исследований, начатых в [1]. Вычислительные эксперименты проводились на моделях сетевых систем, представленных на рис. 1 и 2. В каждой сети 69 узлов. Пропускные способности ребер – значения  $d_k$  – выбирались случайным образом из отрезка  $[900, 999]$  и совпадали для ребер, присутствующих в обеих сетях. В кольцевой сети пропускная способность каждого из добавленных ребер равнялась 900.

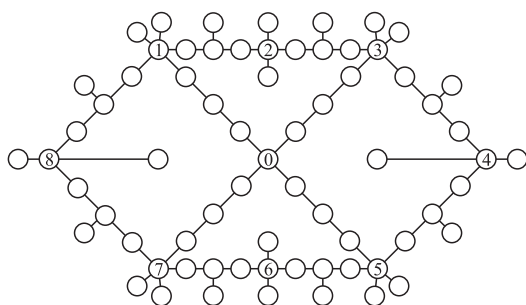


Рис. 1 Базовая сеть

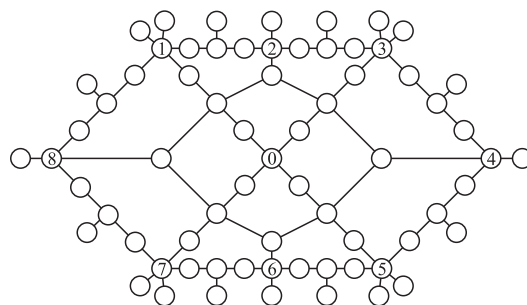


Рис. 2 Кольцевая сеть

Для базовой сети исходная сумма пропускных способностей:  $D^+(0) = 68\,256$ , а для кольцевой сети  $D^{++}(0) = 75\,456$ . Соответствующие значения  $\rho^+(0)$  и  $\rho^{++}(0)$  указаны в «нулевой» строке в табл. 1 и 2, где собраны результаты вычислительных экспе-

риментов. В ходе эксперимента при уравнительном распределении остаточных ресурсов соблюдается *равномерное* убывание остаточной пропускной способности и «длины» вектора  $\rho^+(t)$ . Однако между 44–46 итерациями для базовой и 50–52 для кольце-

Таблица 1 Базовая сеть

$t$	$\rho^-(t)$	$\rho^+(t)$	$d^+(t+1)$	$\rho^=(t)$	$\rho^{++}(t)$	$d^{++}(t+1)$
0	0	8048	68256	0	8048	68256
1	63	4182	26544	95	3881	24476
...	...	...	...	...	...	...
11	70	3975	21469	101	3707	20155
...	...	...	...	...	...	...
22	83	3861	19623	122	3586	18260
...	...	...	...	...	...	...
33	103	3778	18827	139	3522	17601
...	...	...	...	...	...	...
44	<b>190</b>	<b>3553</b>	<b>17503</b>	<b>203</b>	<b>3285</b>	<b>16201</b>
45	<b>1452</b>	<b>2166</b>	<b>7069</b>	<b>1376</b>	<b>2020</b>	<b>6584</b>
46	<b>1498</b>	<b>2158</b>	<b>6707</b>	<b>1388</b>	<b>2017</b>	<b>6483</b>
...	...	...	...	...	...	...
52	1535	2155	6413	1442	2011	6059

Таблица 2 Кольцевая сеть

$t$	$\rho^-(t)$	$\rho^+(t)$	$d^+(t+1)$	$\rho^=(t)$	$\rho^{++}(t)$	$d^{++}(t+1)$
0	0	8440	75456	0	8440	75456
1	68	5317	43038	92	5045	40716
...	...	...	...	...	...	...
11	95	3608	20459	124	3397	19080
...	...	...	...	...	...	...
22	101	3540	19530	130	3350	18338
...	...	...	...	...	...	...
33	135	3346	17561	154	3220	17003
...	...	...	...	...	...	...
44	234	3094	14881	269	2918	13848
...	...	...	...	...	...	...
50	<b>413</b>	<b>2770</b>	<b>12901</b>	<b>329</b>	<b>2792</b>	<b>13079</b>
51	<b>1040</b>	<b>2299</b>	<b>8801</b>	<b>334</b>	<b>2784</b>	<b>13034</b>
52	<b>1062</b>	<b>2297</b>	<b>8672</b>	<b>974</b>	<b>2262</b>	<b>8768</b>
...	...	...	...	...	...	...
55	1069	2297	8630	1010	2259	8553



вой сети наблюдается резкий скачок величин  $\rho^-(t)$ ,  $\rho^+(t)$  и  $d^{++}(t)$ .

На указанных шагах полностью используется пропускная способность ребер в центральной части сети. Сеть *распадается* на несвязные компоненты, и для 80% корреспондентов пропадают пути соединения, а остаточный ресурс распределяется поровну между оставшимися парами узлов.

Анализ результатов показал, что почти равные значения потоков достигаются для 80% корреспондентов и требуют 60%–70% ресурсов. Однако для 2% смежных пар узлов междуузловые потоки на два порядка выше медианных значений, а затраты пропускной способности составляют 20%–30%.

## 5 Заключение

Предложенный метод и проведенные вычислительные эксперименты показали, что уравнительное поэтапное распределение приводит к неравномерному распределению потоков для разных групп корреспондентов. Метрические оценки, полученные в ходе экспериментов, продемонстрировали *деформацию* множества достижимых потоков. В рамках модели предполагалось, что все корреспонденты равноправны, а потоки невзаимозаменяемы, однако при уравнительном предельном распределении смежные пары узлов оказывались в привилегированном положении при использовании остаточной пропускной способности. Пропускные способности ребер рассматривались как вектор ресурсов различных типов, которые распределяются между корреспондентами при передаче потоков различных видов. По построению, на каждом шаге норма вектора смежных междуузловых потоков численно равна модулю вектора остаточ-

ных пропускных способностей. Полученные метрические значения можно использовать для оценки функциональных возможностей сети в режиме предельной загрузки.

## Литература

1. Малашенко Ю. Е., Назарова И. А. Неоднородность распределения потоков при предельной загрузке многопользовательской сети // Известия РАН. Теория и системы управления, 2022. № 3. С. 81–96.
2. Luss H. Equitable resource allocation: Models, algorithms, and applications. — Hoboken, NJ, USA: John Wiley & Sons, 2012. 420 p.
3. Ogryczak W., Luss H., Pioro M., Nace D., Tomaszewski A. Fair optimization and networks: A survey // J. Appl. Math., 2014. Vol. 2014. Art. ID 612018. 25 p. doi: 10.1155/2014/612018.
4. Salimifard K., Bigharaz S. The multicommodity network flow problem: State of the art classification, applications, and solution methods // J. Oper. Res., 2020. Vol. 18. Iss. 3. P. 1–47.
5. Balakrishnan A., Li G., Mirchandani P. Optimal network design with end-to-end service requirements // Oper. Res., 2017. Vol. 65. Iss. 3. P. 729–750.
6. Nace D., Doan L. N., Klopfenstein O., Bashllari A. Max-min fairness in multicommodity flows // Comput. Oper. Res., 2008. Vol. 35. Iss. 2. P. 557–573.
7. Ros-Giralt J., Tsai W. K. A lexicographic optimization framework to the flow control problem // IEEE T. Inform. Theory, 2010. Vol. 56. Iss. 6. P. 2875–2886.
8. Baier G., Kohler E., Skutella M. The k-splittable flow problem // Algorithmica, 2005. Vol. 42. Iss. 3–4. P. 231–248.
9. Bialon P. A. Randomized rounding approach to a k-splittable multicommodity flow problem with lower path flow bounds affording solution quality guarantees // Telecommun. Syst., 2017. Vol. 64. Iss. 3. P. 525–542.

Поступила в редакцию 10.06.22

---



---

# SEQUENTIAL ANALYSIS AND METRIC ESTIMATES OF PEAK LOAD FLOWS IN THE MULTIUSER NETWORK

Yu. E. Malashenko

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

**Abstract:** The set of vectors of internodal flows in a multiuser communication network under peak load is analyzed. Within the framework of the multicommodity model, the throughput capacities of edges are considered as the components of a vector of resources of various types that are required for the transmission of various kinds of flows. When conducting computational experiments, at each iteration, the norms of vectors of jointly permissible internodal flows are calculated, during the transmission of which the capacity of all network edges is fully used.

The proposed method and computational experiments have shown that the equalizing phased distribution leads to an uneven distribution of flows for different groups of correspondents. Metric values obtained during experiments indicate deformation of the sets of accessible flows. Within the framework of the model, all correspondents are tantamount and the flows are noninterchangeable; however, in the case of an equalizing peak load distribution, adjacent pairs of nodes are in a privileged position when using residual capacity. The obtained metric values can be used to evaluate the functional characteristics of the transmission network in the finite capacity loading mode.

**Keywords:** multicommodity flow network model; set of achievable internodal flows; peak load distribution

**DOI:** 10.14357/19922264220306

## References

1. Malashenko, Yu. E., and I. A. Nazarova. 2022. Heterogeneous flow distribution at the peak load in the multiuser network. *J. Comput. Sys. Sc. Int.* 61:372–387.
2. Luss, H. 2012. *Equitable resource allocation: Models, algorithms, and applications*. Hoboken, NJ: John Wiley & Sons. 420 p.
3. Ogryczak, W., H. Luss, M. Pioro, D. Nace, and A. Tomaszewski. 2014. Fair optimization and networks: A survey. *J. Appl. Math.* 2014:612018. 25 p. doi: 10.1155/2014/612018.
4. Salimifard, K., and S. Bigharaz. 2020. The multicommodity network flow problem: State of the art classification, applications, and solution methods. *J. Oper. Res.* 18(3): 1–47.
5. Balakrishnan, A., G. Li, and P. Mirchandani. 2017. Optimal network design with end-to-end service requirements. *Oper. Res.* 65(3):729–750.
6. Nace, D., L. N. Doan, O. Klopfenstein, and A. Bashllari. 2008. Max-min fairness in multicommodity flows. *Comput. Oper. Res.* 35(2):557–573.
7. Ros-Giralt, J., and W. K. Tsai. 2010. A lexicographic optimization framework to the flow control problem. *IEEE T. Inform. Theory* 56(6):2875–2886.
8. Baier, G., E. Kohler, and M. Skutella. 2005. The k-splittable flow problem. *Algorithmica* 42(3-4):231–248.
9. Bialon, P. 2017. A randomized rounding approach to a k-splittable multicommodity flow problem with lower path flow bounds affording solution quality guarantees. *Telecommun. Syst.* 64(3):525–542.

*Received June 10, 2022*

## Contributor

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

# ВЫБОР МОДЕЛИ ПРИ ФАКТОРИЗАЦИИ МАТРИЦЫ ДАННЫХ С ПРОПУСКАМИ

М. П. Кривенко<sup>1</sup>

**Аннотация:** Работа посвящена задаче факторизации матрицы с отсутствующими компонентами в произведение двух матриц более низкого ранга. Исследовано влияние интенсивности пропусков на выбор модели факторизации. Рассматриваются два алгоритма оценивания параметров: альтернирующий наименьших квадратов (alternating least square, ALS) и Вайберга (Wiberg) — для двух моделей факторизации: со средними и без. Обосновывается, что в использовании модели со средними нет необходимости: она является частным случаем другой модели, приводит в отдельных случаях к неоднозначным решениям. При проведении экспериментов предпочтение отдано более устойчивому ALS-алгоритму. Продемонстрированы преимущества метода вставки перед случайным заполнением при начальных установках итерационных алгоритмов оценивания параметров модели. Выявлены причины негативных свойств существующей редакции алгоритма Вайберга. На основании экспериментов установлено: с ростом вероятности пропуска точность представления имеющихся данных увеличивается, что приводит к занижению истинного значения размерности модели.

**Ключевые слова:** понижающая ранг аппроксимация матрицы; сингулярное разложение; пропущенные данные; ALS-алгоритм; алгоритм Вайберга

**DOI:** 10.14357/19922264220307

## 1 Введение

Факторизация матриц данных хорошо зарекомендовала себя как метод снижения размерности. Множество его приложений включает обработку изображений, сжатие передаваемой информации, разведочный анализ данных, визуализацию, распознавание образов и прогнозирование временных рядов, а популярность обусловлена, во-первых, оптимальностью (обычно с точки зрения среднеквадратичной ошибки) линейной схемы сжатия совокупности высокоразмерных векторов в набор векторов меньшей размерности с возможностью последующей реконструкции, во-вторых, тем, что сжатие и восстановление — относительно простые операции, включающие лишь действия с матрицами.

Трудности при факторизации начинают проявляться при возникновении пропусков в данных, причем отбрасывание (игнорирование) или заполнение (вменение, приписывание) недостающих элементов данных с последующим применением обычных процедур анализа перестают эффективно работать при возрастании доли пропусков. Поэтому возрастает роль методов обработки наблюдаемых и пропущенных данных в совокупности, которые требуют разработки оригинальных процедур. Среди них как самостоятельное сформировалось направление низкоранговой аппроксимации матрицы данных с пропусками. С общей харак-

теристикой этой проблемы можно ознакомиться, например, в [1].

Можно выделить две базовые модели факторизации  $(m \times n)$ -матрицы наблюдений  $\mathbf{Y}$ :

- (1) представление ее при помощи  $\tilde{\mathbf{Y}} = \mathbf{U}\mathbf{V}^T$ , где  $\mathbf{U} —  $(m \times r)$ -матрица,  $\mathbf{V} —  $(n \times r)$ -матрица;$$
- (2) представление ее с учетом вектора средних при помощи  $\tilde{\mathbf{Y}}_\mu = \mathbf{U}\mathbf{V}^T + \mathbf{1}_m\boldsymbol{\mu}^T \equiv \mathbf{U}\mathbf{V}^T + \mathbf{M}$ , где в дополнение к  $\mathbf{U}$  и  $\mathbf{V}$  введен  $n$ -вектор  $\boldsymbol{\mu}$ ;  $\mathbf{1}_m$  — единичный  $m$ -вектор.

Каждая модель принимает конкретный вид при одном из значений структурного параметра  $r = 1, \dots, n$ . Наличие пропусков в данных будет отражаться в  $(m \times n)$ -матрице  $\mathbf{H}$ :

$$h_{ij} = \begin{cases} 1, & y_{ij} \text{ присутствует;} \\ 0, & y_{ij} \text{ пропущено.} \end{cases}$$

Для суммы ее элементов примем обозначение  $p$  (число присутствующих элементов матрицы наблюдений  $\mathbf{Y}$ ).

**Задача построения  $\tilde{\mathbf{Y}}$**  формулируется как минимизация целевой функции:

$$\varphi(\mathbf{U}, \mathbf{V}) = \left\| \mathbf{H} \odot (\mathbf{Y} - \tilde{\mathbf{Y}}) \right\|_F^2 \rightarrow \min_{\mathbf{U}, \mathbf{V}}.$$

Перепишем целевую функцию в более удобном для аналитических преобразований виде, исключив  $\mathbf{H}$ .

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

Построчная запись матрицы  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m]^T$ , где  $\mathbf{u}_i$  суть  $r$ -векторы, даст  $mr$ -вектор  $\mathbf{u} = [\mathbf{u}_1^T, \dots, \mathbf{u}_m^T]^T$ . Также для матрицы  $\mathbf{V}$  определим  $n(r+1)$ -вектор  $\mathbf{v}$ . В результате  $\varphi(\mathbf{U}, \mathbf{V})$  можно переписать как

$$\varphi(\mathbf{U}, \mathbf{V}) \equiv \varphi(\mathbf{u}, \mathbf{v}) = |\mathbf{F}\mathbf{u} - \mathbf{y}|^2 = |\mathbf{G}\mathbf{v} - \mathbf{y}|^2, \quad (1)$$

где  $p$ -вектор  $\mathbf{y}$  формируется из соответствующих элементов  $\mathbf{Y}$ , а  $(p \times mr)$ -матрица  $\mathbf{F}$  — из векторов  $\mathbf{v}_i$  и  $(p \times n(r+1))$ -матрица  $\mathbf{G}$  — из векторов  $\mathbf{u}_i$  согласно только значениям  $h_{ij} = 1$ . Смысл (1) в том, что параметры модели (векторы  $\mathbf{u}$  и  $\mathbf{v}$ ) разделены в двух эквивалентных представлениях целевой функции.

Для нахождения минимума (1) существует достаточно обширный набор алгоритмов: если в [1] он включает 5 основных процедур, то в [2] их уже 16. Для целей данной статьи были выбраны и реализованы два варианта: альтернирующий наименьших квадратов (ALS) — как наиболее очевидный алгоритм — и канонический, но постоянно подвергающийся доработке алгоритм Вайберга [3]. Они относятся к итерационным и подразумевают задание начального значения вектора  $\mathbf{v}$ , что дает конкретную матрицу  $\mathbf{F}$ . Далее согласно алгоритму ALS происходит пересчет по формулам:

$$\mathbf{u} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{y}, \quad (2)$$

что дает новое значение  $\mathbf{G}$ , и

$$\mathbf{v} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{y},$$

что завершает очередной шаг итерации.

Для полноты описания последовательности действий достаточно добавить критерий завершения итераций.

Последняя из опубликованных полная версия алгоритма Вайберга [4] имеет следующий вид.

1. Задать начальное значение  $\mathbf{v}$ .
2. Сформировать  $\mathbf{F}$  из  $\mathbf{v}$  и найти вектор  $\mathbf{u}$ , минимизирующий  $|\mathbf{F}\mathbf{u} - \mathbf{y}|^2$ , по формуле (2).
3. Завершить вычисления, если достигнута требуемая точность аппроксимации; в противном случае продолжить и выполнить шаг 4.
4. Сформировать  $\mathbf{G}$  из  $\mathbf{u}$  и найти вектор  $\Delta_v$ , минимизирующий  $|\mathbf{Q}\mathbf{G}\Delta_v - \mathbf{Q}\mathbf{y}|^2$  при  $\mathbf{Q} = \mathbf{I} - \mathbf{F}(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T$ , используя при этом формулу  $\Delta_v = (\mathbf{Q}\mathbf{G})^+ \mathbf{Q}\mathbf{y}$ . Обновить вектор  $\mathbf{v}$  на  $\mathbf{v} + \Delta_v$  и перейти к шагу 2.

Решение задачи минимизации на шаге 4 усложняется из-за того, что  $\mathbf{Q}\mathbf{G}$  всегда является матрицей неполного ранга. Это дает толчок для уточнения общих идей [3] и рождения вариантов под

общим наименованием алгоритма Вайберга. Согласно [4], перспективным вариантом оказывается использование псевдообращения Мура–Пенроуза матрицы  $\mathbf{Q}\mathbf{G}$ : если сингулярное разложение  $\mathbf{Q}\mathbf{G} = \mathbf{S}\mathbf{D}\mathbf{T}^T$ , то

$$(\mathbf{Q}\mathbf{G})^+ = \mathbf{T}\tilde{\mathbf{D}}^{-1}\mathbf{S}^T,$$

где  $\tilde{\mathbf{D}}^{-1} = \text{diag}(1/d_1, \dots, 1/d_q, 0, \dots)$  и  $d_1, \dots, d_q$  суть ненулевые сингулярные числа матрицы  $\mathbf{Q}\mathbf{G}$ . При этом надо выбирать  $q = (n-r)r$ .

В заключение надо уточнить два момента: задание начального значения  $\mathbf{v}$  и критерия завершения итераций. Относительно  $\mathbf{v}$  обычно предлагается только лишь его случайное заполнение. Но не следует забывать о возможностях более целенаправленных методов: если размерности задачи и интенсивность пропусков позволяют, можно привлечь один из методов обработки неполных данных (например, использование подмножества данных без пропусков или заполнение пропусков средними значениями наблюдаемых значений) с последующим нахождением сингулярного разложения теперь уже полной матрицы данных и формирование начального значения  $\mathbf{v}$  из правых сингулярных векторов. Задание критерия завершения в работах по алгоритмам матричной факторизации умалчивается; в данной работе контролировалось относительное изменение значений целевой функции на последовательных шагах итераций, которое должно стать меньше заданного значения Tol; при этом число шагов итерации не должно было превышать  $t_{\max}$ .

**Задача построения  $\tilde{\mathbf{Y}}_\mu$**  формулируется как минимизация по  $\mathbf{U}$ ,  $\mathbf{V}$  и  $\mu$  целевой функции:

$$\varphi(\mathbf{U}, \mathbf{V}, \mu) = \left\| \mathbf{H} \odot (\mathbf{Y} - \tilde{\mathbf{Y}}_\mu) \right\|_F^2 \rightarrow \min_{\mathbf{U}, \mathbf{V}, \mu}.$$

Шаги построения оценок параметров модели со средними подобны случаю  $\tilde{\mathbf{Y}}$ . Построчная запись матрицы  $\mathbf{U}$  остается прежней. Введем матрицу  $\hat{\mathbf{V}} = [\mathbf{V}, \mu]$ , для которой каждую строку обозначим как  $\hat{v}_j$ , и определим  $n(r+1)$ -вектор  $\hat{\mathbf{v}}$ . В результате для  $\varphi(\mathbf{U}, \mathbf{V}, \mu)$  получим

$$\begin{aligned} \varphi(\mathbf{U}, \mathbf{V}, \mu) \equiv \varphi(\mathbf{u}, \hat{\mathbf{v}}) &= |\mathbf{F}\mathbf{u} + (\mathbf{y} - \mu)|^2 = \\ &= \left| \hat{\mathbf{G}}\hat{\mathbf{v}} - \mathbf{y} \right|^2, \end{aligned}$$

где  $p$ -векторы  $\mathbf{y}$  и  $\mu$  формируются из соответствующих элементов  $\mathbf{Y}$  и  $\mathbf{M}$ , а  $(p \times mr)$ -матрица  $\mathbf{F}$  — из векторов  $\mathbf{v}_i$  и  $(p \times n(r+1))$ -матрица  $\hat{\mathbf{G}}$  — из векторов  $[\mathbf{u}_i, \mathbf{1}]$  согласно только значениям  $h_{ij} = 1$ . Алгоритмы оценивания параметров модели подобны случаю без средних. Соответственно при

реализации 4-го шага алгоритма Вайберга рекомендовано для матрицы  $\mathbf{Q}\hat{\mathbf{G}}$  выбирать  $q = (n-r)(r+1)$ .

Далее под алгоритмом Вайберга, согласно сложившимся традициям, будем понимать именно кратко описанный вариант с рекомендациями относительно рангов.

## 2 Эксперименты по сравнительному анализу моделей

Литература по матричной факторизации при наличии пропусков не содержит даже ориентировочных рекомендаций по выбору подходящей модели данных (со средними или без, какова сниженная размерность  $r$ ), ограничиваясь лишь демонстрацией работоспособности предлагаемых решений. В данной работе с помощью экспериментов делается попытка хотя бы частично прояснить ситуацию.

Сравнение моделей осуществлялось с помощью нормированной формы целевой функции  $\varphi_N(\cdot) = \varphi(\cdot)/p$ , что позволяло сопоставлять результаты оценивания параметров модели для различных значений интенсивности пропусков. Основу выводов о свойствах моделей составляли результаты статистических испытаний, что давало возможность проследить за качеством восстановления пропущенных наблюдений с помощью функции  $\psi_N(\cdot) = \psi(\cdot)/(mn-p)$ , где функция  $\psi(\cdot)$  аналогична  $\varphi(\cdot)$ , но вместо  $h_{ij}$  использует множители  $(h_{ij}+1) \bmod 2$ .

Эксперименты с выбором конкретного варианта алгоритма проводились по следующей схеме: задание объемов анализируемых данных  $m$  и  $n$ , назначение сниженной размерности  $r_M$ , генерирование случайных данных с фиксированными средним и дисперсией, получение из них матрицы данных  $\mathbf{Y}$  в подпространстве сниженной размерности, применение того или иного метода оценивания параметров модели для проведения сравнительного анализа результатов. Необходимость продлевать это многократно и при различных комбинациях условий привела к вынужденному ограничению значений основных параметров:  $m \leq 100$ ,  $n \leq 20$ ,  $r_M \leq 10$ . Для определенности рассматривалось ограничение  $n < m$ . Относительно итерационных алгоритмов были приняты значения  $\text{ToI} = 0,01$  и  $t_{\max} = 100$ .

**Выбор модели факторизации в зависимости от учета вектора средних.** Если при определенном значении  $r$  и использовании факторизации со средними получено представление  $\tilde{\mathbf{Y}}_\mu = \mathbf{U}\mathbf{V}^T + \mathbf{1}_m\boldsymbol{\mu}^T$ , то при  $r < n$  имеет место следующая комбинация:

$$\begin{aligned} \text{rank}(\mathbf{U}\mathbf{V}^T) &= r; \\ \text{rank}(\mathbf{1}_m\boldsymbol{\mu}^T) &= 1; \\ \text{rank}(\mathbf{U}\mathbf{V}^T + \mathbf{1}_m\boldsymbol{\mu}^T) &= r + 1. \end{aligned}$$

Тогда, используя сингулярное разложение матрицы

$$\mathbf{W} = (\mathbf{U}\mathbf{V}^T + \mathbf{1}_m\boldsymbol{\mu}^T) = \mathbf{U}_W\boldsymbol{\Sigma}_W\mathbf{V}_W^T,$$

получаем факторизацию

$$\tilde{\mathbf{Y}} = (\mathbf{U}_W\boldsymbol{\Sigma}_W)\mathbf{V}_W^T$$

без явного учета средних, но уже порядка  $(r+1)$ . При  $r = n$  картина меняется: если  $\text{rank}(\mathbf{U}\mathbf{V}^T) = r$  и  $\text{rank}(\mathbf{1}_m\boldsymbol{\mu}^T) = 1$  для любого вектора  $\boldsymbol{\mu}$ , то  $\text{rank}(\mathbf{U}\mathbf{V}^T + \mathbf{1}_m\boldsymbol{\mu}^T) = r$ . Тогда для любого отличного от  $\boldsymbol{\mu}$  вектора  $\tilde{\boldsymbol{\mu}}$  того же размера, что и  $\boldsymbol{\mu}$ , верно

$$\begin{aligned} \mathbf{U}\mathbf{V}^T + \mathbf{1}_m\boldsymbol{\mu}^T &= (\mathbf{U}\mathbf{V}^T + \mathbf{1}_m\boldsymbol{\mu}^T - \mathbf{1}_m\tilde{\boldsymbol{\mu}}^T) + \mathbf{1}_m\tilde{\boldsymbol{\mu}}^T = \\ &= \mathbf{U}_{\tilde{W}}\boldsymbol{\Sigma}_{\tilde{W}}\mathbf{V}_{\tilde{W}}^T + \mathbf{1}_m\tilde{\boldsymbol{\mu}}^T, \end{aligned}$$

где  $\mathbf{U}_{\tilde{W}}\boldsymbol{\Sigma}_{\tilde{W}}\mathbf{V}_{\tilde{W}}^T$  — сингулярное разложение  $\tilde{\mathbf{W}} = \mathbf{U}\mathbf{V}^T + \mathbf{1}_m(\boldsymbol{\mu}^T - \tilde{\boldsymbol{\mu}}^T)\boldsymbol{\mu}^T$ . Таким образом получена эквивалентная исходной факторизация, но с произвольными средними. Следовательно, целесообразно не учитывать в модели средние и ограничиться лишь аппроксимацией  $\tilde{\mathbf{Y}}$ , тем более что в этом случае алгоритм оценивания параметров несколько проще.

**Выбор способов реализации итерационного процесса** проводился на основании ряда экспериментов. Сравнение для обоих алгоритмов случайного задания начального значения  $v$  и заполнение пропусков доступными средними с последующей оценкой  $v$  позволили отдать предпочтение процедуре вставки из-за несколько более низких значений и более стабильного поведения целевых функций. Негативные свойства случайного заполнения вектора  $v$  более отчетливо выявлялись с помощью  $\psi_N$  при  $r \neq r_M$ . При этом неустойчивость (увеличение значения  $\varphi_N$  или  $\psi_N$  для некоторых значений  $r$  могло достигать нескольких порядков) в большей степени была характерна алгоритму Вайберга.

Преимущество ALS-процедуры перед принятой реализацией алгоритма Вайберга подтвердилось в ходе дополнительных более тщательных испытаний, так как чаще проявлялся спонтанный рост  $\varphi_N$  или  $\psi_N$  для отдельных значений  $r$  при практически одинаковых значениях целевых функций в типичных ситуациях. Кроме этого, реализация ALS-шагов проще и поэтому более доступна для алгоритмической оптимизации.



**Таблица 1** Значения отклонений эмпирических рангов от теоретических для матрицы **F**

$g$	$r$									
	1	2	3	4	5	6	7	8	9	10
0,2	0	0	0	0	0	4	5	26	6	0
0,4	0	0	0	9	16	22	11	5	0	0
0,6	0	1	8	38	12	6	1	0	0	0
0,8	0	11	8	2	0	0	0	0	0	0

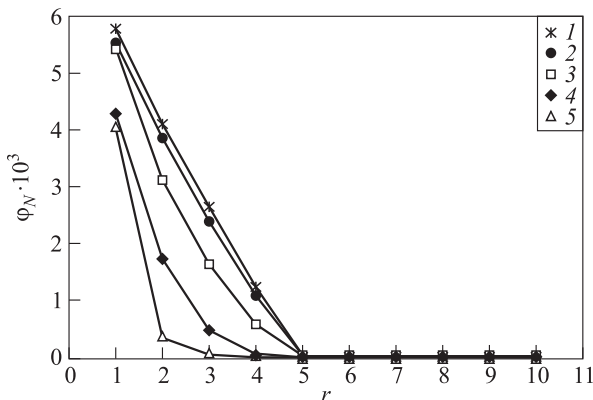
**Таблица 2** Значения отклонений эмпирических рангов от теоретических для матрицы **G**

$g$	$r$									
	1	2	3	4	5	6	7	8	9	10
0,2; 0,4; 0,6	0	0	0	0	0	0	0	0	0	0
0,8	0	11	8	2	0	1	2	1	16	11

Выяснение причин нестабильного поведения алгоритма Вайберга привело к необходимости дополнительных исследований поведения рангов матриц, участвующих в оценке параметров факторизации. Согласно [4], ключевым является предположение о том, что матрицы **U**, **V**, **F** и **G** — полного ранга. В случае модели без средних речь идет о следующих величинах:  $\text{rank } \mathbf{U} = \min(r, m)$ ;  $\text{rank } \mathbf{V} = \min(r, n)$ ;  $\text{rank } \mathbf{F} = \min(p, mr)$ ;  $\text{rank } \mathbf{G} = \min(p, nr)$ , будем называть их теоретическими. Проведенные эксперименты показали, что это не совсем так:

- для матриц **U** и **V** отклонений не выявлено;
- для матриц **F** и **G** экспериментальные значения рангов подчас не совпадают с теоретическими значения, причем реже для **G** и чаще для **F**.

В табл. 1 и 2 для случая  $m = 50, n = 10, r_M = 5$ , а также различных значений  $r$  и вероятности



**Рис. 1** Зависимости точности аппроксимации присутствующих данных  $\varphi_N$  от  $r$  при разных вероятностях пропуска: 1 —  $g = 0,0$ ; 2 —  $0,2$ ; 3 —  $0,4$ ; 4 —  $0,6$ ; 5 —  $g = 0,8$

пропуска  $g$  отражены результаты оценивания параметров факторизации. Для матрицы **F** значения отклонений эмпирических рангов от теоретических приведены в табл. 1. Для сравнения: в случае матрицы **G** несовпадения экспериментальных и теоретических значений обозначились лишь при большом числе пропусков (см. табл. 2).

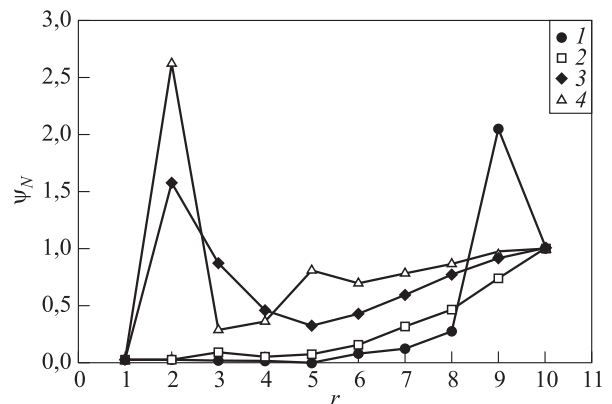
Важно обратить внимание, что во всех проведенных экспериментах отклонения в значениях рангов зависят от реализации случайных пропусков, но неизменны в ходе итераций.

Выявленные контрпримеры поведения рангов основополагающих матриц объясняют неустойчивое поведение привычных алгоритмов оценивания параметров модели и приводят к необходимости их коррекции. Если кратко, то она свелась к полному контролю рангов обрабатываемых матриц и использованию псевдообращения Мура–Пенроуза.

С точки зрения выбора размерности модели интерес представляют исследования зависимости точности аппроксимации данных от  $r$ , результаты которых приведены на рис. 1. Из них, в частности, следует:

- при умеренных значениях  $g < 0,5$  уверенно можно говорить о том, что увеличение  $r$  сверх значения  $t_M = 5$  существенного улучшения качества аппроксимации не дает;
- с увеличением  $g$  точность представления присутствующих элементов данных растет, незначительно изменяясь при небольшом смещении  $g$ ;
- увеличение точности аппроксимации приводит к занижению истинного значения  $r_M$ .

Точность восстановления пропущенных данных (рис. 2), характеризуемая с помощью  $\psi_N$ , падает



**Рис. 2** Зависимости точности восстановления пропущенных данных  $\psi_N$  от  $r$  при разных вероятностях пропуска: 1 —  $g = 0,2$ ; 2 —  $0,4$ ; 3 —  $0,6$ ; 4 —  $g = 0,8$

с ростом  $g$ . Проявляется это ярче при оценке параметров факторизации для больших значений  $r$ .

### 3 Эксперименты по обработке реальных данных

Для апробации разработанных методов факторизации на реальных данных использовалась матрица объект–признак, фигурирующая в исследованиях возможности прогнозировать химический состав мочевых камней у пациентов с уролитиазом по метаболическим показателям мочи и сыворотки крови (см. ссылку на проводимые работы и краткое описание данных в [5]). При этом объекты — это пациенты, а признаки включали, в частности, группу метаболических и антропологических признаков пациента.

Общая характеристика данных в обозначениях этой статьи:  $m = 471$ ,  $n = 17$ , количество пропусков — 475, тогда  $p = 7532$  и оценка  $g = 0,06$ . Далее понадобится размер так называемой комплектной (полной) матрицы данных  $\mathbf{Y}_C$ , состоящей из строк исходной матрицы без пропусков:  $m_C = 304$ .

Относительно небольшое число строк с пропусками (35,5%) позволило выделить комплектные данные и использовать их для следующих целей: формирование начального приближения для вектора  $v$  при обработке всей совокупности данных, обработка в качестве самостоятельного набора данных для построения статистических выводов относительно объекта исследований.

В этой связи важно упомянуть теорему Экарта–Янга (Eckart–Young, см. [6]), касающуюся приближения некоторой матрицы  $\mathbf{A}$  ранга  $r$  другой матрицей  $\mathbf{A}_s$  меньшего ранга  $s$ . Если потребовать, чтобы такое приближение было наилучшим в смысле нормы Фробениуса разности  $\mathbf{A}$  и  $\mathbf{A}_r$  (в более общем случае любой унитарно инвариантной нормы), то оказывается, что  $\mathbf{A}_s$  получается из сингулярного разложения  $\mathbf{A} = \mathbf{P}\mathbf{\Sigma}\mathbf{Q}^T$  по формуле  $\mathbf{A}_s = \mathbf{P}\mathbf{\Sigma}_s\mathbf{Q}^T$ , где  $\mathbf{\Sigma}_s$  — результат замены в  $\mathbf{\Sigma}$  нулями всех ее  $(s-r)$  наименьших диагональных элементов  $\sigma_{s+1}^2, \dots, \sigma_r^2$ , при этом

$$\|\mathbf{A} - \mathbf{A}_s\|_F^2 = \sigma_{s+1}^2 + \dots + \sigma_r^2.$$

Размеры матрицы  $m_C = 304$  позволяют легко найти сингулярное разложение и тем самым сформировать начальные приближения для различных  $r$  за счет усечения до первых  $r$  столбцов матрицы  $\mathbf{Q}$ . Таким образом комплектную матрицу можно эффективно использовать при оценивании качества аппроксимации представленных данных с пропусками.

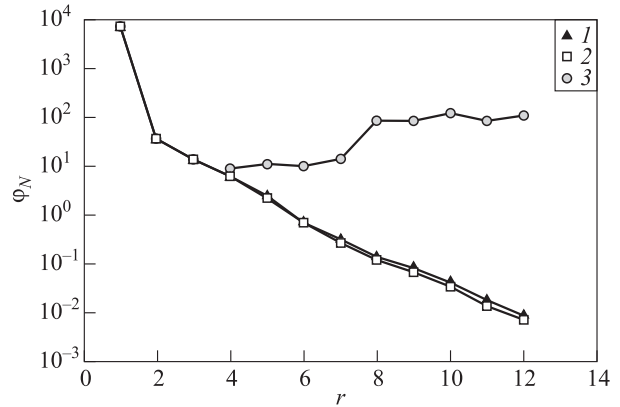


Рис. 3 Зависимость качества аппроксимации от размерности модели  $r$  в виде  $\varphi_N$  для комплектных данных (1) и для данных с пропусками (2), а также числа  $N_A$  аномальных восстановленных элементов матрицы данных (3)

В принципе, можно попробовать хотя бы грубо оценивать качество восстановления и тех данных, которых нет. В рассматриваемом случае анализируемая матрица данных содержит результаты измерений, принимающие неотрицательные значения, поэтому появление отрицательных элементов в  $\tilde{\mathbf{Y}}$  свидетельствует об их аномальности и может стать мерой качества восстановления пропущенных данных. Число подобных ситуаций обозначим как  $N_A$ .

Результаты экспериментов по оцениванию параметров факторизации и качества аппроксимации данных отражены на рис. 3. Из-за высокой временной сложности соответствующих алгоритмов при нахождении  $\varphi_N$  для данных с пропусками пришлось ограничиться диапазоном значений  $r = 1, \dots, 12$ . Наряду с этими значениями приведены числа  $N_A$ , для  $r = 1, 2, 3$  они нулевые. Анализ полученных результатов позволяет сделать следующие выводы:

- значения  $\varphi_N$  для  $\tilde{\mathbf{Y}}$  несколько лучше, чем  $\tilde{\mathbf{Y}}_C$ ; это говорит о том, что дополнение  $\mathbf{Y}_C$  данными, пусть и с пропусками, улучшает аппроксимацию присутствующих наблюдений;
- значения  $\varphi_N$  для  $\tilde{\mathbf{Y}}$  и  $\tilde{\mathbf{Y}}_C$  практически не отличаются, но вычислять  $\varphi_N$  при больших  $r$  становится затруднительным; поэтому в рассматриваемом случае можно обойтись сингулярным разложением комплектной матрицы и ее производными величинами, тем более что оно не существенно хуже;
- как указано в начальной части статьи, восстановление пропущенных данных ухудшается с ростом принятых значений  $r$ ; в частности, если появится необходимость использовать  $\tilde{\mathbf{Y}}$  в полной мере, то следует ожидать появления аномальных данных.

Обратим внимание на следующее. Если принять  $g = 0,06$  в качестве вероятности пропуска, то строка матрицы вообще не содержит пропусков с вероятностью  $0,941^{17}$ . Для случайных пропусков при  $m = 471$  число строк без пропусков распределено приблизительно нормально со средним 168 и стандартным отклонением 10, что никак не согласуется с наблюдаемым значением  $m_C = 304$ . Это ставит под сомнение возможность применения для анализа данных специфических вероятностных методов и широко используемого механизма случайных пропусков (Missing at Random, MAR) и побуждает отдать предпочтение общим методам факторизации матрицы данных.

## 4 Заключение

В работе используемые методы факторизации матриц данных с пропусками дополнены деталями, выявляющими и исправляющими ошибки в доступных публикациях. Они имеют важное прикладное значение и углубляют понимание особенностей моделей и методов аппроксимации данных. Эксперименты с синтетическими и реальными данными позволили обратить внимание на возможности использования комплектных подмножеств исходных данных, на необходимость учитывать негативные свойства восстановленных данных.

Наличие достаточно обширного набора алгоритмов факторизации матрицы данных с пропус-

ками в первую очередь говорит об отсутствии универсальной эффективной процедуры оценивания параметров модели. Поэтому так важно продолжать работу по исследованию существующих и созданию новых алгоритмов, активному внедрению решений, учитывающих специфические особенности обрабатываемых матриц.

## Литература

1. *Chen P.* Optimization algorithms on subspaces: Revisiting missing data problem in low-rank matrix // *Int. J. Comput. Vision*, 2008. Vol. 80. Iss. 1. P. 125–142.
2. *Hong J. H., Fitzgibbon A.* Secrets of matrix factorization: Approximations, numerics, manifold optimization and random restarts. [http://www.cv-foundation.org/openaccess/content\\_iccv\\_2015/papers/Hong\\_Secrets\\_of\\_Matrix\\_ICCV\\_2015\\_paper.pdf](http://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Hong_Secrets_of_Matrix_ICCV_2015_paper.pdf).
3. *Wiberg T.* Computation of principal components when data are missing // *2nd Symposium of Computational Statistics Proceedings*. — Wien: Physica-Verlag, 1976. P. 229–236.
4. *Okatani T., Deguchi K.* On the Wiberg algorithm for matrix factorization in the presence of missing components // *Int. J. Comput. Vision*, 2007. Vol. 72. Iss. 3. P. 329–337.
5. *Кривенко М. П.* Критерии значимости отбора признаков классификации // *Информатика и её применения*, 2016. Т. 10. Вып. 3. С. 32–40.
6. *Seber G. A. F.* A matrix handbook for statisticians. — Hoboken, NJ, USA: Wiley, 2008. 560 p.

Поступила в редакцию 15.04.22

---

# MODEL SELECTION FOR MATRIX FACTORIZATION WITH MISSING COMPONENTS

M. P. Krivenko

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

**Abstract:** The work is dedicated to the problem of factorizing a matrix with missing components into a product of two lower-rank matrices. The influence of the intensity of missing on the choice of the factorization model is studied. Two algorithms for parameter estimation are considered: alternating least squares (ALS) and Wiberg — for two factorization models: with and without means. It is substantiated that there is no need to use a model with averages: it is a special case of another model and, in some cases, leads to ambiguous solutions. During the experiments, the preference was given to a more stable ALS algorithm. The advantages of the insertion method over random filling in the initial settings of iterative algorithms for estimating model parameters are demonstrated. The reasons for the negative properties of the existing version of the Wiberg algorithm are revealed. Based on the experiments, it was found that with an increase in the probability of missing, the accuracy of the presentation of the available data increases which leads to an underestimation of the true value of the model dimension.

**Keywords:** lower-rank matrix approximation; singular value decomposition; missing data; ALS algorithm; Wiberg algorithm

**DOI:** 10.14357/19922264220307

## References

1. Chen, P. 2008. Optimization algorithms on subspaces: Revisiting missing data problem in low-rank matrix. *Int. J. Comput. Vision* 80(1):125–142.
2. Hong, J. H., and A. Fitzgibbon. 2015. Secrets of matrix factorization: Approximations, numerics, manifold optimization and random restarts. Available at: [http://www.cv-foundation.org/openaccess/content\\_iccv\\_2015/papers/Hong\\_Secrets\\_of\\_Matrix\\_ICCV\\_2015\\_paper.pdf](http://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Hong_Secrets_of_Matrix_ICCV_2015_paper.pdf) (accessed May 26, 2022).
3. Wiberg, T. 1976. Computation of principal components when data are missing. *2nd Symposium of Computational Statistics Proceedings*. Wien: Physica-Verlag. 229–236.
4. Okatani, T., and K. Deguchi. 2007. On the Wiberg algorithm for matrix factorization in the presence of missing components. *Int. J. Comput. Vision* 72(3):329–337.
5. Krivenko, M. P. 2016. Kriterii znachimosti otbora priznakov klassifikatsii [Significance tests of feature selection for classification]. *Informatika i ee Primeneniya — Inform. Appl.* 10(3):32–40.
6. Seber, G. A. F. 2008. *A matrix handbook for statisticians*. Hoboken, NJ, USA: Wiley. 560 p.

Received April 15, 2022

## Contributor

**Krivenko Michail P.** (b. 1946) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; [mkrivenko@ipiran.ru](mailto:mkrivenko@ipiran.ru)

# СРЕДОВЫЕ МОДЕЛИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ: ТЕОРЕТИЧЕСКИЕ ОСНОВАНИЯ ПОСТРОЕНИЯ\*

И. М. Зацман<sup>1</sup>

**Аннотация:** Рассматриваются варианты модели ИТО (information technology oriented model — ИТО model), которые используются при проектировании информационных технологий (ИТ) извлечения из текстов лингвистического и медицинского знания в рамках двух проектов по грантам РФФИ и ГФЕН (КНР). Варианты модели были созданы в рамках парадигмы деления предметной области информатики на среды различной природы. Цель статьи состоит в описании исходных данных и теоретических оснований обобщения этих вариантов. Эти основания планируется использовать для создания обобщенной модели ИТ, ее видов и частных случаев, образующих новый класс моделей, которые предлагается назвать средовыми. Основная идея обобщения состоит в распределении этапов проектируемой ИТ, их входов и выходов по выделенным средам предметной области информатики и границам между ними в интересах моделирования ИТ. На границах между средами размещаются сущности, неоднородные по своей природе (семиотические знаки на границе между ментальной и информационной средами, таблицы кодирования Unicode на границе между информационной и цифровой средами и пр.). Такие ИТ предлагается назвать полисредовыми. Средовые модели ориентированы, в частности, на решение актуальной проблемы создания таблиц унифицированного кодирования в базах знаний семантического веба значений многозначных слов при обработке больших объемов текстов (по аналогии с таблицами Unicode для кодирования литер и символов).

**Ключевые слова:** средовые модели; среды предметной области информатики; варианты модели ИТО; теоретические основания; полисредовые информационные технологии

**DOI:** 10.14357/19922264220308

## 1 Введение

Отчет с прогнозом будущего технологического развития, который был подготовлен под эгидой Всемирного экономического форума (Давос, Швейцария), содержит следующие положения [1]:

- компьютеры и развитие *цифровой* среды способствуют приумножению когнитивного и креативного потенциала человека (с. 3), сочетая возможности *ментальной* и других сред, в том числе за счет использования *нейрокомпьютерных* интерфейсов (brain-computer interfaces) (с. 8);
- органичная (seamless) интеграция физического и *цифрового* миров с помощью... программного обеспечения изменяет модели промышленного производства (с. 4);
- формирование институциональных структур общества в значительной степени определяют... доступными *технологиями* (с. 32).

В подготовке материалов для этого отчета приняли участие около 800 экспертов и руководите-

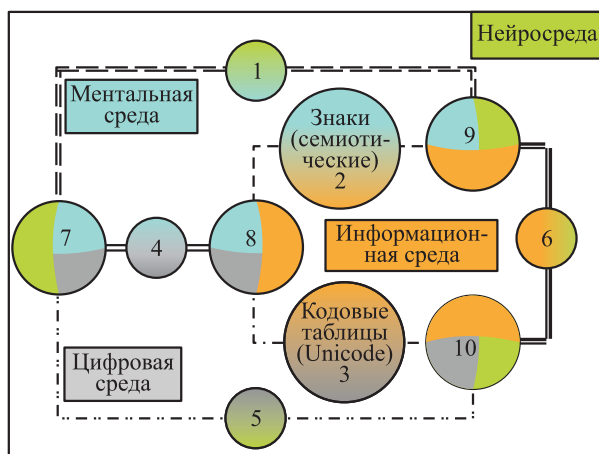
лей отрасли информационных и коммуникационных технологий. В отчете с результатами этого прогноза приведен перечень новых ИТ, которые определяют кардинальный характер преобразования общества и экономики, получившего название «Четвертая промышленная революция» [2]. Прогнозируемый характер преобразования во многом будет обусловлен теми ИТ, которые охватывают сущности, принадлежащие средам разной природы (далее — *полисредовые информационные технологии*, или ПИТ), включая нейросреду, ментальную, информационную, цифровую и другие среды предметной области информатики [3]. Точка зрения экспертов о формировании институциональных структур общества и определяющей роли в их развитии ИТ расширяет предметную область информатики и ставит перед ней задачу создания теоретических оснований проектирования и моделирования ПИТ, а также изучения и применения интерфейсов на границах между средами и в точках их соприкосновения.

Рисунок 1 иллюстрирует разнообразие интерфейсов, которые теоретически могут встретиться-

\* Исследование выполнено с использованием ЦКП «Информатика» ФИЦ ИУ РАН при поддержке РФФИ (проект 20-012-00166) и по гранту РФФИ и Государственного фонда естественных наук (ГФЕН) Китая № 21-57-53018.

<sup>1</sup> Федеральное исследовательское учреждение «Информатика и управление» Российской академии наук, izatsman@yandex.ru





**Рис. 1** Четыре среды разной природы, границы между ними и точки соприкосновения

ся при проектировании ПИТ, охватывающих сущности нейросреды, ментальной, информационной и цифровой сред. На границах между двумя смежными средами кружками условно обозначены интерфейсы 2-го порядка (они пронумерованы числами от 1 до 6), а в точках соприкосновения трех сред показаны интерфейсы 3-го порядка (7–10), которые, по определению в работе [3], обеспечивают связи между объектами (сущностями), принадлежащими трем средам (интерфейс 4-го порядка не показан).

В информационных системах широко используются кодовые таблицы, например Unicode, которые служат наиболее распространенным способом реализации интерфейса 2-го порядка № 3. В системах организации знания интерфейс 2-го порядка № 2 используется, например, для разрешения лексической неоднозначности и может быть реализован с помощью тезаурусов. Из интерфейсов 3-го порядка № 7 применяется в технологиях управления роботизированной рукой [3]. Интерфейс № 8 явно или неявно применяется при кодировании значений языковых единиц в информационных системах.

Цель статьи состоит в описании исходных данных и теоретических оснований построения средовых моделей ПИТ в рамках парадигмы деления предметной области информатики на среды. В разд. 2 рассмотрены исходные данные, с использованием которых описываются парадигма деления предметной области информатики на среды и затем, в разд. 3, теоретические основания построения средовых моделей. При описании теоретических

оснований их построения используются примеры частных случаев этих моделей, применяемых при проектировании ПИТ в двух проектах: извлечение из текстов новых значений модальных глаголов и медицинских терминов.

## 2 Исходные данные обобщения

При рассмотрении исходных данных используется характеристика информатики как фундаментальной науки из обзора [4]. На основе этого обзора была разработана Европейская стратегия преподавания информатики [5], которая предусматривает диадический подход к преподаванию: (1) информатики как самостоятельного предмета; (2) методов и средств информатики в предметных областях других наук [6]. Для данной статьи наибольший интерес представляет вторая позиция этого подхода. Она обуславливает существенное расширение предметной области информатики по сравнению с традиционными границами ее содержания<sup>1</sup>. Причина ее расширения — необходимость рассматривать и моделировать информационные процессы не только в искусственных, но и в живых и в социальных системах. Такой спектр информационных процессов будет охватывать объекты (сущности) нескольких сред предметной области информатики и, что важно отметить, *сред принципиально разной природы*.

В обзоре [4] его авторы предложили характеристику современной информатики, отметив, что она кардинально изменилась за последние 50 лет: «В то время как естественные науки определяются применительно к миру (world), в котором мы живем, информатику как научную дисциплину определить сложнее; у нее нет эмпирических основ, как у естественных наук; это нечто большее, чем формально-символьное мышление, как в математике; и это далеко не просто компиляция инженерных принципов и технологий».

Авторы этого обзора не претендуют на строгое и детальное определение предметной области информатики. Они предлагают косвенную ее характеристику, используя три отобранных ими ключевые работы [8–10], которые, с их точки зрения, дополняют друг друга при описании информатики как фундаментальной науки.

Первая публикация — это доклад Кристена Нюгорда (Университет Осло), который был прочитан им на Всемирном конгрессе IFIP (International Federation of Information Processing — Международная

<sup>1</sup> В 1967 г. А. Ньюэлл, А. Дж. Перлис и Х. А. Саймон написали в журнале «Science»: «Объектам и явлениям соответствуют те науки, которые их изучают. Появились компьютеры. Следовательно, назначение информатики — это изучение компьютеров» [7]. Такой подход к определению ее содержания доминировал долгие годы в информатике как компьютерной науке.

федерация по обработке информации) в Дублине в 1986 г. Второй раздел «The definition of the term “informatics”» этого доклада посвящен определению термина «информатика». В этом разделе он пишет: «Термин “computer science” следует заменить на “informatics”. Несколько лет назад выбор между этими двумя терминами казался, скорее всего, несущественным. Обсуждения терминологии часто считают праздными, но иногда они могут отражать ключевые различия во мнениях или по крайней мере акценты [в понимании терминов]. Сегодня, к сожалению, используется термин “computer science”. Этот термин имеет тенденцию поддерживать слишком узкое представление об информационных системах, которые в настоящее время связывают коллективы людей и интегрируют разнообразные средства обработки информации. При этом коллективы взаимодействуют как посредством межличностных связей людей, так и с использованием электронных каналов связи» [8].

После констатации важности четкого определения терминов «информатика» и «информация» для формирования адекватного представления об информационных системах Кристен Нюгорд предлагает свою дефиницию первого термина: «Информатика — это наука, которая имеет своей областью [исследований] информационные процессы и *связанные с ними феномены в артефактах, обществе и природе* (курсив мой — И. З.)» [8].

Далее дается ссылка на определение понятия «феномен» в словаре Webster 1960 г. («любой факт, обстоятельство или событие, которые сенсорно воспринимаются и которые могут быть научно описаны или оценены» [11]). Затем Кристен Нюгорд дает расширенное толкование этого понятия, которое кардинально отличается от его определения в словаре Webster 1960 г.: «Важными примерами феноменов являются: живые организмы, неодушевленные объекты (включая артефакты, такие как, например, машины), события и процессы (например, выполнение компьютерных программ). Мы также можем говорить о *когнитивных феноменах, происходящих в сознании людей*, в отличие от явных [сенсорно воспринимаемых] *феноменов, находящихся вне сознания* (курсив мой — И. З.)» [8].

Толкование Кристена Нюгорда с делением феноменов, *связанных с информационными процессами*, на когнитивные, происходящие в сознании людей и формирующие концепты знания человека, и сенсорно воспринимаемые (например, тексты), существующие вне сознания, является ключевым. Такое деление уже неявно, но существенно рас-

ширяет предметную область информатики и создает основание для нового ее позиционирования в системе современного научного знания. Имплитно такое толкование вводит в предметную область информатики сущности разной природы: ментальной и информационной (= знаковой сенсорно воспринимаемой в данной статье). Отметим, что толкование понятия «феномен», расширенное по объему его значения, позже было включено и в современный онлайн-словарь Merriam-Webster<sup>1</sup> (в новом определении удалена *сенсорная восприимчивость* для феноменов, поддающихся научному описанию).

Таким образом, смысловое содержание термина «информатика» по Нюгорду зависит только от двух понятий: «феномен» и «информационный процесс», — при этом в определении второго понятия как единого словосочетания Кристен Нюгорд не использует слово «информация». Далее он комментирует свою дефиницию термина «информатика» следующим образом: «Приведенная выше дефиниция информатики принципиально отличается от дефиниций, которые направлены на определение информатики как формальной дисциплины, сродни математике. Однако формальные задачи могут относиться к предметной области информатики — например, доказательство правильности программ... Формирование дефиниции термина «информатика» никоим образом не является тривиальным. Когда утверждается, что информатика является только формальной дисциплиной, то «согласно такому определению, влияние информационной системы на социум, в который она интегрирована, находится за пределами предметной области информатики. Кроме того, исследования того, как осуществляется обработка данных в организациях, также выходят за рамки информатики в этом узком смысле» [13]» [8].

Можно предположить, что выбор авторами обзора [4] доклада Кристена Нюгорда, опубликованного более 35 лет назад, был обусловлен, скорее всего, включением в предметную область широкого спектра феноменов, связанных с информационными процессами в *технических, живых и социальных системах*.

Из трех публикаций, выбранных авторами обзора [4], вторая — это книга Дэвида Харела (Научный институт Вейцмана), опубликованная в 1987 г., с изложением основ информатики, ряда важных и базовых ее тем с алгоритмической точки зрения. В книге подчеркивается фундаментальный характер этой науки. В ней анализируются три вида

<sup>1</sup> Современный онлайн-словарь Merriam-Webster приводит следующую дефиницию термина «феномен»: «Факт или событие, представляющие научный интерес, поддающиеся научному описанию и объяснению» [12], которая соответствует толкованию Кристена Нюгорда, но без явного деления феноменов в этом словаре на когнитивные и сенсорно воспринимаемые.

сложности проблем информатики: вычислительная, системная и когнитивная [9].

Для обобщения вариантов модели ИТО существенный интерес представляет сложность третьего вида — когнитивная, которая порождается использованием в ИТ сущностей, принадлежащих ментальной среде (концепты знания в сознании людей), или сущностей двойной природы, принадлежащих ее границам с нейросредой, цифровой, информационной и другими средами предметной области информатики. Этот вид сложности связан с задачами, которые не удается решить с применением формальных алгоритмов, так как часто они не поддаются точной постановке и четкому формальному определению этих сущностей. Примером такой задачи может служить специфицирование в базе знаний дефиниций новых значений модальных глаголов в контекстах, для которых уже известные базе знаний их значения оказываются нерелевантными контекстам [14].

Дэвид Харел называет такие задачи «псевдоалгоритмическими» проблемами (pseudo-algorithmic problems), подчеркивая в своей книге, что он следует формальной трактовке термина «алгоритм». Пристальное внимание к таким проблемам обусловлено желанием достичь эффекта от их компьютерного решения, подобного человеческому. Такое стремление ставит принципиально новые задачи — разработать системы, которые будут демонстрировать «человекоподобные интеллектуальные действия». Главный вопрос при решении подобных проблем состоит в том, чтобы описать сложное знание человека так, чтобы его представление поддавалось алгоритмической обработке, применению систем и средств информатики для их решения [9, с. 402].

Из трех публикаций, выбранных авторами обзора [4], третья работа, комплементарная с [8, 9], — это статья Питера Деннинга и Пола Розенблюма [10]. В ней предложен новый вариант позиционирования информатики в системе современного научного знания и обосновывается необходимость существенного расширения ее предметной области. Один из подходов к ее расширению и предлагаемая парадигма деления предметной области на среды различной природы приведены в работах [3, 15–17].

В работах [3, 16] по развитию варианта Деннинга и Розенблюма были описаны пять сред предметной области информатики и границы между ними. Отметим, что предлагаемое развитие этого варианта охватывает следующие ключевые идеи и Кристена Ньюгорда, и Дэвида Харела:

- деление феноменов, связанных с информационными процессами, на когнитивные, происходящие в сознании людей (ментальная среда), и сенсорно воспринимаемые (информационная среда), находящиеся вне сознания, что имплицитно уже вводит в предметную область информатики сущности разной природы [8];
- определение класса проблем когнитивной сложности, для решения которых с применением систем и средств информатики необходимо представить сложное знание так, чтобы оно поддавалось алгоритмической обработке в цифровой среде [9].

Рассмотренные работы Деннинга, Розенблюма, Ньюгорда, Харела, Касперсена<sup>1</sup> и его соавторов, обсуждающие вопросы позиционирования информатики в системе современного научного знания, были опубликованы в период с 1986 по 2019 гг. В настоящее время эти вопросы сохраняют свою актуальность и продолжается их активное обсуждение, в том числе с позиции компьютерного образования, качество которого существенно зависит от их решения [18].

Контекст рассмотренных работ дает возможность сформулировать следующие исходные позиции для разработки теоретических оснований создания обобщенной модели ПИТ, ее видов и частных случаев:

- деление предметной области информатики на среды разной природы;
- распределение объектов одной природы по средам (концепты в ментальной среде, компьютерные коды в цифровой и т. д. [3]);
- распределение объектов двойной природы по границам между двумя средами соответствующей природы (семиотические знаки на границе между ментальной и информационной средами, таблицы кодирования Unicode на границе между информационной и цифровой средами и т. д.);
- распределение объектов тройной природы (и более высокого порядка) в точках соприкосновения трех и большего числа сред, соответствующих природе объектов; например, таблица кодирования значений слов (см. интерфейс № 8 на рис. 1) принадлежит точке соприкосновения трех сред: ментальной (значения слов),

<sup>1</sup>С 2020 г. Майкл Касперсен возглавляет Европейский комитет по компетенциям в области информатики (European Informatics Competence Framework Committee). С 2022 г. он специальный советник по компьютерному образованию вице-президента Европейской комиссии Маргрет Вестагер.

информационной (слова как последовательно-сти литер) и цифровой (компьютерные коды значений слов) [3, 16]).

### 3 Теоретические основания

В данном разделе на базе рассмотренных исходных позиций опишем теоретические основания создания обобщенной модели ПИТ, используя ранее разработанные варианты модели ИТО [14, 19–21]. Первый вариант этой модели использовался при проектировании технологии извлечения новых значений немецких модальных глаголов из параллельных текстов художественных произведений с целью пополнения словарных статей немецко-русских словарей. Сначала для каждого этапа технологии были определены его вход и выход. Затем *этапы*, их *входы* и *выходы* были распределены по средам

и границам между ними (рис. 2). Описание применения этого варианта модели ИТО дано в [14]. Используя рис. 2 и исходные данные предыдущего раздела, перечислим предлагаемые теоретические основания создания обобщенной модели ПИТ:

- выделение в предметной области информатики тех сред разной природы, которые необходимы для моделирования проектируемой ПИТ, определяется целью ее создания (на рис. 2 показаны три среды, достаточные для моделирования технологии пополнения словарных статей);
- распределение моноприродных этапов ПИТ, их входов и выходов по средам определяется их природой<sup>1</sup>;
- распределение этапов двойной природы, их входов и выходов по границам между двумя средами определяется природой смежных сред<sup>2</sup>;

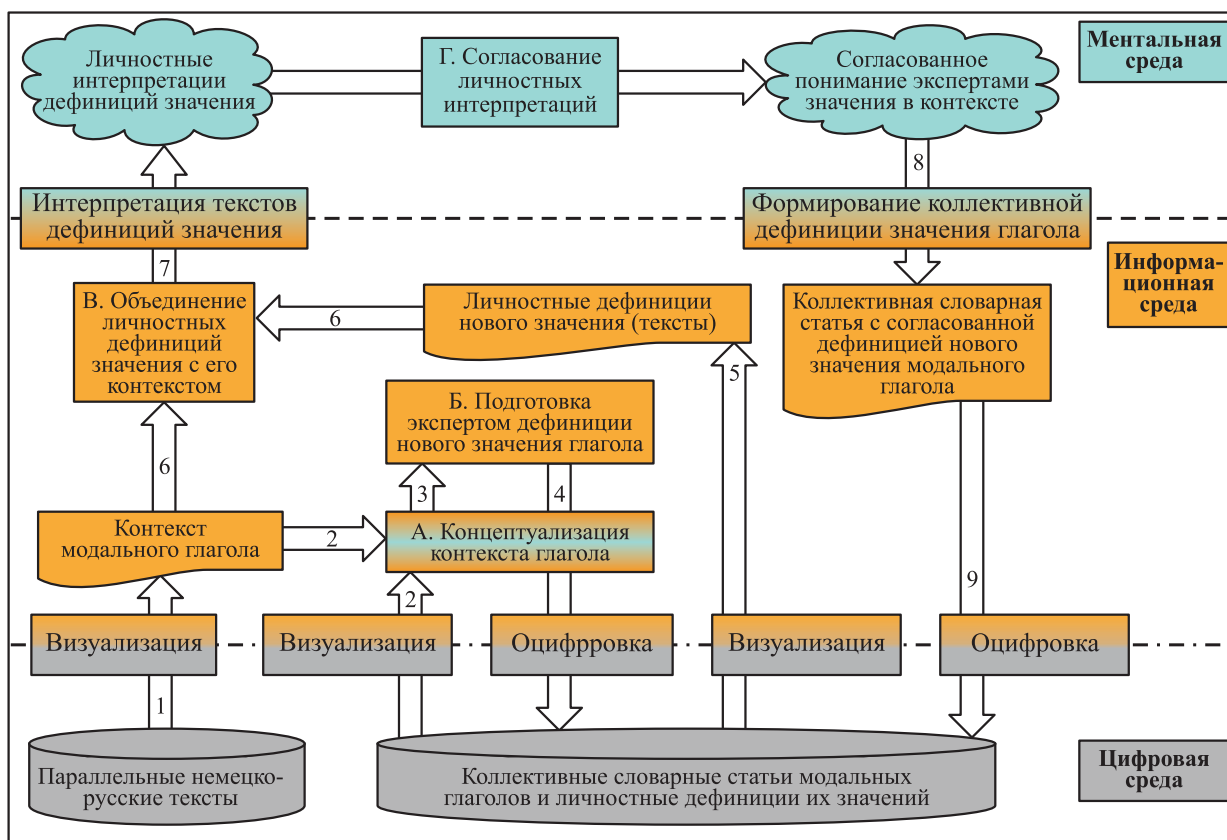


Рис. 2 Первый вариант модели ИТО

<sup>1</sup> На рис. 2 показаны три моноприродных этапа (Б, В и Г). Строго говоря, только этап В является таковым. При выполнении этапа Б эксперт использует свое понимание смысла глагола, сформированное им на этапе А, а при выполнении этапа Г эксперты могут обмениваться сообщениями в процессе согласования личностных интерпретаций. Однако для упрощения рис. 2 этапы Б–Г, их входы/выходы считаются моноприродными и обозначены одним цветом (оранжевым или бирюзовым).

<sup>2</sup> На рис. 2 показаны пять этапов двойной природы. Этапы «Интерпретация текстов дефиниций значения» и «Формирование коллективной дефиниции значения глагола» размещены на границе между ментальной и информационной средами. Этапы «Визуализация» и «Оцифровка» — между информационной и цифровой средами. Этап А размещен в информационной среде, чтобы упростить рис. 2. Его двойная природа условно обозначена оранжево-бирюзовой цветовой гаммой.



— распределение полиприродных этапов, их входов и выходов по точкам соприкосновения трех и более сред определяется природой сред, которые сходятся в этих точках (см. интерфейсы 7–10 на рис. 1).

В первом варианте модели ИТО использовались только интерфейсы второго порядка и реализующие их этапы двойной природы. Например, интерфейсу № 2 на рис. 1 соответствуют этапы «Интерпретация текстов дефиниций значения» и «Формирование коллективной дефиниции значения глагола», а интерфейсу № 3 — этапы «Визуализация» и «Оцифровка».

Второй вариант модели ИТО использовался при проектировании технологии извлечения новых терминов из текстов медицинских документов с целью формирования и актуализации в базе знаний терминологических портретов хронических болезней. Графическое представление этого варианта приведено на рис. 5 в [20]. Третий вариант модели ИТО, адаптированный для одновременного извлечения новых терминов и индикаторов настроений, названный ИТО-Sent, рассмотрен в работе [21]. Графическое представление этого варианта модели приведено на рис. 2 в [21]. Сопоставительное описание вариантов модели ИТО, приведенное в работах [17, 21], позволяет определить те их характеристики [22], которые планируется использовать при описании на последующих этапах обобщенной модели, ее видов и частных случаев, образующих новый класс средовых моделей ПИТ.

## 4 Заключение

Формирование класса средовых моделей ПИТ планируется осуществить на основе рассмотренных теоретических оснований. На первом этапе его формирования основное внимание будет уделено виду средовых моделей технологий кодирования значений многозначных слов в базах знаний. Модели этого вида ориентированы в первую очередь на решение актуальной проблемы создания (по аналогии с таблицами Unicode для кодирования литер и символов) таблиц одновременного кодирования значений многозначных языковых единиц (слов, коллокаций и пр.) и форм представления их значений в информационной среде в виде последовательностей литер (символов).

Сама идея одновременного кодирования значений слов и форм их представления не нова. Например, она используется в тезаурусе RuWordNet [23] и проекте анализа процесса возникновения новых значений слов английского языка [24]. Отличительная черта этого вида средовых моделей ПИТ

состоит в кодировании кроме самого значения степени и этапов его социализации. Это даст возможность кодировать в базах знаний значения новых слов, различая личностные, коллективные и конвенциональные их значения [14, 17, 19–21].

На последующих этапах обобщения предполагается описать еще три вида средовых моделей: информационно-технологически ориентированные (средовые модели ИТО — medium ITO models), познавательные-объяснительные (образовательные средовые модели), а также прогностические [25]. В заключение отметим, что частные случаи средовых моделей с ментальной, информационной и цифровой средами начали применяться в 2012 г. при проектировании и моделировании технологий формирования индикаторов мониторинга и оценки научно-исследовательских программ [26, 27].

## Литература

1. Deep shift — technology tipping points and societal impact // World Economic Forum. — Geneva, Switzerland, 2015. [http://www3.weforum.org/docs/WEF\\_GAC15\\_Technological\\_Tipping\\_Points\\_report\\_2015.pdf](http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf).
2. Шваб К. Четвертая промышленная революция / Пер. с англ. — М.: Эксмо, 2018. 288 с. (*Schwab K. The fourth industrial revolution.* — Geneva, Switzerland: World Economic Forum, 2016. 172 p.)
3. Зацман И. М. Интерфейсы третьего порядка в информатике // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 82–89.
4. Informatics education in Europe: Are we all in the same boat? — New York, NY, USA: ACM, 2017. Technical Report of the Committee on European Computing Education. 251 p. doi: 10.1145/3106077.
5. Caspersen M. E., Gal-Ezer J., McGettrick A., Nardelli E. Informatics for all: The strategy. — New York, NY, USA: ACM, 2018. 16 p.
6. Caspersen M. E., Gal-Ezer J., McGettrick A., Nardelli E. Informatics as a fundamental discipline for the 21st century // Commun. ACM, 2019. Vol. 62. No. 4. P. 58–63.
7. Newell A., Perlis A., Simon H. Computer science // Science, 1967. Vol. 157. No. 3795. P. 1373–1374.
8. Nygaard K. Program development as a social activity // 10th World Computer Congress Proceedings / Ed. H.-J. Kugler. — North Holland: Elsevier Science Pubs. B.V., IFIP, 1986. P. 189–198.
9. Harel D. Algorithmics — the spirit of computing. — Reading, MA, USA: Addison-Wesley, 1987. 514 p.
10. Denning P., Rosenbloom P. Computing: The fourth great domain of science // Commun. ACM, 2009. Vol. 52. No. 9. P. 27–29.
11. Webster's New World dictionary of the American language / Eds. D. B. Guralnik, J. H. Friend. — New York, NY, USA: The World Publishing Co., 1960. 1760 p.



12. Merriam-Webster's dictionary. Definition of phenomenon. <https://www.merriam-webster.com/dictionary/phenomenon>.
13. Nygaard K., Håndlykken P. The system development process — its setting, some problems and needs for methods // Software Engineering Environments Symposium Proceedings. — Amsterdam, 1981. P. 157–172.
14. Zatsman I. A model of goal-oriented knowledge discovery based on human–computer symbiosis // 16th Forum (International) on Knowledge Asset Dynamics Proceedings. — Matera, Italy: Arts for Business Institute, 2021. P. 297–312.
15. Зацман И. М. Методология обратимой генерализации в контексте классификации информационных трансформаций // Системы и средства информатики, 2018. Т. 28. № 2. С. 128–144.
16. Зацман И. М. Кодирование концептов в цифровой среде // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 97–106.
17. Зацман И. М. Модель процесса извлечения новых терминов и тональных слов из текстов // Системы и средства информатики, 2022. Т. 32. № 2. С. 115–127.
18. Tedre M., Pajunen J. Grand theories or design guidelines? Perspectives on the role of theory in computing education research // ACM T. Comput. Educ., 2021. doi: 10.1145/3487049.
19. Zatsman I. Finding and filling lacunas in linguistic typologies // 15th Forum (International) on Knowledge Asset Dynamics Proceedings. — Matera, Italy: Arts for Business Institute, 2020. P. 780–793.
20. Zatsman I., Khakimova A. New knowledge discovery for creating terminological profiles of diseases // 22nd European Conference on Knowledge Management Proceedings. — Reading, U.K.: Academic Publishing International Ltd., 2021. P. 837–846.
21. Зацман И. М., Золотарев О. В., Хакимова А. Х. Средовые модели извлечения из текста новых терминов и индикаторов настроений // Информатика и её применения, 2022. Т. 16. Вып. 2. С. 60–67.
22. Зацман И. М. Проблемно-ориентированная актуализация словарных статей двуязычных словарей и медицинской терминологии: сопоставительный анализ // Информатика и её применения, 2021. Т. 15. Вып. 1. С. 94–101.
23. Bolshina A., Loukachevitch N. All-words word sense disambiguation for Russian using automatically generated text collection // Cybernetics Information Technologies, 2020. Vol. 20. Iss. 4. P. 90–107.
24. Ramiro C., Srinivasan M., Malt B. C., Xu Y. Algorithms in the historical emergence of word senses // P. Natl. Acad. Sci. USA, 2018. Vol. 115. Iss. 10. P. 2323–2328.
25. Зацман И. М. Таблица интерфейсов информатики как информационно-компьютерной науки // Научно-техническая информация. Сер. I: Организация и методика информационной работы, 2014. № 11. С. 1–15.
26. Zatsman I. Tracing emerging meanings by computer: Semiotic framework // 13th European Conference on Knowledge Management Proceedings. — Reading, U.K.: Academic Publishing International Ltd., 2012. Vol. 2. P. 1298–1307.
27. Zatsman I. Denotatum-based models of knowledge creation for monitoring and evaluating R&D program implementation // 11th IEEE Conference (International) on Cognitive Informatics and Cognitive Computing Proceedings. — Los Alamitos, CA, USA: IEEE Computer Society Press, 2012. P. 27–34.

Поступила в редакцию 14.07.22

---

## INFORMATICS' MEDIUM MODELS OF INFORMATION TECHNOLOGY: THEORETICAL FOUNDATIONS FOR THEIR CREATING

I. M. Zatsman

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

**Abstract:** The variants of the information technology-oriented (ITO) model are considered which are used in the design of information technologies (IT) for discovering linguistic and medical knowledge from texts within the framework of two projects under the grants from the Russian Foundation for Basic Research (RFBR) and the Natural Science Foundation of China (NSFC). The variants of the ITO model were created within the paradigm of dividing the subject domain of informatics into media of different nature. The aim of the paper is to describe the initial data and the theoretical foundations for the generalization of the ITO model. These foundations are planned to be used to create a generalized model of IT, its types, and particular cases as constituents of the new class of models that are proposed to be called the informatics' medium ones. The principal idea of generalization is to distribute the stages of the IT being designed, their inputs, and outputs according to media of the subject domain of informatics and the boundaries between them in the interests of IT modeling. The boundaries locate entities of

multiple nature (semiotic signs on the boundary between the mental and information media, Unicode coding tables on the boundary between the information and digital media, etc.). Such IT is proposed to be called the multimedium one. Medium models are focused, first of all, on solving the topical problem of creating unified tables for coding meanings (= senses) of ambiguous words in semantic web knowledge bases when processing large volumes of texts (by analogy with Unicode tables for encoding letters and symbols).

**Keywords:** medium models in informatics; media of informatics subject domain; ITO model variants; theoretical foundations; multimedium information technologies

**DOI:** 10.14357/19922264220308

## Acknowledgments

The research was carried out using infrastructure of shared research facilities CKP “Informatics” of FRC CSC RAS. The reported study was funded by RFBR, project number 20-012-00166, and by RFBR and NSFC, project number 21-57-53018.

## References

1. Deep shift — technology tipping points and societal impact. 2015. *World Economic Forum*. Geneva, Switzerland. Available at: [http://www3.weforum.org/docs/WEF\\_GAC15\\_Technological\\_Tipping\\_Points\\_report\\_2015.pdf](http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf) (accessed July 27, 2022).
2. Schwab, K. 2016. *The fourth industrial revolution*. Geneva, Switzerland: World Economic Forum. 172 p.
3. 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.
4. The Committee on European Computing Education. 2017. Informatics education in Europe: Are we all in the same boat? New York, NY: ACM. Technical Report. 251 p. Available at: <https://dl.acm.org/citation.cfm?id=3106077> (accessed July 27, 2022).
5. Caspersen, M. E., J. Gal-Ezer, A. McGettrick, and E. Nardelli. 2018. *Informatics for all: The strategy*. New York, NY: ACM. 16 p.
6. Caspersen, M. E., J. Gal-Ezer, A. McGettrick, and E. Nardelli. 2019. Informatics as a fundamental discipline for the 21st century. *Commun. ACM* 62(4):58–63.
7. Newell, A., A. Perlis, and H. Simon. 1967. Computer science. *Science* 157(3795):1373–1374.
8. Nygaard, K. 1986. Program development as a social activity. *10th World Computer Congress Proceedings*. Ed. H.-J. Kugler. North Holland: Elsevier Science Pubs. B.V., IFIP. 189–198.
9. Harel, D. 1987. *Algorithmics — the spirit of computing*. Reading, MA: Addison-Wesley. 514 p.
10. Denning, P., and P. Rosenbloom. 2009. Computing: The fourth great domain of science. *Commun. ACM* 52(9):27–29.
11. Guralnik, D. B., and J. H. Friend, eds. 1960. *Webster's New World dictionary of the American language*. New York, NY: The World Publishing Co. 1760 p.
12. Merriam-Webster's dictionary. Definition of phenomenon. Available at: <https://www.merriam-webster.com/dictionary/phenomenon> (accessed July 27, 2022).
13. Nygaard, K., and P. Håndlykken. 1981. The system development process — its setting, some problems and needs for methods. *Software Engineering Environments Symposium Proceedings*. Amsterdam. 157–172.
14. Zatsman, I. 2021. A model of goal-oriented knowledge discovery based on human–computer symbiosis. *16th Forum (International) on Knowledge Asset Dynamics Proceedings*. Matera, Italy: Arts for Business Institute. 297–312.
15. Zatsman, I. 2018. Metodologiya obratimoy generalizatsii v kontekste klassifikatsii informatsionnykh transformatsiy [Methodology of reversible generalization in context of classification of information transformations]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(2):128–144.
16. Zatsman, I. 2019. Kodirovanie kontseptov v tsifrovoy srede [Digital encoding of concepts]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):97–106.
17. Zatsman, I. 2022. Model' protsessa izvlecheniya novykh terminov i tonal'nykh slov iz tekstov [A model of discovering novel terms and sentiments in texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 32(2):115–127.
18. Tedre, M., and J. Pajunen. 2021. Grand theories or design guidelines? Perspectives on the role of theory in computing education research. *ACM T. Comput. Educ.* doi: 10.1145/3487049.
19. Zatsman, I. 2020. Finding and filling lacunas in linguistic typologies. *15th Forum (International) on Knowledge Asset Dynamics Proceedings*. Matera, Italy: Arts for Business Institute. 780–793.
20. Zatsman, I., and A. Khakimova. 2021. New knowledge discovery for creating terminological profiles of diseases. *22nd European Conference on Knowledge Proceedings*. Reading, U.K.: Academic Publishing International Ltd. 837–846.
21. Zatsman, I., O. Zolotarev, and A. Khakimova. 2022. Sredovye modeli izvlecheniya iz teksta novykh terminov i indikatorov nastroeniya [Medium models for discovering novel terms and sentiments from texts]. *Informatika i ee Primeneniya — Inform. Appl.* 16(2):60–67.
22. Zatsman, I. 2021. Problemno-orientirovannaya aktualizatsiya slovarnykh statey dvuyazychnykh slovarey i me-

- ditsinskoy terminologii: sopostavitel'nyy analiz [Problem-oriented updating of dictionary entries of bilingual dictionaries and medical terminology: Comparative analysis]. *Informatika i ee Primeneniya — Inform. Appl.* 15(1):94–101.
23. Bolshina, A., and N. Loukachevitch. 2020. All-words word sense disambiguation for Russian using automatically generated text collection. *Cybernetics Information Technologies* 20(4):90–107.
24. Ramiro, C., M. Srinivasan, B. C. Malt, and Y. Xu. 2018. Algorithms in the historical emergence of word senses. *P. Natl. Acad. Sci. USA* 115(10):2323–2328.
25. Zatsman, I. 2014. A table of interfaces of informatics as computer and information science. *Scientific Technical Information Processing* 41(4):233–246.
26. Zatsman, I. 2012. Tracing emerging meanings by computer: Semiotic framework. *13th European Conference on Knowledge Management Proceedings*. Reading, U.K.: Academic Publishing International Ltd. 2:1298–1307.
27. Zatsman, I. 2012. Denotatum-based models of knowledge creation for monitoring and evaluating R&D program implementation. *11th IEEE Conference (International) on Cognitive Informatics and Cognitive Computing Proceedings*. Los Alamitos, CA: IEEE Computer Society Press. 27–34.

Received July 14, 2022

## Contributor

**Zatsman Igor M.** (b. 1952) — Doctor of Science in technology, head of department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; izatsman@yandex.ru

# ПЕРЕВОДЧЕСКИЙ АНАЛИЗ ТЕКСТА С ПРИМЕНЕНИЕМ ИНФОРМАЦИОННЫХ РЕСУРСОВ: РЕДУЦИРОВАНИЕ СПЕКТРА МОДЕЛЕЙ ПЕРЕВОДА В НАДКОРПУСНЫХ БАЗАХ ДАННЫХ\*

В. А. Нуриев<sup>1</sup>

**Аннотация:** Уточняется подход к редуцированию спектра моделей перевода в надкорпусных базах данных (НБД). Надкорпусные базы данных — это информационный ресурс широкого профиля: они обеспечивают решение задач, представляющих интерес для информатики, компьютерной лингвистики, медицины и т. д. В данной статье НБД рассматриваются с точки зрения корпусного переводоведения. Показано, как этот автоматизированный инструмент может применяться для «пристального и дальнего чтения» — подхода, выступающего за внедрение современных информационных ресурсов в деятельность литературного переводчика. Особое внимание уделено возможностям использования НБД для редуцирования спектра моделей перевода. При переводе ввиду синонимического потенциала естественных языков часто имеется некоторое множество относительно взаимозаменяемых по смыслу эквивалентов (слов, словосочетаний, синтаксических конструкций и т. д.). Выбирая один конечный эквивалент, переводчик стремится максимально сузить диапазон своего поиска. Цель статьи — уточнить подход к применению НБД, позволяющий сузить диапазон поиска релевантной модели перевода.

**Ключевые слова:** корпусное переводоведение; цифровые гуманитарные науки; компьютерный анализ текста; дальнее чтение; параллельные тексты; перевод; модели перевода; надкорпусная база данных; многовариантный выбор

**DOI:** 10.14357/19922264220309

## 1 Введение

Взаимодействие информатики и языкознания интенсифицировалось в последние несколько десятилетий и привело к оформлению целого научного направления, получившего название «corpus-based translation studies» (корпусное переводоведение) [1–5]. Оно занимается проблемой корпусных информационных ресурсов и возможностями их внедрения в переводческий процесс.

Электронные текстовые корпуса и базы данных, где хранится информация о естественных языках, призваны изменить подходы к переводу, возникшие в докорпусную эпоху (см., например, [6]). Новые информационные ресурсы открывают иные возможности для переводчиков (исследователей, писателей, преподавателей и т. д.), помогая изучать текст-объект в большем многообразии его уникальных характеристик<sup>2</sup> путем применения компьютеризированных инструментов.

Один из таких подходов предлагает английский ученый Р. Юдейл [8]. В нем сочетается традицион-

ное пристальное/внимательное чтение (close reading) с количественным анализом, куда вовлечены компьютерные инструменты и который обозначается как дальнее чтение (distant reading)<sup>3</sup>. Подход получил название «пристальное и дальнее чтение» (close and distant reading approach). Он не предполагает применения каких-либо средств автоматизации перевода, пытаясь при этом привлечь внимание к возможностям внедрения современных информационных технологий в практику художественного перевода.

Внимательное чтение понимается в привычном ключе: оно является неременным условием художественного перевода, выступая основой для стилистической оценки текста и интерпретации его коммуникативной интенции. Глубина внимания при чтении зависит от опыта переводчика, накопленного им знания о языках и культурах.

Вместе с тем в художественных текстах большой протяженности имеются языковые элементы, необходимые для понимания текста и его органи-

\* Работа выполнена в Федеральном исследовательском центре «Информатика и управление» Российской академии наук с использованием ЦКП «Информатика» ФИЦ ИУ РАН.

<sup>1</sup> Федеральний исследовательский центр «Информатика и управление» Российской академии наук, nurieff.v@gmail.com

<sup>2</sup> О характеристиках текста, подлежащего переводу, а также о многообразии факторов, определяющих качество переводного текста, см., например, [7, с. 96–169].

<sup>3</sup> Термин «distant reading» заимствован у Ф. Моретти; перевод термина как «дальнее чтение» предложен в русскоязычной версии книги [9].

зации, но труднодоступные для отслеживания человеком «вручную»: паттерны использования высокочастотных служебных слов, средняя величина предложения, принципы кластеризации слов и конструкций, повторяющиеся союзы и т. д. Дальнее чтение, сконцентрированное не на тексте целиком, а на конкретных его структурных параметрах и элементах, с помощью современных автоматизированных инструментов<sup>1</sup> способно обеспечить их анализ на фоне большого числа других текстов. Тем самым осуществляется «отдаление» переводчика от отдельно взятого текста. Оно выводит на подкрепленные количественными данными обобщения, помогая сделать заключения об устройстве текста, подлежащего переводу, и о том, каким должен быть конечный переводной текст: «Количественные данные могут указать на языковые феномены, о которых читатель не задумывается» [11, с. 389].

Примером информационного инструмента, помогающего осуществлять дальнейшее чтение, служат НБД (подробнее см. [12]), функционирующие на основе параллельных корпусов Национального корпуса русского языка (НКРЯ). Они разрабатываются с 2012 г. в Институте проблем информатики ФИЦ ИУ РАН [13].

Данная статья продолжает серию исследований о возможностях применения НБД в переводной деятельности. Ее основная цель — уточнить подход к редуцированию спектра моделей перевода<sup>2</sup>, которые регистрируются и хранятся в НБД и которые в своей деятельности применяет переводчик. Эти возможности иллюстрируются на примере НБД коннекторов<sup>3</sup> (см. разд. 3). Но прежде чем перейти к уточнению указанного подхода, следует уделить внимание некоторым свойствам НБД и дать краткое описание ее структуры (см. разд. 2).

## 2 Надкорпусная база данных: свойства и структурная схема

Наряду с лексикографическими изданиями и контрастивными грамматиками спектр моделей для перевода разных языковых единиц может регистрироваться и аннотироваться (описываться в соответствии с определенным набором рубрик) в электронных базах данных, например в НБД коннекторов<sup>4</sup>.

<sup>1</sup> О преимуществах современного тестового анализа с применением компьютеризированных инструментов см. высказывание пионера в области цифровых гуманитарных наук Р. Бузы [10, р. xxvi].

<sup>2</sup> О моделях перевода см. подробнее разд. 3.

<sup>3</sup> О НБД коннекторов см. [13, 14].

<sup>4</sup> Коннектор — языковая единица, функция которой состоит в выражении логико-семантических отношений (ЛСО) между двумя соединенными с ее помощью текстовыми фрагментами (подробнее см. [15–17]).

Надкорпусная база данных функционирует на основе параллельных текстов, и в ней можно интерактивно осуществить обратный переход от модели перевода к примерам ее использования. Для этого в НБД каждый пример перевода связан с соответствующей динамически сформированной моделью перевода, которая имеет гипертекстовую ссылку на список всех примеров ее реализации.

Аннотации изучаемых языковых единиц и их переводных соответствий наполняют надкорпусную часть НБД, предназначенную для формирования, хранения и поиска аннотаций. Они интегрированы с параллельными текстами, которые содержатся в корпусной части НБД, связанной с надкорпусной. Выделяются два основных класса информационных объектов, к которым обеспечивает доступ НБД:

- (1) аннотационные рубрики, аннотации, фрагменты параллельных текстов и другие данные, хранящиеся в таблицах НБД;
- (2) генерализуемые объекты (содержащиеся в НБД модели перевода коннекторов, сгруппированные по тому или иному параметру).

Два основных класса информационных объектов обуславливают два типа многоуровневых связей в НБД:

- (1) связи, задаваемые структурой НБД, т. е. связи между таблицами баз данных SQL-сервера;
- (2) вычисляемые связи, обеспечивающие многоуровневые переходы от аннотаций к генерализуемым объектам и обратно.

Подробнее о НБД и ее структурной схеме см. [18].

## 3 Редуцирование спектра моделей перевода (на материале коннектора «si» и его русских эквивалентов)

В данном разделе предлагается развитие подхода к редуцированию спектра моделей перевода в НБД. В основе этого подхода лежит представление о переводе как о ситуации многовариантного выбора, где переводчик постоянно оперирует некоторым набором относительно равноценных языковых



альтернатив для оформления переводного варианта. При продуцировании конечного текста переводчику нужно добиться максимально возможного редуцирования спектра моделей перевода, в идеале сузив его до единичного соответствия. Формирование такого подхода предполагает ответ на два главных вопроса:

1. Насколько детальной в НБД должна быть классификационная система рубрик-признаков при аннотировании переводимого текстового фрагмента, чтобы максимально возможно сузить множество альтернатив при выборе конечной модели перевода?
2. Можно ли сформировать универсальный набор рубрик, позволяющий редуцировать модели перевода разных языковых единиц, или система рубрик-признаков для каждого класса языковых единиц должна проходить специализацию и формироваться таргетированно?

Ранее эмпирические данные, на основе которых проводилась экспериментальная часть исследования, включали в себя хранящиеся в НБД коннекторов аннотации параллельных текстовых фрагментов, содержащих один из самых частотных<sup>1</sup> французских коннекторов «с'est-à-dire» и его переводные соответствия (204 аннотации). Источник текстового материала — французо-русский параллельный подкорпус<sup>2</sup> НКРЯ. Объем французо-русских параллельных текстов — 2 184 492 словоупотреблений (= токенов). Анализ этих эмпирических данных позволил сделать ряд выводов [19, с. 122–123]. На данном этапе они подлежат верификации и уточнению. Для этого привлекаются хранящиеся в НБД коннекторов аннотации параллельных текстовых фрагментов, содержащих другой частотный французский коннектор — «si» и его русские эквиваленты (всего 255 аннотации). Употребление «si» характеризуется относительной равномерностью: он реализуется в разножанровых текстах, принадлежащих разным авторам (О. де Бальзак, Ф. Бегбедер, Ж. Верн, Р. Госинни, Ж. Женетт, Ж. Кокто, П. Модигано, Г. де Мопассан, А. де Сент-Экзюпери). Для «si» зафиксировано большое многообразие моделей перевода (табл. 1, № 34), этот коннектор может сигнализировать о разных типах ЛСО (условные / условные с ограничительным оттенком, отношения пропозициональной причины, временные, уступительные, сопоставительные).

<sup>1</sup> В НБД для таких коннекторов зафиксировано наибольшее число аннотаций.

<sup>2</sup> Аннотации коннекторов в НБД формируются посредством обработки параллельных текстов, включающих исходные тексты (ИТ) и их переводы. Для настоящего эксперимента привлекались переводы, официально опубликованные, выполненные человеком-профессионалом.

На основе аннотаций в НБД для исследуемого коннектора был сформирован экспериментальный массив моделей перевода в виде кортежей <коннектор; вариант перевода>. Для наиболее продуктивной модели перевода кортеж имеет вид <si; если> (см. все варианты перевода в табл. 1). Генерализованная модель перевода имеет вид <si; {вариант перевода}>, где {вариант перевода} обозначает весь спектр таких вариантов во втором столбце табл. 1.

Затем было определено число аннотаций с различными сочетаниями рубрик, использованных для аннотирования текстовых фрагментов, где употреблен коннектор (табл. 2). Таблица 2 показы-

**Таблица 1** Варианты перевода французского коннектора «si»

№	Вариант перевода	Число аннотаций
1	если	123
2	если  то	56
3	если бы	21
4	если только	6
5	хотя	4
6	если и	4
7	если бы  то бы	4
8	только если	3
9	зего ( <i>коннектор не переведен</i> )	2
10	если и  то	2
11	если  значит	2
12	если и  то	2
13	же	2
14	даже если	2
15	а	2
16	но	2
17	ежели	1
18	если  значит	1
19	если бы только	1
20	если при этом	1
21	если и	1
22	едва  как	1
23	тогда	1
24	и	1
25	в то время как	1
26	когда	1
27	когда  то	1
28	поскольку  то	1
29	пусть и  во всяком случае	1
30	особенно	1
31	х_при N	1
32	xv_fin.Impératif+только	1
33	xv_fin.Impératif	1
34	x.gérondif passé	1
Всего аннотаций:		255

**Таблица 2** Число аннотаций с моделями перевода для «si» (первые два столбца) при указанном сочетании рубрик (третий столбец)

ИТ	Вариант перевода	Сочетание рубрик	Число аннотаций
si	если	CNT q p + начальная + с предикацией + условные	57
	если  то	CNT q p + начальная + с предикацией + условные	35
	если	p CNT q + начальная + с предикацией + условные	19
	если	CNT q p + SubCNT + начальная + с предикацией + условные	18
	если бы	CNT q p + начальная + с предикацией + условные	10
	—	Другие сочетания рубрик	116
Итого:			255

вает, какие рубрики характеризуют модель перевода в НБД, а также насколько продуктивна эта модель при том или ином их сочетании. Экспериментальные данные включают следующие рубрики: порядок следования связываемых коннектором текстовых фрагментов (p CNT q, CNT q p); встраивающий/встроенный коннектор (SubCNT / SuperCNT); маркирует часть предложения без предикации / с предикацией; позиция (начальная/не-начальная/конечная); ЛСО, устанавливаемые коннектором (условные и т. д.); маркирует вставку и т. д.

Релевантность приведенных рубрик определялась в ходе предыдущих корпусных контрастных исследований, где инструментом служила НБД коннекторов. Результаты некоторых исследований см. в [15, 17, 20]. Приведенные в табл. 2 рубрики, характеризующие коннектор и его контекст, были проанализированы на предмет их релевантности для выбора переводного варианта<sup>1</sup>. Следовало определить, позволяет ли имеющийся в НБД<sup>2</sup> набор рубрик сузить область многовариантного переводческого поиска, т. е. максимально редуцировать спектр моделей перевода, анализируя сочетания рубрик.

Согласно табл. 1, французский коннектор «si» имеет более чем одно переводное русскоязычное соответствие: для него в НБД зарегистрированы 34 варианта перевода. Чтобы переводчик мог редуцировать спектр представленных моделей перевода, указанных в табл. 2 рубрик оказалось недостаточно. Исходя из задач переводной деятельности, число рубрик необходимо увеличить. Этот вывод сделан на основе нескольких наблюдений:

- несмотря на явную продуктивность самого частотного русскоязычного эквивалента «si» — «если», число случаев его употребления не достигает 50% (48%) от общего числа употреблений (123 аннотации из 255), причем лишь в двух

переводах этот эквивалент занимает лидирующую позицию (у Н. Галь в «Маленьком принце» А. де Сент-Экзюпери — 92% и у Ю. Яхниной в «Утраченном мире» П. Модино — 81%), в остальных переводах на него приходится самое большее 50% употреблений, при этом «если» отличается универсальностью и востребовано в текстах любой стилистической и жанровой направленности;

- в сборнике научных статей Ж. Женетта «Фигуры» 33% случаев перевода «si» приходится на «если», что, по-видимому, мотивировано стремлением к синонимическому варьированию в подборе эквивалентов для оформления научного высказывания;
- вторым по продуктивности русскоязычным эквивалентом «si» является «если||то», на него приходится 22% случаев употребления (56 аннотаций из 255, зарегистрированных в НБД), для него не характерна такая универсальность, как для «если», это эквивалент с явно выраженной функционально-прагматической нагрузкой — он употребляется в высказываниях, имитирующих живую разговорную речь (диалогические/монологические реплики, внутренний монолог), где за счет частицы «то» нагнетается иллюкутивная сила (об этом подробнее см. [20]), а также в синтаксически насыщенных предложениях с развернутой сложной структурой, где «то» служит средством демаркации, помогая лучше обозначить границы разных синтаксических блоков (такие предложения прежде всего востребованы в нехудожественных или пограничных с художественными жанрами);
- для «si» в НБД зафиксированы 18 единичных русских эквивалентов, из них 14 не содержат

<sup>1</sup> Анализ проводился доктором филологических наук В. А. Нуриевым — переводчиком, имеющим опубликованные переводы с французского и английского языков.

<sup>2</sup> Набор рубрик первоначально формировался лингвистами, исходя из определенных задач, напрямую не связанных с переводческой деятельностью.

в качестве строительного элемента «если» (что свидетельствует об интерпретативной диффузности коннектора «si»), а некоторые имеют очевидные ограничения на употребление, например устаревшее «ежели» или характерное для письменной речи «едва||как»;

- при переводе текстов художественных жанров наблюдается тенденция к большей синтаксической вариативности, здесь реализуются такие варианты перевода «si», как частица «же» (2 вхождения), конструкции типа *x\_при N* (1 вхождение), *xv\_fin\_Impératif+только* (1 вхождение), *xv\_fin\_Impératif* (1 вхождение), *x\_gérondif passé* (1 вхождение) и т. д.

Эти наблюдения подтверждают сделанный на предыдущем этапе вывод о том, что при простановке рубрик, характеризующих коннектор и его переводной эквивалент в НБД, необходимо ввести дополнительную рубрику «художественный/нехудожественный текст», позволяющую редуцировать спектр моделей перевода.

## 4 Заключение

В заключение на основе приведенных выше наблюдений можно сделать два основополагающих вывода, имеющих принципиальное значение для редуцирования спектра моделей перевода в НБД при переводе.

1. Разнообразие в использовании моделей перевода во многом задается средовыми условиями. Некоторые переводные эквиваленты в явном виде тяготеют к реализации в текстах с определенными характеристиками, будь то тексты, имеющие усложненную синтаксическую организацию типа литературоведческих трудов, как в случае Ж. Женетта, или беллетризованные документально-дневниковые заметки, как в случае Ж. Кокто. Например, «если||то», указывающее на условные и сопоставительные ЛСО, употребляется в силу необходимости более рельефно отделить условие наступления какого-либо положения дел от самого положения дел, с помощью частицы «то» проводя более отчетливую границу между двумя синтаксическими частями в сложносоставных предложениях-высказываниях. Как уже говорилось выше, в диалогических репликах «если||то» употребляется с другой целью — чтобы придать сказанному большую выразительность, используя для этого частицу «то». «Ежели» имеет явную тенденцию употребляться там, где переводчик стремится стилистически «состарить»

текст. А употребление «поскольку||то» может отвечать задачам наращивания синонимического ряда в научном тексте (переводы Ж. Женетта).

2. Первый вывод можно дополнить вторым. В НБД при простановке рубрик, характеризующих текстовый фрагмент с коннектором, наряду с рубрикой «художественный/нехудожественный текст» необходимо ввести такие рубрики, как «дата создания оригинала» (для маркировки пригодных для старения текста переводных вариантов) и «речь, имитирующая живую / разговорный регистр / письменная речь» (для маркировки контекстов, имеющих форму диалога, внутреннего монолога или, наоборот, возможных только в письменной речи)<sup>1</sup>. О значимости таких рубрик говорят и дополнительные наблюдения, связанные с переводом на русский французских языковых единиц, функционирующих в живой разговорной речи, — междометий и обращений: существует явная асимметрия в употреблении этих классов речевых единиц во французском и русском языках. Таким образом, при переводе не коннекторов, а некоторых других языковых единиц рубрики, характеризующие принадлежность к устному/письменному регистру речи, будут не опциональны, а обязательны. Также целесообразно предусмотреть рубрики, направленные на нюансирование стилевой и жанровой амплитуды текстовых фрагментов. Таким образом, дальнейшая детализация классификационной системы рубрик при аннотировании в НБД необходима. Дополнительные рубрики позволят более эффективно сужать поле переводческого поиска, существенно редуцируя множество альтернатив при выборе моделей перевода обрабатываемой языковой единицы.

Насколько детальной и универсальной должна быть эта классификационная система, в настоящее время определить сложно. Чтобы ответить на этот вопрос, нужно применить предлагаемый подход к языковым единицам, отличным от коннекторов.

## Литература

1. Corpus-based translation studies: Research and applications / Eds. A. Kruger, K. Wallmach, J. Munday. — London: Bloomsbury Academic, 2011. 320 p.
2. New directions in corpus-based translation studies / Eds. C. Fantinuoli, F. Zanettin. — Berlin: Language Science Press, 2015. 168 p.

<sup>1</sup>Приводимые названия рубрик носят рабочий характер.

3. *Kaibao Hu*. Introducing corpus-based translation studies. — Heidelberg: Springer, 2016. 245 p.
4. New perspectives on corpus translation studies / Eds. V.X. Wang, L. Lim, D. Li. — Singapore: Springer, 2021. 318 p.
5. Extending the scope of corpus-based translation studies / Eds. S. Granger, M.-A. Lefer. — London: Bloomsbury Academic, 2022. 288 p.
6. *Ляшевская О. Н.* Корпусные инструменты в грамматических исследованиях русского языка. — М.: ЯСК: Рукописные памятники Древней Руси, 2016. 520 с.
7. *Bittner H.* Evaluating the evaluator: A novel perspective on translation quality assessment. — London, New York: Routledge, 2021. 296 p.
8. *Youdale R.* Using computers in the translation of literary style: Challenges and opportunities. — London, New York: Routledge, 2020. 242 p.
9. *Моретти Ф.* Дальнее чтение / Пер. с англ. — М.: Изд-во Института Гайдара, 2016. 352 с. ( *Moretti F.* Distant reading. — Verso, 2013. 224 p.)
10. *Busa R.* Foreword: Perspectives on the digital humanities // A companion to digital humanities / Eds. S. Schreibman, R. Siemens, J. Unsworth. — Malden, Oxford, Carlton: Blackwell, 2004. P. xvi–xxii.
11. *Mahlberg M.* Corpus stylistics // The Routledge handbook of stylistics / Ed. M. Burke. — London, New York: Routledge, 2014. P. 378–392.
12. *Егорова А. Ю., Зацман И. М., Мамонова О. С.* Надкорпусные базы данных в лингвистических проектах // Системы и средства информатики, 2019. Т. 29. № 3. С. 77–91.
13. *Зализняк А. А., Зацман И. М., Инькова О. Ю.* Надкорпусная база данных коннекторов: построение системы терминов // Информатика и её применения, 2017. Т. 11. Вып. 1. С. 100–108.
14. *Инькова О. Ю.* Надкорпусная база данных как инструмент формальной вариативности коннекторов // Компьютерная лингвистика и интеллектуальные технологии: по мат-лам ежегодной Междунар. конф. «Диалог». — М.: РГГУ, 2018. Вып. 17(24). С. 240–253.
15. Семантика коннекторов: контрастивное исследование / Под ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2018. 396 с.
16. Структура коннекторов и методы ее описания / Под ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2019. 316 с.
17. *Инькова-Манзотти О. Ю.* Коннекторы противопоставления во французском и русском языках: Сопоставительное исследование. — М.: Информэлектро, 2001. 429 с.
18. *Зацман И. М., Кружков М. Г.* Надкорпусная база данных коннекторов: развитие системы терминов проектирования // Системы и средства информатики, 2018. Т. 28. № 4. С. 156–167.
19. *Нурiev В. А., Зацман И. М.* Редуцирование спектра моделей перевода в надкорпусных базах данных // Информатика и её применения, 2020. Т. 14. Вып. 2. С. 119–126.
20. *Нурiev В.* Коннектор *раз... то* и модели его перевода на французский язык: между условием, причиной и следствием // Сьпоставительно езикознание (Сопоставительное языкознание / Contrastive Linguistics), 2017. Т. 42. № 4. С. 63–76.

Поступила в редакцию 14.07.22

---

## COMPUTER-ASSISTED TEXTUAL ANALYSIS IN TRANSLATION: REDUCING THE SPECTRUM OF TRANSLATION MODELS IN SUPRACORPORA DATABASES

V. A. Nuriev

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

**Abstract:** The paper refines the approach aimed at reducing the spectrum of translation models in supracorpora databases (SDBs). Being an information resource of broad potential application, SDBs can be used to research on problems in the field of information science, computer linguistics, medicine, etc. Here, SDBs are regarded from the perspective of the corpus-based translation studies. It is shown how this automated instrument can be applied in ‘close and distant reading’ — an approach that advocates the idea of using modern information resources in literary translation. The special focus is on opportunities that SDBs could offer for reducing the spectrum of translation models. Due to the synonymic potential, characteristic of natural languages, in translation, instead of the only possible solution, one has to choose between relatively interchangeable alternatives (words, collocations, syntactic constructions, etc.). Choosing the only one output equivalent, a translator seeks to narrow the choice set. Hence, the goal of the paper is to refine the approach that would allow using SDBs for narrowing the choice set of relevant translation models.



**Keywords:** corpus-based translation studies; digital humanities; computer-assisted textual analysis; distant reading; parallel texts; translation; translation models; supracorpora database; multiple choice

**DOI:** 10.14357/19922264220309

## 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

- Kruger, A., K. Wallmach, and J. Munday, eds. 2011. *Corpus-based translation studies: Research and applications*. London: Bloomsbury Academic. 320 p.
- Fantinuoli, C., and F. Zanettin, eds. 2015. *New directions in corpus-based translation studies*. Berlin: Language Science Press. 168 p.
- Hu, K. 2016. *Introducing corpus-based translation studies*. Heidelberg: Springer. 245 p.
- Wang, V. X., L. Lim, and D. Li, eds. 2021. *New perspectives on corpus translation studies*. Singapore: Springer. 318 p.
- Granger, S., and M.-A. Lefer, eds. 2022. *Extending the scope of corpus-based translation studies*. London: Bloomsbury Academic. 288 p.
- Lyashevskaya, O. N. 2016. *Korpusnye instrumenty v grammaticheskikh issledovaniyakh russkogo yazyka* [Corpus instruments in grammatical studies of Russian]. Moscow: YaSK: Rukopisnye pamyatniki Drevney Rusi. 520 p.
- Bittner, H. 2020. *Evaluating the evaluator: A novel perspective on translation quality assessment*. London, New York: Routledge. 296 p.
- Youdale, R. 2020. *Using computers in the translation of literary style: Challenges and opportunities*. London, New York: Routledge. 242 p.
- Moretti, F. 2013. *Distant reading*. Verso. 224 p.
- Busa, R. 2004. Foreword: Perspectives on the digital humanities. *A companion to digital humanities*. Eds. S. Schreibman, R. Siemens, and J. Unsworth. Malden, Oxford, Carlton: Blackwell. xvi–xxii.
- Mahlberg, M. 2014. Corpus stylistics. *The Routledge handbook of stylistics*. Ed. M. Burke. London, New York: Routledge. 378–392.
- Egorova, A. Yu., I. M. Zatsman, and O. S. Mamonova. 2019. Nadkorpussnye bazy dannykh v lingvisticheskikh proektakh [Supracorpora databases in linguistic projects]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(3):77–91.
- Zaloznyak, A. A., I. M. Zatsman, and O. Yu. In'kova. 2017. Nadkorpussnaya baza dannykh konnektorov: postroeniye sistemy terminov [Supracorpora database of connectives: Term system development]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):100–108.
- Inkova, O. Yu. 2018. Nadkorpussnaya baza dannykh kak instrument formal'noy variativnosti konnektorov [Supracorpora database as an instrument of the study of the formal variability of connectives]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: po mat-lam ezhegodnoy Mezhdunar. konf. "Dialog"* [Computer Linguistic and Intellectual Technologies: Conference (International) “Dialog” Proceedings]. Moscow. 17(24):240–253.
- In'kova, O. Yu., ed. 2018. *Semantika konnektorov: kontrastivnoe issledovanie* [Semantics of connectives: Contrastive study]. Moscow: TORUS PRESS. 396 p.
- Inkova, O. Yu., ed. 2019. *Struktura konnektorov i metody ee opisaniya* [Structure of connectives and methods of its description]. Moscow: TORUS PRESS. 316 p.
- In'kova-Manzotti, O. Yu. 2001. *Konnektory protivopostavleniya vo frantsuzskom i russkom yazykakh. Sopostavitel'noe issledovanie* [Connectors of opposition in French and Russian: A comparative study]. Moscow: Informelektro. 429 p.
- Zatsman, I. M., and M. G. Kruzhkov. 2018. Nadkorpussnaya baza dannykh konnektorov: razvitiye sistemy terminov proektirovaniya [Supracorpora database of connectives: Design-oriented evolution of the term system]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(4):156–167.
- Nuriev, V. A., and I. M. Zatsman. 2020. Redutsirovaniye spektra modeley perevoda v nadkorpussnykh bazakh dannykh [Reducing the spectrum of translation models in supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 14(2):119–126.
- Nuriev, V. 2017. Konnektor *raz... to* i modeli ego perevoda na frantsuzskiy yazyk: mezhdum usloviem, prichinoy i sledstviem [The connector *raz... to* and variants of rendering it into French: Between condition, cause, and effect]. *Sopostavitel'noe yazykoznanie* [Contrastive Linguistics] 42(4):63–76.

Received July 14, 2022

## Contributor

**Nuriev Vitaly A.** (b. 1980) — Doctor of Science in philology, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; nurieff.v@gmail.com



# ОБ ОДНОМ ПОДХОДЕ К ОЦЕНИВАНИЮ СКОРОСТИ СХОДИМОСТИ НЕСТАЦИОНАРНЫХ МАРКОВСКИХ МОДЕЛЕЙ СИСТЕМ ОБСЛУЖИВАНИЯ\*

И. А. Ковалёв<sup>1</sup>, Я. А. Сатин<sup>2</sup>, А. В. Сеницина<sup>3</sup>, А. И. Зейфман<sup>4</sup>

**Аннотация:** Рассмотрено преобразование прямой системы Колмогорова, позволяющее получать простые оценки скорости сходимости для марковских цепей (м.ц.) с непрерывным временем, описывающих системы массового обслуживания. Кроме того, используется понятие логарифмической нормы операторной функции и связанные с ней оценки нормы матрицы Коши. Полученные результаты позволяют оценивать скорость сходимости для новых классов моделей, у которых преобразованная матрица не является существенно неотрицательной и применение метода логарифмической нормы не гарантирует возможность получения оценок скорости сходимости. Ранее для таких ситуаций применялся достаточно трудоемкий общий метод неравенств. Сформулирована теорема о получении скорости сходимости при изменении интенсивностей матрицы, при сравнении с процессом рождения и гибели. В качестве примера исследована специальная нестационарная модель с групповым обслуживанием требований (обслуживание парами).

**Ключевые слова:** скорость сходимости; эргодичность; логарифмическая норма; системы массового обслуживания

**DOI:** 10.14357/19922264220310

## 1 Введение

Одна из важнейших проблем при исследовании марковских моделей систем массового обслуживания — построение их вероятностных характеристик. Точное их вычисление является достаточно трудной задачей даже для стационарных моделей, а в нестационарном случае ситуация становится гораздо более сложной, в связи с чем неизбежно применение аппроксимационных подходов с теми или иными гарантиями их точности. В частности, многие авторы используют методы дискретизации для построения переходных вероятностей неоднородных марковских цепей на конечных временных интервалах [1–5] или применяют предельные вероятностные характеристики [6].

Первый подход в принципе не гарантирует точность на произвольных временных интервалах, особенно в случае медленной сходимости, а второй требует умения вычислять эти предельные характеристики и, самое главное, наличия оценок, гарантирующих близость переходных характеристик

к предельным. Для их получения необходимо количественное исследование скорости сходимости для эргодических марковских цепей с непрерывным временем. Ранее (см., например, обзор в [7]) был развит подход, при котором прямая система Колмогорова

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

для вероятностей состояний цепи «подвергается» следующим стандартным преобразованиям: вначале исключается нулевое состояние, полагая

$$p_0(t) = 1 - \sum_{i \geq 1} p_i(t),$$

получаем из (1)

$$\frac{d\mathbf{z}}{dt} = B(t)\mathbf{z}(t) + \mathbf{f}(t),$$

затем изучается соответствующая «однородная» система  $d\mathbf{x}/dt = B(t)\mathbf{x}(t)$ , затем применяется трехчленное преобразование  $w = D\mathbf{x}$ , где

\* Исследования в разд. 4 выполнены А. В. Сенициной за счет гранта Российского научного фонда (проект 21-71-30011).

<sup>1</sup>Вологодский государственный университет; Московский центр фундаментальной и прикладной математики, kovalev.iv96@yandex.ru

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

<sup>3</sup>Ярославский государственный университет им. П. Г. Демидова, a\_korotysheva@mail.ru

<sup>4</sup>Вологодский государственный университет; Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Вологодский научный центр Российской академии наук; Московский центр фундаментальной и прикладной математики, a\_zeifman@mail.ru

$$D = \begin{pmatrix} d_1 & d_1 & d_1 & \cdots \\ 0 & d_2 & d_2 & \cdots \\ 0 & 0 & d_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

с положительными  $d_k$ , после чего изучается уже система

$$\frac{d\mathbf{w}}{dt} = DB(t)D^{-1}\mathbf{w}(t).$$

Если при этом матрица  $B_1(t) = DB(t)D^{-1}$  оказывается существенно неотрицательной, то метод логарифмической нормы (см. [7]) позволяет получить точные оценки для нормы решений этой системы, а значит, и для скорости сходимости исходной м. ц.

Если же ее матрица не является существенно неотрицательной, то применение метода логарифмической нормы не гарантирует возможность получения оценок скорости сходимости, для таких ситуаций был разработан более общий метод неравенств (см., например, [8–10]), применение которого очень трудоемко.

В настоящей работе описывается другой, существенно более простой подход, позволяющий в ряде случаев получать оценки скорости сходимости, а значит, и дающий возможность исследовать и строить предельные характеристики моделей.

Впервые этот подход, без подробностей, был описан в работе [11] и применен к исследованию одного класса суперкомпьютерных систем в [12].

## 2 Основные понятия и описание подхода

Пусть  $X(t)$  — вообще говоря, неоднородная марковская цепь с непрерывным временем  $t \geq 0$  и не более чем счетным пространством состояний. Обозначим через

$$\mathbf{p}(t) = (p_0(t), p_1(t), \dots)^T$$

ее вектор вероятностей состояний. Положим

$$a_{ii}(t) = - \sum_{j \neq i} a_{ji}(t)$$

и рассмотрим транспонированную матрицу интенсивностей

$$A(t) = \begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} & \cdots \\ a_{10} & a_{11} & a_{12} & a_{13} & \cdots \\ a_{20} & a_{21} & a_{22} & a_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

предполагая, что все  $a_{ij}(t)$  при  $i \neq j$  локально интегрируемы на  $[0, \infty)$  и неотрицательны при всех  $t \geq 0$ , а кроме того, выполнено условие ограниченности

$$\sup_i |a_{ii}(t)| \leq L < \infty$$

почти для всех  $t \geq 0$ .

Обозначим через  $\|\cdot\|$   $l_1$ -норму вектора,  $\|\mathbf{x}\| = \sum |x_i|$ ,  $\|H\| = \sup_j \sum_i |h_{ij}|$ , если  $H = (h_{ij})_{i,j=0}^\infty$ , и обозначим через  $\Omega$  множество всех векторов из  $l_1$  с неотрицательными координатами и единичной нормой. Тогда имеем

$$\|A(t)\| = 2 \sup_k |a_{kk}(t)| \leq 2L$$

почти для всех  $t \geq 0$ .

Можем рассматривать прямую систему Колмогорова как дифференциальное уравнение в пространстве последовательностей  $l_1$ , где  $A(t)$  является ограниченным почти для всех  $t \geq 0$  линейным оператором из  $l_1$  в себя.

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

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

для любых начальных условий  $\mathbf{p}^1(0) = \mathbf{p}^1 \in \Omega$  и  $\mathbf{p}^2(0) = \mathbf{p}^2 \in \Omega$ .

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

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

а для оператора Коши соответствующего дифференциального уравнения  $dx/dt = H(t)x$  справедлива оценка

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

Пусть  $\mathbf{p}^*(t)$  и  $\mathbf{p}^{**}(t)$  являются решениями прямой системы Колмогорова (1) с соответствующими (различными) начальными условиями  $\mathbf{p}^*(0) \in \Omega$  и  $\mathbf{p}^{**}(0) \in \Omega$ , тогда их разность  $\mathbf{z}(t) = \mathbf{p}^*(t) - \mathbf{p}^{**}(t)$  также является решением системы (1), но при этом сумма всех координат вектора  $\mathbf{z}(t)$  равна нулю при всех  $t \geq 0$ ,  $\sum_{i \geq 0} z_i(t) = 0$ .

Возьмем теперь положительное число  $c$  и вычтем из правой части уравнения  $z'_0(t) = \sum_{j \geq 0} a_{0j}(t)z_j$  выражение  $c \sum_{j \geq 0} z_j = 0$ . Запишем получившуюся при этом систему в виде

$$\frac{dz}{dt} = W(t)\mathbf{z}(t), \quad t \geq 0, \quad (2)$$

где  $W(t) = A(t) - C$  и матрица  $C$  имеет вид:

$$\begin{pmatrix} c & c & c & c & \dots \\ 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (3)$$

Для получения простейших оценок скорости сходимости можно применять и исследовать систему (2).

### 3 Оценки скорости сходимости

Для получения нужных свойств и оценок потребуются некоторые вспомогательные «взвешенные» нормы. Рассмотрим последовательность положительных чисел  $\{d_i\}$  такую, что  $d = \inf d_i > 0$ , диагональную матрицу  $D = \text{diag}(d_0, d_1, d_2, \dots)$  и пространство последовательностей  $l_{1D} = \{z / \|z\|_{1D} = \|Dz\|_1 < \infty\}$ .

Положим

$$\gamma_*(t) = \inf_i \left( |w_{ii}(t)| - \sum_{j \neq i} \frac{d_j}{d_i} |w_{ji}(t)| \right).$$

Рассмотрим (2) как дифференциальное уравнение в пространстве последовательностей  $l_{1D}$ . Тогда

$$\begin{aligned} \|W(t)\|_{1D} &= \|DW(t)D^{-1}\| = \\ &= \sup_i \left( |w_{ii}(t)| + \sum_{j \neq i} \frac{d_j}{d_i} |w_{ji}(t)| \right) = \\ &= \sup_i \left( 2|w_{ii}(t)| + \sum_{j \neq i} \frac{d_j}{d_i} |w_{ji}(t)| - |w_{ii}(t)| \right) \leq \\ &\leq 2 \sup_i |w_{ii}(t)| - \gamma_*(t) \leq 2L - \gamma_*(t); \end{aligned}$$

следовательно, операторная функция  $W(t)$  ограничена в пространстве  $l_{1D}$ . Получаем

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

Далее оценим логарифмическую норму оператора  $DW(t)D^{-1}$ . Обозначим через  $-\alpha_k(t)$  сумму всех элементов  $k$ -го столбца матрицы  $DW(t)D^{-1}$ ,

тогда в предположении, что почти при всех  $t \geq 0$   $c \leq a_{01}(t)$ , получаем

$$\alpha_k(t) = \begin{cases} \sum_{i>0} \left(1 - \frac{d_i}{d_0}\right) a_{i0} + c, & k = 0; \\ \sum_{i \neq 1} \left(1 - \frac{d_i}{d_1}\right) a_{i1} + c \frac{d_0}{d_1}, & k = 1; \\ \sum_{i \neq k} \left(1 - \frac{d_i}{d_k}\right) a_{ik} - c \frac{d_0}{d_k}, & k > 1 \end{cases} \quad (4)$$

и  $\gamma_*(t) = \inf_k \alpha_k$ , а значит, справедливо следующее утверждение.

**Теорема 1.** Пусть найдутся число  $c > 0$ ,  $c \leq a_{01}(t)$ , а также последовательность положительных чисел  $\{d_i\}$ , для которой  $d = \inf d_i > 0$  такие, что

$$\int_0^\infty \gamma_*(t) dt = +\infty.$$

Тогда  $X(t)$  слабо эргодична, причем

$$\begin{aligned} \|\mathbf{p}^1(t) - \mathbf{p}^2(t)\|_{1D} &\leq \\ &\leq e^{-\int_0^t \gamma_*(u) du} \|\mathbf{p}^1(0) - \mathbf{p}^2(0)\|_{1D} \end{aligned} \quad (5)$$

для любых начальных условий  $\mathbf{p}^1(0) \in \Omega$ ,  $\mathbf{p}^2(0) \in \Omega$  и всех  $t \geq 0$ .

**Замечание 1.** Отметим, что описываемый подход не гарантирует точных оценок. Например, для одно-родной цепи с двумя состояниями и интенсивностями переходов  $\lambda$  и  $\mu$  матрица  $A$  имеет вид:

$$A = \begin{pmatrix} -\lambda & \mu \\ \lambda & -\mu \end{pmatrix}.$$

Одно из собственных значений этой матрицы нулевое, а второе равно  $-(\lambda + \mu)$ , так что скорость сходимости к стационарному режиму имеет порядок  $e^{-(\lambda + \mu)t}$ . С другой стороны, пользуясь теоремой 1 и выбирая максимально возможное  $c = \mu$ , получаем оценку скорости сходимости  $e^{-\mu t}$ .

**Замечание 2.** В общем случае возможно использование как другой строки в матрице  $C$ , так и рассмотрение функции  $c(t)$  вместо положительной константы (см. пример далее).

**Замечание 3.** В частности, в случае марковской цепи с катастрофами (т. е. при возможности перехода из любого состояния в нулевое с ненулевой интенсивностью  $a_{0j}(t) = \gamma_j(t)$  для  $j \geq 1$ ) можно выбрать в качестве  $c(t)$  функцию  $c(t) = \inf_k \gamma_k(t)$ .

**Теорема 2.** Рассмотрим марковскую цепь  $X^*(t)$  с интенсивностями  $a_{i,j}^*(t)$ . Пусть при выполнении условий теоремы 1 последовательность  $\{d_k\}$  монотонно возрастает. Если для всех  $t \geq 0$  выполнены условия:

- (1)  $a_{i,j}(t) \leq a_{i,j}^*(t)$  при  $i < j$ ;
- (2)  $a_{i,j}(t) \geq a_{i,j}^*(t)$  при  $i > j$ ,

то цепь  $X^*(t)$  также слабо эргодична, причем  $\gamma_*^*(t) \geq \gamma_*(t)$ .

**Доказательство.** Достаточно убедиться, что  $\alpha_k^*(t) \geq \alpha_k(t)$ . Имеем

$$\alpha_k^*(t) = \begin{cases} \sum_{i>0} \left(1 - \frac{d_i}{d_0}\right) a_{i0}^* + c, & k = 0; \\ \sum_{i \neq 1} \left(1 - \frac{d_i}{d_1}\right) a_{i1}^* + c \frac{d_0}{d_1}, & k = 1; \\ \sum_{i \neq k} \left(1 - \frac{d_i}{d_k}\right) a_{ik}^* - c \frac{d_0}{d_k}, & k > 1 \end{cases} = \begin{cases} \sum_{i>0} \left(1 - \frac{d_i}{d_0}\right) a_{i0}^* + c, & k = 0; \\ \left(1 - \frac{d_0}{d_1}\right) a_{01}^* + \sum_{i>1} \left(1 - \frac{d_i}{d_1}\right) a_{i1}^* + c \frac{d_0}{d_1}, & k = 1; \\ \sum_{i<k} \left(1 - \frac{d_i}{d_k}\right) a_{ik}^* + \sum_{i>k} \left(1 - \frac{d_i}{d_k}\right) a_{ik}^* - c \frac{d_0}{d_k}, & k > 1 \end{cases} \geq \begin{cases} \sum_{i>0} \left(1 - \frac{d_i}{d_0}\right) a_{i0} + c, & k = 0; \\ \left(1 - \frac{d_0}{d_1}\right) a_{01} + \sum_{i>1} \left(1 - \frac{d_i}{d_1}\right) a_{i1} + c \frac{d_0}{d_1}, & k = 1; \\ \sum_{i<k} \left(1 - \frac{d_i}{d_k}\right) a_{ik} + \sum_{i>k} \left(1 - \frac{d_i}{d_k}\right) a_{ik} - c \frac{d_0}{d_k}, & k > 0 \end{cases} \geq \alpha_k(t).$$

Из этого следует и утверждение теоремы.

Пусть теперь имеется процесс рождения и гибели с интенсивностями рождения  $\lambda_k(t)$  и гибели  $\mu_k(t)$  и пусть для него подобраны нужные параметры так, что выполнены условия теоремы 1, а последовательность  $\{d_k\}$  монотонно возрастает. Тогда можно применить подход из рассуждений теоремы 2 и получить оценки скорости сходимости для так называемых single birth processes (см., например, [13, 14]), для которых интенсивности  $a_{ij}(t) \equiv 0$  при всех  $i - j \geq 2$ , а именно: будем предполагать,

что транспонированная матрица интенсивностей для марковской цепи  $X^*(t)$  имеет вид

$$A^*(t) = \begin{pmatrix} a_{00}^*(t) & a_{01}^*(t) & a_{02}^*(t) & a_{03}^*(t) & \dots \\ \lambda_0(t) & a_{11}^*(t) & a_{12}^*(t) & a_{13}^*(t) & \dots \\ 0 & \lambda_1(t) & a_{22}^*(t) & a_{23}^*(t) & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

**Теорема 3.** Пусть интенсивности марковской цепи  $X^*(t)$  удовлетворяют условию  $a_{ii}^*(t) \equiv a_{ii}(t)$  при всех  $i$ . Пусть при выполнении условий теоремы 1 последовательность  $\{d_k\}$  монотонно возрастает. Тогда  $X^*(t)$  слабо эргодична и справедлива оценка скорости сходимости (5).

Для доказательства достаточно отметить, что все суммы, входящие в (4), в условиях теоремы 3 заведомо не уменьшаются. В частности, если взять процесс рождения и гибели с интенсивностями  $\lambda_k = \lambda$  и  $\mu_k = \mu$  так, что

$$K = \left(\sqrt{\lambda} - \sqrt{\mu}\right)^2 - \frac{\lambda(\mu - \sqrt{\lambda\mu})}{\lambda + \mu} > 0,$$

то можно взять

$$c = \frac{\lambda(\mu - \sqrt{\lambda\mu})}{\lambda + \mu}; \quad d_k = \left(\sqrt{\frac{\mu}{\lambda}}\right)^k,$$

тогда получится  $\gamma_* = K$ .

### 4 Пример

Рассмотрим модель системы обслуживания, изученную в [8], где требования поступают по одному, а обслуживаются только парами. Транспонированная матрица интенсивностей  $A(t)$  имеет вид:

$$\begin{pmatrix} -\lambda(t) & 0 & \mu(t) & 0 \\ \lambda(t) & -\lambda(t) & 0 & \mu(t) \\ 0 & \lambda(t) & -(\lambda(t) + \mu(t)) & 0 \\ 0 & 0 & \lambda(t) & -(\lambda(t) + \mu(t)) \\ 0 & 0 & 0 & \lambda(t) \\ \vdots & \vdots & \vdots & \vdots \\ & & 0 & 0 \dots \\ & & 0 & 0 \dots \\ & & \mu(t) & 0 \dots \\ & & 0 & \mu(t) \dots \\ & & -(\lambda(t) + \mu(t)) & 0 \dots \\ & & \vdots & \vdots \dots \end{pmatrix}.$$

В работе [8] показано, что стандартный метод, упомянутый в начале настоящей статьи, не позволяет получить содержательную оценку скорости сходимости и применен метод неравенств, который удалось применить к данной модели (однако в общей ситуации для этого метода требуется перебор бесконечного числа случаев комбинаций знаков).

Попробуем применить описанный выше метод, выбрав для этого в матрице  $C(t)$  ненулевой первую строку (считая с нуля), т. е.

$$C(t) = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ c(t) & c(t) & c(t) & c(t) & \dots \\ 0 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

и выпишем суммы по столбцам для получившейся матрицы  $DW(t)D^{-1}$ , взятые с противоположным знаком:

$$\alpha_k(t) = \begin{cases} \lambda(t) - \frac{d_1}{d_0} |\lambda(t) - c(t)|, & k = 0; \\ c(t) - \left(\frac{d_2}{d_1} - 1\right) \lambda(t), & k = 1; \\ \left(1 - \frac{d_0}{d_2}\right) \mu(t) - \left(\frac{d_3}{d_2} - 1\right) \lambda(t) - c(t) \frac{d_1}{d_2}, & k = 2; \\ \mu(t) - \left(\frac{d_4}{d_3} - 1\right) \lambda(t) - \frac{d_1}{d_3} |\mu(t) - c(t)|, & k = 3; \\ \left(1 - \frac{d_{k-2}}{d_k}\right) \mu(t) - \left(\frac{d_{k+1}}{d_k} - 1\right) \lambda(t) - c(t) \frac{d_1}{d_k}, & k > 3. \end{cases} \quad (6)$$

Пусть  $d_0 = d_1 = 1$  и  $d_k = \delta^{k-1}$  для  $k \geq 2$  и  $\delta > 1$  и  $c = (\delta - 1)\mu(t)/(\delta + 1) < \lambda(t)$ . Подставим эти значения в (6) и получим:

$$\alpha_k(t) = \begin{cases} c(t), & k = 0; \\ c(t) - (\delta - 1) \lambda(t), & k = 1; \\ \left(1 - \frac{1}{\delta}\right) \mu(t) - (\delta - 1) \lambda(t) - c(t) \frac{1}{\delta}, & k = 2; \\ \mu(t) - (\delta - 1) \lambda(t) - \frac{1}{\delta^2} |\mu(t) - c(t)|, & k = 3; \\ \left(1 - \frac{1}{\delta^2}\right) \mu(t) - (\delta - 1) \lambda(t) - c(t) \frac{1}{\delta^{k-1}}, & k > 3 \end{cases} =$$

$$= \begin{cases} \frac{\delta - 1}{\delta + 1} \mu(t), & k = 0; \\ \frac{\delta - 1}{\delta + 1} \mu(t) - (\delta - 1) \lambda(t), & k = 1; \\ \frac{\delta - 1}{\delta + 1} \mu(t) - (\delta - 1) \lambda(t), & k = 2; \\ \mu(t) - (\delta - 1) \lambda(t) - \frac{1}{\delta^2} \frac{2}{\delta + 1} \mu(t), & k = 3; \\ \left(1 - \frac{1}{\delta^2}\right) \mu(t) - (\delta - 1) \lambda(t) - \frac{\delta - 1}{\delta + 1} \mu(t) \frac{1}{\delta^{k-1}}, & k > 3 \end{cases} \geq$$

$$\geq \begin{cases} \frac{\delta - 1}{\delta + 1} \mu(t) - (\delta - 1) \lambda(t); \\ \left(1 - \frac{1}{\delta^2}\right) \mu(t) - (\delta - 1) \lambda(t) + \frac{1}{\delta^2} \frac{\delta - 1}{\delta + 1} \mu(t); \geq \\ \left(1 - \frac{1}{\delta^2}\right) \mu(t) - (\delta - 1) \lambda(t) - \frac{1}{\delta^3} \frac{\delta - 1}{\delta + 1} \mu(t) \end{cases}$$

$$\geq \begin{cases} \frac{\delta - 1}{\delta + 1} \mu(t) - (\delta - 1) \lambda(t); \\ \left(1 - \frac{1}{\delta^2}\right) \mu(t) - (\delta - 1) \lambda(t) - \frac{1}{\delta^3} \frac{\delta - 1}{\delta + 1} \mu(t) \end{cases}$$

и  $\alpha_k \geq (\delta - 1)\mu(t)/(\delta + 1) - (\delta - 1)\lambda(t)$ .

**Теорема 4.** Пусть существует  $\delta > 1$  такое, что

$$(\delta - 1) \lambda(t) \leq \frac{(\delta - 1)\mu(t)}{\delta + 1} \leq \lambda(t)$$

и

$$\int_0^\infty (\mu(t) - (\delta + 1)\lambda(t)) dt = +\infty.$$

Тогда процесс, описывающий число требований в системе, слабо эргодичен, причем

$$\gamma_*(t) \geq \frac{(\delta - 1)\mu(t)}{\delta + 1} - (\delta - 1) \lambda(t).$$

Рассмотрим теперь конкретную модель с интенсивностями

$$\lambda(t) = 2 + \frac{1}{2} \sin(2\pi t), \quad \mu(t) = 8 - 2 \cos(2\pi t).$$

Положим  $\delta = 9/7$ . Тогда

$$\alpha_k \geq \frac{1}{8}(8 - 2 \cos(2\pi t)) - \frac{2}{7}(2 + \frac{1}{2} \sin(2\pi t)) \geq \frac{3}{22}$$

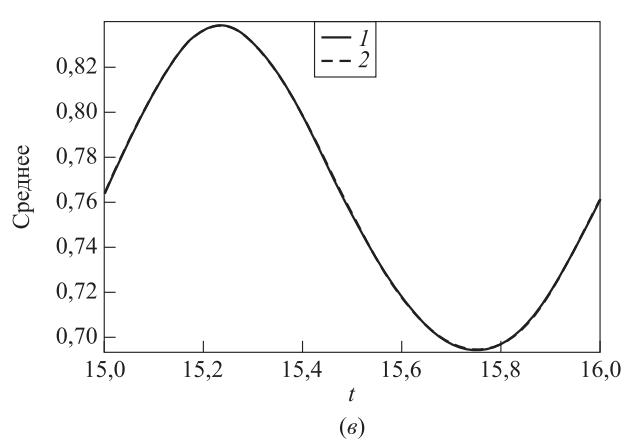
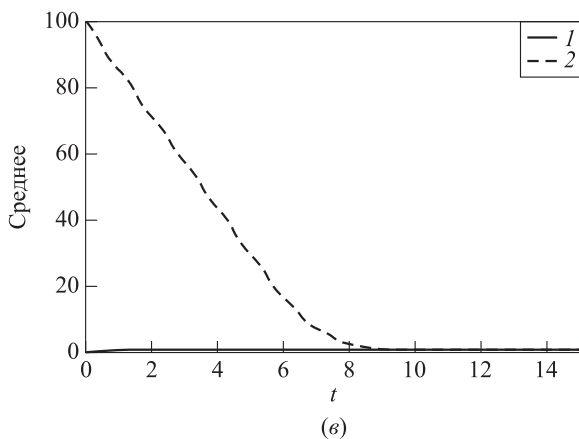
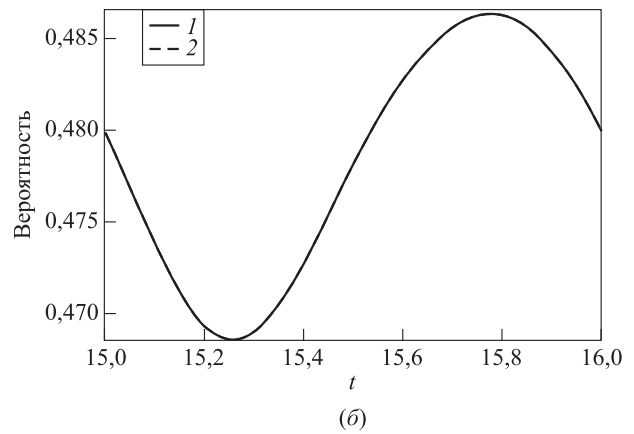
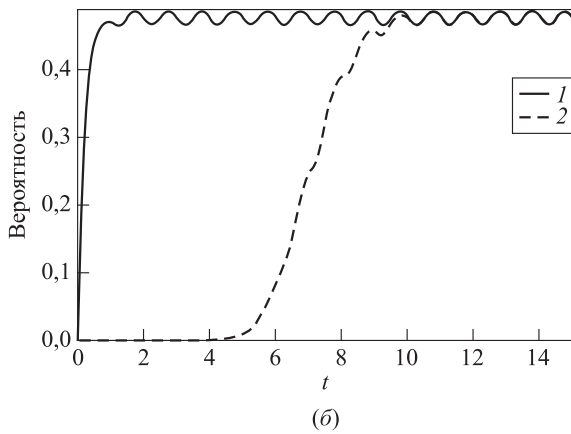
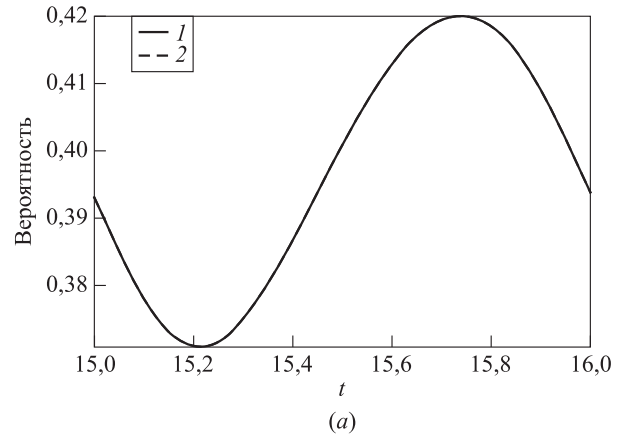
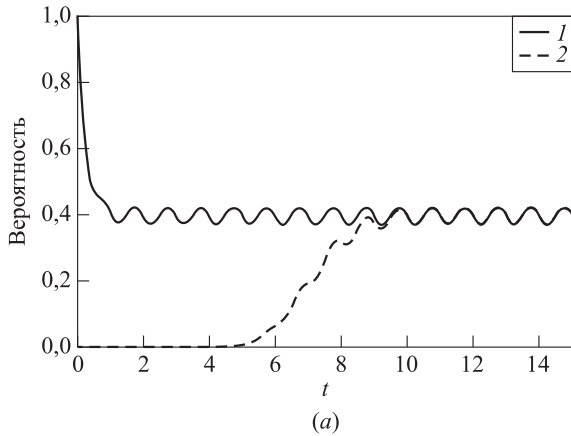
и оценка скорости сходимости имеет вид:

$$\|p^{*1}(t) - p^{*2}(t)\|_{1D} \leq e^{-3t/22} \|p^{*1}(0) - p^{*2}(0)\|_{1D}.$$



Далее следуем методу, который был подробно описан в [15], а именно: выбираем размерность усеченного процесса (в данном случае 100), интервал, на котором достигается желаемая точность

([0, 15]), и сам предельный интервал ([15, 16]), строим графики ожидаемого числа требований в системе и некоторых наиболее вероятных состояний (рис. 1 и 2).



**Рис. 1** Вероятность пустой системы массового обслуживания (а), вероятность  $p_1(t)$  системы массового обслуживания (б) и среднее  $E(t, k)$  (е) для  $t \in [0, 15]$

**Рис. 2** Аппроксимации предельной вероятности пустой системы массового обслуживания (а), предельной вероятности  $p_1(t)$  системы массового обслуживания (б) и предельного среднего значения  $E(t, k)$  (е) для  $t \in [15, 16]$

## Литература

1. Sidje R. B., Burrage K., Macnamara S. Inexact uniformization method for computing transient distributions of Markov chains // SIAM J. Sci. Comput., 2007. Vol. 29. P. 2562–2580.
2. Arns M., Buchholz P., Panchenko A. On the numerical analysis of inhomogeneous continuous-time Markov chains // Inform. J. Comput., 2010. Vol. 22. P. 416–432.
3. Andreychenko A., Sandmann W., Wolf V. Approximate adaptive uniformization of continuous-time Markov chains // Appl. Math. Model., 2018. Vol. 61. P. 561–576.
4. Burak M. R., Korytkowski P. Inhomogeneous CTMC birth-and-death models solved by uniformization with steady-state detection // ACM T. Model. Comput. S., 2020. Vol. 30. P. 1–18.
5. Inoue Y. A new approach to computing the transient-state probabilities in time-inhomogeneous Markov chains // J. Oper. Res. Soc. Jpn., 2022. Vol. 65. No. 1. P. 48–66.
6. Chakravarthy S. R. A catastrophic queueing model with delayed action // Appl. Math. Model., 2017. Vol. 46. P. 631–649.
7. Zeifman A. On the study of forward Kolmogorov system and the corresponding problems for inhomogeneous continuous-time Markov chains // Differential and difference equations with applications / Eds. S. Pinelas, J. R. Graef, S. Hilger, P. Kloeden, and C. Schinas. — Springer proceedings in mathematics and statistics ser. — Springer, 2020. Vol. 333. P. 21–39.
8. Satin Ya., Zeifman A., Kryukova A. On the rate of convergence and limiting characteristics for a nonstationary queueing model // Mathematics, 2019. Vol. 7. Iss. 8. Art. No. 678. 11 p.
9. Zeifman A., Satin Y., Kryukova A., Razumchik R., Kiseleva K., Shilova G. On three methods for bounding the rate of convergence for some continuous-time Markov chains // Int. J. Appl. Math. Comp., 2020. Vol. 30. No. 2. P. 251–266.
10. Zeifman A., Satin Y., Sipin A. Bounds on the rate of convergence for  $M_t^X/M_t^X/1$  queueing models // Mathematics, 2021. Vol. 9. Iss. 1. Art. No. 1752. 11 p.
11. Сатин Я. А. Исследование модели типа  $M_t/M_t/1$  с двумя различными классами требований // Системы и средства информатики, 2021. Т. 31. № 1. С. 17–27.
12. Razumchik R., Rumyantsev A. Some ergodicity and truncation bounds for a small scale Markovian supercomputer model // 36th ECMS Conference (International) on Modelling and Simulation Proceedings. — Saarbrücken—Dudweiler, Germany: Digitaldruck Pirrot GmbH, 2022. P. 324–330.
13. Chen M. Single birth processes // Chinese Annals Mathematics, 1999. Vol. 20. No. 1. P. 77–82.
14. Mao Y. H., Zhang Y. H. Exponential ergodicity for single-birth processes // J. Appl. Probab., 2004. Vol. 41. No. 4. P. 1022–1032.
15. Zeifman A., Satin Y., Korolev V., Shorgin S. On truncations for weakly ergodic inhomogeneous birth and death processes // Int. J. Appl. Math. Comp., 2014. Vol. 24. No. 3. P. 503–518.

Поступила в редакцию 25.06.22

## ON AN APPROACH FOR ESTIMATING THE RATE OF CONVERGENCE FOR NONSTATIONARY MARKOV MODELS OF QUEUEING SYSTEMS

I. A. Kovalev<sup>1,2</sup>, Y. A. Satin<sup>1</sup>, A. V. Sinitcina<sup>3</sup>, and A. I. Zeifman<sup>1,2,4,5</sup>

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

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

<sup>3</sup>P. G. Demidov Yaroslavl State University, 14 Sovetskaya Str., Yaroslavl 150003, Russian Federation

<sup>4</sup>Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

<sup>5</sup>Vologda Research Center of the Russian Academy of Sciences, 56A Gorky Str., Vologda 160014, Russian Federation

**Abstract:** The transformation of the forward Kolmogorov system is considered which allows one to obtain simple estimates on the rate of convergence for Markov chains with continuous time describing queueing systems. In addition, the concept of the logarithmic norm of the operator function and the associated estimates of the norm of the Cauchy matrix are used. The results obtained make it possible to estimate the rate of convergence for new classes of models in which the matrix is not significantly nonnegative and the use of the logarithmic norm method does not guarantee the possibility of obtaining estimates of the rate of convergence. Previously, a rather laborious more general method of inequalities was used for such situations. A theorem is formulated on obtaining the rate of convergence when the intensities of the matrix change. An estimate was obtained for the process of

birth and death with constant intensities. As an example, a special nonstationary model with group service of requirements (service in pairs) is investigated.

**Keywords:** rate of convergence; ergodicity bounds; logarithmic norm; queuing systems

**DOI:** 10.14357/19922264220310

## Acknowledgments

The results of Section 4 were obtained by A. V. Sinitcina supported by the Russian Science Foundation (grant No. 21-71-30011).

## References

1. Sidje, R. B., K. Burrage, and S. Macnamara. 2007. Inexact uniformization method for computing transient distributions of Markov chains. *SIAM J. Sci. Comput.* 29:2562–2580.
2. Arns, M., P. Buchholz, and A. Panchenko. 2010. On the numerical analysis of inhomogeneous continuous-time Markov chains. *Inform. J. Comput.* 22:416–432.
3. Andreychenko, A., W. Sandmann, and V. Wolf. 2018. Approximate adaptive uniformization of continuous-time Markov chains. *Appl. Math. Model.* 61:561–576.
4. Burak, M. R., and P. Korytkowski. 2020. Inhomogeneous CTMC birth-and-death models solved by uniformization with steady-state detection. *ACM T. Model. Comput. S.* 30:1–18.
5. Inoue, Y. 2022. A new approach to computing the transient-state probabilities in time-inhomogeneous Markov chains. *J. Oper. Res. Soc. Jpn.* 65(1):48–66.
6. Chakravarthy, S. R. 2017. A catastrophic queueing model with delayed action. *Appl. Math. Model.* 46:631–649.
7. Zeifman, A. 2020. On the study of forward Kolmogorov system and the corresponding problems for inhomogeneous continuous-time Markov chains. *Differential and difference equations with applications*. Eds. S. Pinelas, J. R. Graef, S. Hilger, P. Kloeden, and C. Schinas. Springer proceedings in mathematics and statistics ser. Springer. 333:21–39.
8. Satin, Y., A. Zeifman, and A. Kryukova. 2019. On the rate of convergence and limiting characteristics for a nonstationary queueing model. *Mathematics* 7(8):678. 11 p.
9. Zeifman, A., Y. Satin, A. Kryukova, R. Razumchik, K. Kiseleva, and G. Shilova. 2020. On the three methods for bounding the rate of convergence for some continuous time Markov chains. *Int. J. Appl. Math. Comp.* 30(2):251–266.
10. Zeifman, A., Y. Satin, and A. Sipin. 2021. Bounds on the rate of convergence for  $M_t^X/M_t^X/1$  queueing models. *Mathematics* 9(1):1752. 11 p.
11. Satin, Y. A. 2021. Issledovanie modeli tipa  $M_t/M_t/1$  s dvumya razlichnymi klassami trebovaniy [On the bounds of the rate of convergence for  $M_t/M_t/1$  model with two different requests]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(1):17–27.
12. Razumchik, R., and A. Rumyantsev. 2022. Some ergodicity and truncation bounds for a small scale markovian supercomputer model. *36th ECMS Conference (International) on Modelling and Simulation Proceedings*. Saarbrücken–Dudweiler, Germany: Digitaldruck Pirrot GmbH. 324–330.
13. Chen, M. 1999. Single birth processes. *Chinese Annals Mathematics* 20(1):77–82.
14. Mao, Y. H., and Y. H. Zhang. 2004. Exponential ergodicity for single-birth processes. *J. Appl. Probab.* 41(4):1022–1032.
15. Zeifman, A., Y. Satin, V. Korolev, and S. Shorgin. 2014. On truncations for weakly ergodic inhomogeneous birth and death processes. *Int. J. Appl. Math. Comp.* 24(3):503–518.

Received June 25, 2022

## Contributors

**Kovalev Ivan A.** (b. 1996) — PhD student, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; scientist, Moscow Center for Fundamental and Applied Mathematics, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; kovalev.iv96@yandex.ru

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

**Sinitcina Anna V.** (b. 1988) — Candidate of Science (PhD) in physics and mathematics, scientist, P. G. Demidov Yaroslavl State University, 14 Sovetskaya Str., Yaroslavl 150003, Russian Federation; a.korotyshva@mail.ru

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

# ГИСТЕРЕЗИСНОЕ УПРАВЛЕНИЕ НАГРУЗКОЙ В БЕСПРОВОДНЫХ СЕНСОРНЫХ СЕТЯХ\*

Б. Алию<sup>1</sup>, Е. А. Мачнев<sup>2</sup>, Е. В. Мокров<sup>3</sup>

**Аннотация:** Решается задача анализа перегрузок в беспроводной сенсорной сети. Предложен механизм гистерезисного управления, отсеивающий часть нагрузки в случае перегрузки системы для стабилизации ее функционирования. Построена математическая модель в виде марковского процесса с конечным пространством состояний, и получена формула для анализа показателей эффективности беспроводной сенсорной сети. Численно исследованы вероятностно-временные характеристики гистерезисного управления — вероятность сброса пакета и средняя длина очереди. Сравнение гистерезисного механизма и алгоритма IRED (improved random early detection — улучшенное случайное раннее обнаружение) при варьировании размера окна управления показало, что гистерезисный механизм дает преимущество в области больших нагрузок по сравнению с механизмом IRED.

**Ключевые слова:** беспроводные сенсорные сети; управление нагрузкой; гистерезисное управление; марковский процесс; система массового обслуживания

**DOI:** 10.14357/19922264220311

## 1 Введение

Автономный характер беспроводных сенсорных сетей в сочетании с их способностью воспринимать, обрабатывать и передавать данные об окружающей среде на базовую станцию сотовой сети [1] способствует их широкому распространению. Согласно данным IoT (Internet of Things) Analytics [2], более 14,4 млрд сенсоров взаимодействуют с использованием беспроводных сенсорных сетей. Такое число устройств IoT может привести к перегрузкам в сети (см., например, [3]), особенно с ростом числа устройств, использующих технологию беспроводных сенсорных сетей.

Согласно Cisco Systems [4] к 2023 г. прогнозируется порядка 14,7 млрд активных IoT-соединений, включая датчики, носимые устройства, смартфоны, беспилотные автомобили и беспилотные летательные аппараты (БПЛА). В отсутствие управления трафиком в беспроводных сенсорных сетях возможны частые и продолжительные перегрузки. Эта особенность исследуется в статье с помощью сравнительного анализа производительности сети на основе гистерезисной схемы управления нагрузкой [5, 6] и улучшенной схемы управления перегрузкой со случайным ранним обнаружением (IRED) [7]. Именно перегрузки в беспроводной сенсорной сети и стали предметом исследования настоящей работы.

В статье предложена системная модель для наиболее типичной архитектуры беспроводной сенсорной сети, в марковских предположениях построена математическая модель, проведен ее анализ и предложены формулы для вычисления ее характеристик. Численный анализ направлен на сравнение алгоритма гистерезисного управления нагрузкой с алгоритмом управления IRED [7]. Показано, что в области высокой нагрузки гистерезисное управление дает преимущество по вероятности сброса, т. е. потере пакета при поступлении, и средней длине очереди пакетов на шлюзе.

## 2 Системная модель

Рассматривается сценарий, в котором датчики, расположенные в пределах некоторой заданной области, передают информацию на центральный узел, отвечающий за ее хранение и обработку. Для сбора данных с датчиков используются шлюзы, размещенные на БПЛА, патрулирующих территорию аналогично одному из случаев, описанных в [6], и более подробно рассмотренному в [8].

Системная модель представляет собой беспроводную сенсорную сеть, в которой кластер беспроводных датчиков загружает полученные данные об окружающей среде через шлюз в базу данных (БД), размещенную в облаке. Архитектура сети, как по-

\*Исследование выполнено за счет гранта Российского научного фонда № 21-79-00157, <https://rscf.ru/project/21-79-00157> (получатель Е. В. Мокров, системная и математическая модели), и при поддержке Программы стратегического академического лидерства РУДН (получатель Е. А. Мачнев, введение).

<sup>1</sup> Российский университет дружбы народов, bashaliyuu@gmail.com

<sup>2</sup> Российский университет дружбы народов, 1032143100@rudn.ru

<sup>3</sup> Российский университет дружбы народов, mokrov-ev@rudn.ru

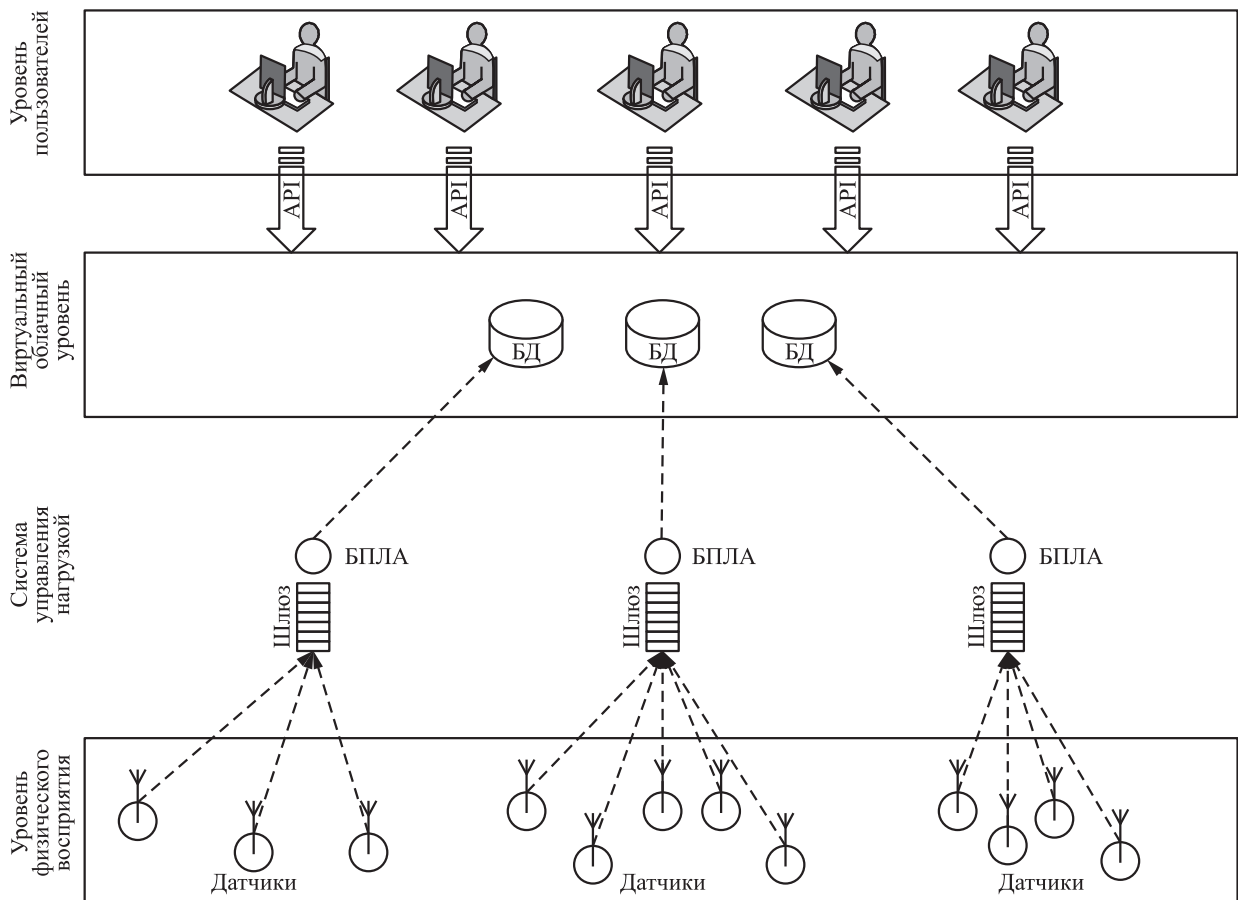


Рис. 1 Схема системной модели беспроводной сенсорной сети

казано на рис. 1, включает три уровня: уровень физического восприятия, где датчики собирают данные об окружающей среде, например температуру, влажность, освещенность; виртуальный облачный уровень, где размещены БД, в которые передается на хранение собранная датчиками информация; уровень пользователей, на котором расположены веб-интерфейсы для взаимодействия с БД. Передача данных между первым и вторым уровнями происходит через шлюз, на котором размещена система управления нагрузкой, осуществляющая мониторинг и снижение нагрузки для предотвращения перегрузок.

В статье рассматривается гистерезисная схема управления нагрузкой с двумя порогами [5] для предотвращения перегрузок шлюза, результатом которых может стать переполнение памяти шлюза.

### 3 Математическая модель

Описанная сеть моделируется системой массового обслуживания типа  $M/M/1/B$  с гистерезисным управлением нагрузкой [9]. На систему

с одним прибором, накопителем емкости  $B$  и экспоненциальным обслуживанием с параметром  $\mu$  поступает пуассоновский поток заявок, соответствующих пакетам данных от сенсоров, с интенсивностью  $\lambda$ . Считаем, что заявки сохраняют за собой место в накопителе до момента окончания обслуживания.

Для контроля перегрузок в системе используются два порога: нижний порог  $L$  и верхний порог  $H$ . Система работает в трех режимах: режиме нормальной работы ( $s = 0$ ), режиме сниженной нагрузки ( $s = 1$ ) и режиме блокировки ( $s = 2$ ). Когда число заявок в системе превышает верхний порог  $H$ , система переходит в режим сниженной нагрузки ( $s = 1$ ). Принятая в систему нагрузка в этом режиме снижается до уровня  $\lambda^* = (1 - p)\lambda$ , при этом часть нагрузки  $p\lambda$  сбрасывается, таким образом  $p$  — это доля сбрасываемой нагрузки при нахождении системы в режиме сниженной нагрузки. Система продолжает находиться в режиме сниженной нагрузки либо до заполнения системы, либо до достижения числом заявок в очереди нижнего порога  $L$ . В первом случае система перейдет в режим



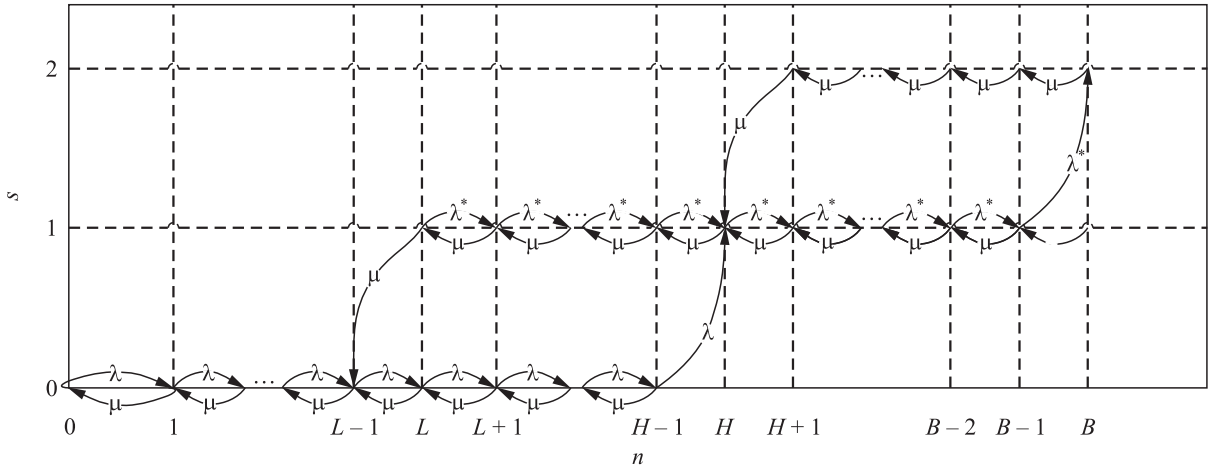


Рис. 2 Диаграмма интенсивностей переходов МП  $X(t)$  для системы с гистерезисным управлением

блокировки ( $s = 2$ ) и перестанет принимать новые, но продолжит обслуживать уже находящиеся в системе заявки. В таком состоянии система будет пребывать до тех пор, пока число заявок не упадет до верхнего порога, после чего возвратится в режим сниженной нагрузки ( $s = 1$ ). Во втором случае нагрузка на систему восстановится до исходной, и система перейдет в режим нормальной работы ( $s = 0$ ).

Функционирование системы описывается марковским процессом (МП)  $X(t)$  над пространством состояний  $\mathcal{X} = \{(n, s) : s \in \{0, 1, 2\}, n \in \{0, \dots, B\}\}$ , которое представимо в виде разбиения

$$\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1 \cup \mathcal{X}_2,$$

где  $\mathcal{X}_0 = \{(0, n) : n = 0, \dots, H - 1\}$  — множество состояний нормальной работы;  $\mathcal{X}_1 = \{(1, n) : n = L, \dots, B - 1\}$  — множество состояний сниженной нагрузки;  $\mathcal{X}_2 = \{(2, n) : n = H + 1, \dots, B\}$  — множество состояний блокировки. На рис. 2 проиллюстрирована диаграмма интенсивностей переходов состояний МП  $X(t)$ .

Решив для МП  $X(t)$  систему уравнений равновесия, получим стационарное распределение  $p_{sn}, (s, n) \in \mathcal{X}$ , с помощью которого в явном виде могут быть представлены следующие вероятностно-временные характеристики:  $B(\mathcal{X}_0) = \sum_{n=0}^{H-1} p_{0,n}$  — вероятность того, что система работает в режиме нормальной нагрузки;  $B(\mathcal{X}_1) = \sum_{n=L}^{B-1} p_{1,n}$  — вероятность того, что система работает в режиме сниженной нагрузки и часть заявок будет сброшена;  $B(\mathcal{X}_2) = \sum_{n=N+1}^B p_{2,n}$  — вероятность того, что система перегружена и не принимает новые заявки;  $B = (1 - p)B(\mathcal{X}_1) + B(\mathcal{X}_2)$  — вероятность сброса заявки;  $MQ = \sum_{(s,n) \in \mathcal{X}} np_{s,n}$  — средняя длина очереди.

Аналогичные характеристики для алгоритма IRED вычисляются согласно [7], что позволяет провести сравнение двух механизмов защиты от перегрузок.

#### 4 Численный анализ

Было проведено численное сравнение вероятностей сброса пакета для обоих рассматриваемых алгоритмов в зависимости от параметров  $B, L, H, p$  и  $\rho = \lambda/\mu$  при значениях исходных данных из [7]. В результате анализа было показано, что алгоритм гистерезисного управления более эффективен по сравнению с алгоритмом IRED при больших значениях нагрузки. Заметим, что при этом с точки зрения средней длины очереди алгоритм гистерезисного управления уступает алгоритму IRED. Данные результаты представлены на рис. 3 при фиксированном значении объема накопителя  $B = 50$  и нижнего порога  $L = 10$  и варьировании остальных параметров в пределах  $0,1 \leq p < 1$  и  $12 \leq H \leq 40$ . График показывает минимальное пороговое значение нагрузки  $\rho_0$ , после которого гистерезисный подход позволяет успешно передать больше пакетов, чем алгоритм IRED, т.е.  $\rho_0 : \forall \rho > \rho_0 B_{\text{hyst}}(\rho) < B_{\text{IRED}}(\rho)$ . Из рис. 3 можно заключить, что пороговое значение нагрузки  $\rho_0$  почти линейно зависит от доли  $p$  сбрасываемой нагрузки в состоянии сниженной нагрузки и слабо зависит от изменения верхнего порога при фиксированном нижнем. Аналогичная картина наблюдается и при других значениях нижнего порога.

На рис. 4 отдельно представлены результаты сравнения производительности системы с гистерезисным управлением и управлением с помощью алгоритма IRED для объема накопителя  $B = 40$ ,

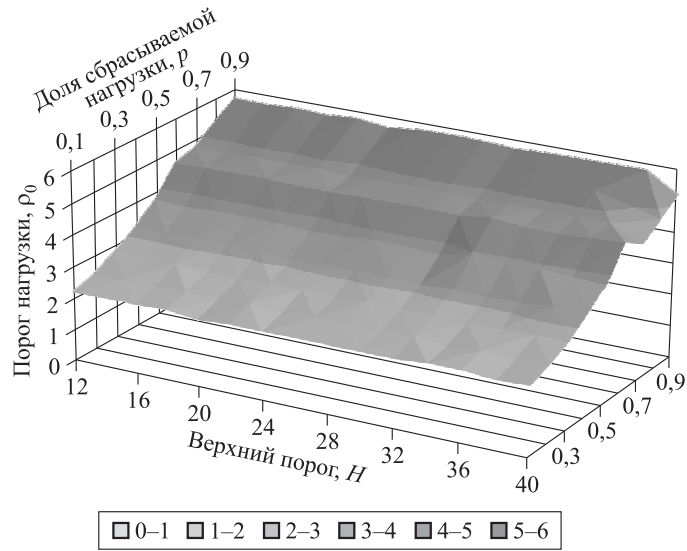


Рис. 3 Пороговое значение нагрузки

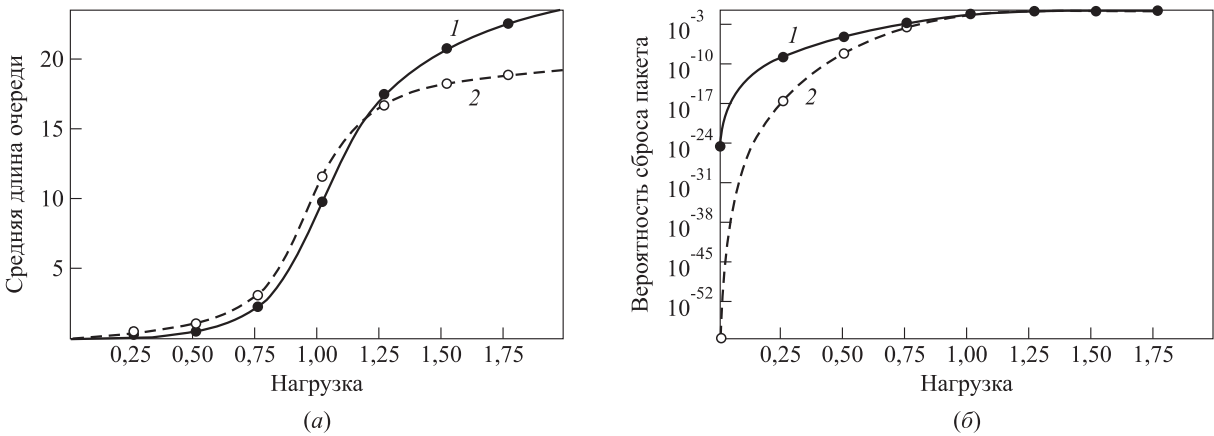


Рис. 4 Средняя длина очереди (а) и вероятность сброса пакета (б): 1 — гистерезисное управление; 2 — управление с помощью алгоритма IRED

доли сбрасываемых пакетов  $\rho = 0,3$ , нижнего и верхнего порогов  $L = 10$  и  $H = 30$  соответственно.

Средняя длина очереди в зависимости от предложенной нагрузки, варьирующейся в пределах  $0 < \rho < 2$ , показана на рис. 4, а. На рис. 4, б в логарифмической шкале показана вероятность сброса пакетов в зависимости от предложенной нагрузки.

Из приведенных результатов видно, что пороговое значение нагрузки также существенно зависит от объема накопителя, поскольку в данном случае алгоритм IRED показывает более хорошие результаты при низкой нагрузке, а также при небольшой перегрузке  $\rho \leq 1,2$ . Однако уже начиная со значения нагрузки  $\rho > 1,2$  лучшие характеристики пока-

зывает гистерезисное управление. Таким образом, в проведенном эксперименте пороговое значение нагрузки уменьшилось по сравнению с предыдущим экспериментом до значения  $\rho_0 = 1,2$  при тех же значениях порогов и доли сбрасываемой нагрузки. Это показывает, что значение порога обратно пропорционально размеру накопителя.

На рис. 5 показано сравнение систем в зависимости от размера окна управления  $(H - L)$  при фиксированном верхнем пороге  $H = 35$ , накопителе объема  $B = 40$ , доле сбрасываемых пакетов  $\rho = 0,3$  и изменении нижнего порога  $L$  от 5 до 35.

Из графиков видно, что, хотя пороговое значение нагрузки достаточно слабо зависит от значения нижнего порога, значение порога достаточно сильно влияет на среднюю длину очереди.

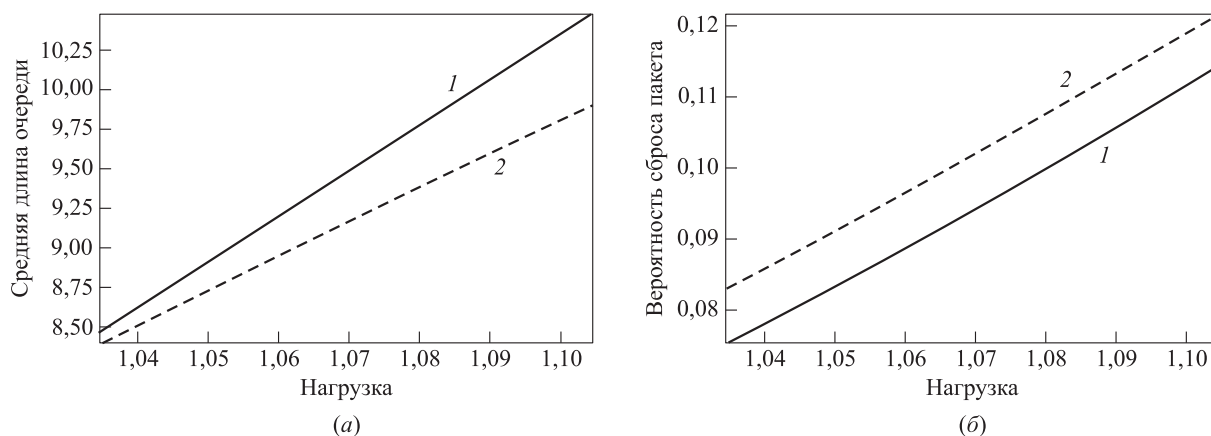


Рис. 5 Средняя длина очереди (а) и вероятность сброса пакета (б) в зависимости от размера окна управления: 1 — гистерезисное управление; 2 — управление с помощью алгоритма IRED

## 5 Заключение

Численный эксперимент показал, что наиболее значимыми параметрами, влияющими на пороговое значение нагрузки, при превышении которого гистерезисный подход дает лучшие результаты по сравнению с алгоритмом IRED, с точки зрения вероятности сброса являются объем накопителя и доля сбрасываемой нагрузки. Также было показано, что эти два параметра наиболее значимы и для фиксированного порогового значения нагрузки.

Помимо этого, для типичного набора исходных данных исследована зависимость показателей качества функционирования системы от размера окна управления. Показано, что для рассмотренного набора данных при размере окна  $H - L \leq 3$  гистерезисное управление дает более хорошие результаты, чем IRED, во всех случаях, когда  $\rho > 1$ , т. е. при перегрузках системы. При  $\rho = 0,8$  в IRED длина очереди меньше, но разница полученных значений для обоих алгоритмов мала, а поведение гистерезисной системы объясняется ее недогруженностью. Таким образом, размер окна управления, хоть и в меньшей степени, но также влияет на пороговое значение нагрузки.

В работе для модели беспроводной облачной сенсорной сети с механизмом гистерезисного управления нагрузкой исследованы основные показатели системы — вероятность сброса пакета и средняя длина очереди. Проведено численное сравнение гистерезисного подхода с алгоритмом IRED и показано, что при перегрузках гистерезисный подход показывает качественно лучшую картину, чем IRED, с точки зрения вероятности сброса пакета. Также введено понятие порогового значения нагрузки, при превышении которого гистерезисный подход более эффективен, чем алгоритм

IRED, и показана зависимость порогового значения нагрузки от параметров системы. Для нагрузки ниже порога гистерезисный подход незначительно проигрывает алгоритму типа IRED в некоторых диапазонах значений параметров системы. При этом алгоритм IRED реализуется на сетевом уровне, а гистерезисное управление нагрузкой — на прикладном уровне, где эти различия оказываются несущественными.

## Литература

1. Yadav S. L., Ujjwal R. L. Mitigating congestion in wireless sensor networks through clustering and queue assistance: A survey // J. Intell. Manuf., 2021. Vol. 32. P. 2083–2098. doi: 10.1007/s10845-020-01640-8.
2. Hasan M. State of IoT 2022: Number of connected IoT devices growing 18% to 14.4 billion globally. Pr., 2022. <https://iot-analytics.com/number-connected-iot-devices>.
3. Shiny S. S. G., Murugan K. TSDN-WISE: Automatic threshold-based low control-flow communication protocol for SDWSN // IEEE Sens. J., 2021. Vol. 21. No. 17. P. 19560–19569. doi: 10.1109/JSEN.2021.3088604.
4. Cisco Annual Internet Report (2018–2023) white paper, 2020. <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>.
5. Bashir A., Machnev E., Mokrov E. Queueing model of hysteretic congestion control for cloud wireless sensor networks // 13th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops. — Piscataway, NJ, USA: IEEE, 2021. P. 104–108. doi: 10.1109/ICUMT54235.2021.9631576.
6. Uyoata U., Mwangama J., Adeogun R. Relaying in the Internet of Things (IoT): A survey // IEEE Access, 2021. Vol. 9. P. 132675–132704. doi: 10.1109/ACCESS.2021.3112940.
7. Huang J., Du D., Duan Q., Zhang Y., Zhao Y., Luo H., Mai Z., Liu Q. Modeling and analysis on congestion con-

- trol for data transmission in sensor clouds // *Int. J. Distrib. Sens. N.*, 2014. Vol. 2014. Art. ID 453983. P. 1–9. doi: 10.1155/2014/453983.
8. Petrov V., Samuylov A., Begishev V., Moltchanov D., Andreev S., Samouylov K., Koucheryavy Y. Vehicle-based relay assistance for opportunistic crowdsensing over Narrowband IoT (NB-IoT) // *IEEE Internet Things*, 2018. Vol. 5. No. 5. P. 3710–3723. doi: 10.1109/JIOT.2017.2670363.
9. Красносельский М. А., Покровский А. В. Системы с гистерезисом. — М. Наука, 1983. 272 с.

Поступила в редакцию 01.07.22

## HYSTERETIC CONGESTION CONTROL IN WIRELESS CLOUD SENSOR NETWORKS

B. Aliyu, E. A. Machnev, and E. V. Mokrov

Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

**Abstract:** The paper considers the problem of congestion analysis in a wireless sensor network. A hysteresis control mechanism is proposed which screens out part of the load in the event of an overload of the system in order to stabilize its operation. A mathematical model is constructed in the form of a Markov process with a finite state space and a formula is obtained for analyzing the performance indicators of a wireless sensor network. The probabilistic-temporal characteristics of hysteresis control, the packet drop probability, and the average queue length are studied numerically. The hysteresis mechanism is compared with the IRED (improved random early detection) algorithm with different control window sizes. It is shown that the hysteresis mechanism provides an advantage in the region of high loads compared to the IRED mechanism.

**Keywords:** wireless sensor networks; load control; hysteresis control; Markov process; queuing system

**DOI:** 10.14357/19922264220311

### Acknowledgments

The reported study was funded by the Russian Science Foundation, project number 21-79-00157 (E. V. Mokrov, system model and mathematical model) and by the RUDN University Strategic Academic Leadership Program (E. A. Machnev, introduction).

### References

1. Yadav, S. L., and R. L. Ujjwal. 2021. Mitigating congestion in wireless sensor networks through clustering and queue assistance: A survey. *J. Intell. Manuf.* 32:2083–2098. doi: 10.1007/s10845-020-01640-8.
2. Hasan, M. 2022. State of IoT 2022: Number of connected IoT devices growing 18% to 14.4 billion globally. Available at: <https://iot-analytics.com/number-connected-iot-devices/> (accessed July 18, 2022).
3. Shiny, S. S. G., and K. Murugan. 2021. TSDN-WISE: Automatic threshold-based low control-flow communication protocol for SDWSN. *IEEE Sens. J.* 21(17):19560–19569. doi: 10.1109/JSEN.2021.3088604.
4. Cisco Annual Internet Report (2018–2023) white paper. 2020. Available at: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html> (accessed July 18, 2022).
5. Bashir, A., E. Machnev, and E. Mokrov. 2021. Queuing model of hysteretic congestion control for cloud wireless sensor networks. *13th Congress (Internatioanal) on Ultra Modern Telecommunications and Control Systems Proceedings*. Piscataway, NJ: IEEE. 104–108. doi: 10.1109/ICUMT54235.2021.9631576.
6. Uyoata, U., J. Mwangama, and R. Adeogun. 2021. Relaying in the Internet of Things (IoT): A survey. *IEEE Access* 9:132675–132704. doi: 10.1109/ACCESS.2021.3112940.
7. Huang, J., D. Du, Q. Duan, Y. Zhang, Y. Zhao, H. Luo, Z. Mai, and Q. Liu. 2014. Modeling and analysis on congestion control for data transmission in sensor clouds. *Int. J. Distrib. Sens. N.* 10(3):453983. doi: 10.1155/2014/453983.
8. Petrov, V., A. Samuylov, V. Begishev, D. Moltchanov, S. Andreev, K. Samouylov, and Y. Koucheryavy. 2017. Vehicle-based relay assistance for opportunistic crowd sensing over Narrowband IoT (NB-IoT). *IEEE Internet Things* 5(5):3710–3723. doi: 10.1109/JIOT.2017.2670363.
9. Krasnoselskii, M., and A. Pokrovskii. 1989. *Systems with hysteresis*. Berlin – Heidelberg – New York – London – Paris – Tokio: Springer Verlag. 410 p.

Received July 1, 2022

## Contributors

**Aliyu Bashir** (b. 1975) — PhD student, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; bashaliyuu@gmail.com

**Machnev Egor A.** (b. 1996) — PhD student, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; 1032143100@rudn.ru

**Mokrov Evgeniy V.** (b. 1988) — Candidate of Science (PhD) in physics and mathematics, assistant professor, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; mokrov-ev@rudn.ru



# УПРАВЛЯЕМАЯ СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С ЭЛАСТИЧНЫМ ТРАФИКОМ И СИГНАЛАМИ ДЛЯ АНАЛИЗА НАРЕЗКИ РЕСУРСОВ В СЕТИ РАДИОДОСТУПА\*

А. С. Власкина<sup>1</sup>, С. А. Бурцева<sup>2</sup>, И. А. Кочеткова<sup>3</sup>, С. Я. Шоргин<sup>4</sup>

**Аннотация:** Построена управляемая система массового обслуживания с нетерпеливым эластичным трафиком и минимально гарантированной скоростью обслуживания для анализа динамической нарезки радиоресурсов беспроводной сети. Управление перераспределением ресурса между двумя сегментами сети осуществляется при поступлении сигнала контроллера и в случае простаивания ресурса одного сегмента и ожидающих начала обслуживания пользователей другого сегмента. В зависимости от состояния системы принимается решение о том, на какой объем перераспределить ресурс. Принятие решения основано на показателях эффективности нарезки, учитывающих использование ресурса каждого из сегментов, отклонение от рекомендуемого «начального» разделения, дополнительную сигнальную нагрузку от контроллера. Предложена соответствующая целевая функция управления, применен итерационный алгоритм расчета параметров оптимальной политики распределения ресурса.

**Ключевые слова:** 5G; нарезка ресурсов; управляемая система массового обслуживания; марковский процесс принятия решения; эластичный трафик

**DOI:** 10.14357/19922264220312

## 1 Введение

Одна из ключевых концепций сетей пятого и последующих поколений — технология нарезки сети (*англ.* network slicing) — позволяет создавать логические сети (сегменты), каждая из которых имеет соответствующие ресурсы для обслуживания пользователей определенных услуг [1–3]. Общая конструкция системы обеспечивает экономичность и гибкость настройки и развертывания сегментов сети [4–6].

В работе исследуется модель сети, где имеются базовый оператор с ресурсом, а также два виртуальных оператора, которые запрашивают доступ к этому ресурсу (рис. 1). В качестве услуг, которые предоставляют виртуальные операторы своим пользователям, рассматриваются услуги передачи данных с минимально допустимой скоростью обслуживания, а также с ограничениями на время ожидания начала обслуживания. Отношения между базовым и виртуальными операторами регулирует соглашение о качестве обслуживания (*англ.* Service Level Agreement, SLA). В соответствии с этим соглашением базовый оператор планирует распределение ресурса в условиях нормальной загрузки

(«начальное» разделение ресурса). Управление заключается в перераспределении ресурсов между классами трафика для предотвращения ухудшения качества обслуживания в одном сегменте сети из-за нехватки ресурса в другом.

Стратегией управления радиоресурсами занимается контроллер, который с определенной периодичностью направляет сигналы о проверке необходимости перераспределения. Задачей является поиск оптимальных способов планирования радиоресурсов (по каким критериям и сколько ресурсов зарезервировать под конкретную услугу) в зависимости от соответствия начальному распределению, максимального использования ресурсов, а также числа сигналов контроллера, приведших к перераспределению.

Данная работа является продолжением исследований [7, 8], анализирующих эффективность динамической нарезки радиоресурсов между сегментами сети, с учетом перечисленных выше трех принципов. Исследована модель управления ресурсами при поступлении сигнала контроллера в виде управляемой системы массового обслуживания, применен итерационный алгоритм вычисления оптимальной стратегии.

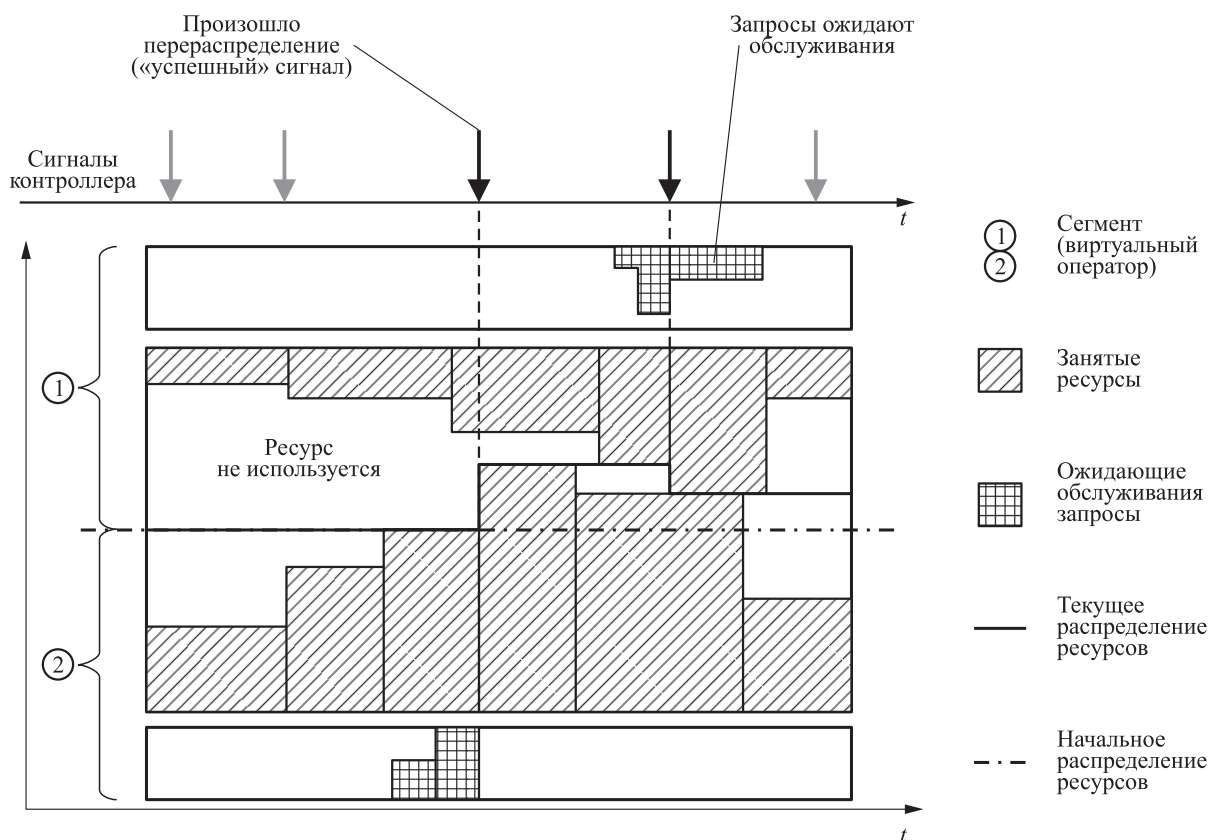
\* Публикация выполнена при поддержке Программы стратегического академического лидерства РУДН.

<sup>1</sup> Российский университет дружбы народов, vlasnikina-as@rudn.ru

<sup>2</sup> Российский университет дружбы народов, sofya\_burceva@inbox.ru

<sup>3</sup> Российский университет дружбы народов; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, kochetkova-ia@rudn.ru

<sup>4</sup> Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru



**Рис. 1** Схема перераспределения ресурса: 1 — сегмент (виртуальный оператор); 2 — занятые ресурсы; 3 — ожидающие обслуживания запросы; 4 — текущее распределение ресурсов; 5 — начальное распределение ресурсов

## 2 Управляемая система массового обслуживания

Рассмотренную модель с двумя классами услуг можно описать в виде управляемой системы массового обслуживания [9] с нетерпеливым эластичным трафиком, минимально гарантированной скоростью обслуживания и потоком сигналов, которые моделируют сообщения от контроллера. Схематическое изображение системы представлено на рис. 2.

Будем считать, что поток заявок на обслуживание пользователей виртуальных операторов пуассоновский с интенсивностями  $\lambda_1$  и  $\lambda_2$ . Обслуживание — экспоненциальное со средним объемом передаваемых данных  $\mu_1^{-1}$  и  $\mu_2^{-1}$ . С учетом минимально гарантированной скорости обслуживания запросов  $b$  рассматривается многоканальная система с  $N$  обслуживающими приборами. По внешнему пуассоновскому потоку сигналов с параметром  $\delta$  начальное распределение ресурсов  $\tilde{m}_1$  (для 1-го класса трафика) может изменяться динамически и в произвольный момент времени равняться  $m_1$ . В случае отсутствия свободных приборов, соответ-

ствующих классу поступившего запроса, последний попадает в очередь длины  $R_1$  или  $R_2$  соответственно. С интенсивностью  $\varepsilon_1$  и  $\varepsilon_2$  запросы покидают очередь по причине превышения допустимого порога на экспоненциальное время задержки в очереди.

Опишем функционирование системы при помощи марковского процесса принятия решений в непрерывном времени, который задается четырьмя параметрами  $(S, A_s, Q_a(s, s'), R(s))$ . Состояние системы может быть представлено в виде вектора  $s = (m_1, l_1, l_2)$ , где  $m_1$  — текущее число приборов для 1-го класса трафика;  $l_1$  и  $l_2$  — число запросов 1-го и 2-го классов трафика в системе, в пространстве

$$S = \{s = (m_1, l_1, l_2) : m_1 = 0, \dots, N, l_1 = 0, \dots, m_1 + R_1, l_2 = 0, \dots, N - m_1 + R_2\}.$$

Перераспределение ресурсов инициируется при поступлении сигнала контроллера, а также проставлении ресурсов одного класса запросов и наличии ожидающих обслуживания запросов другого класса (будем называть такой сигнал «успешным», т. е.

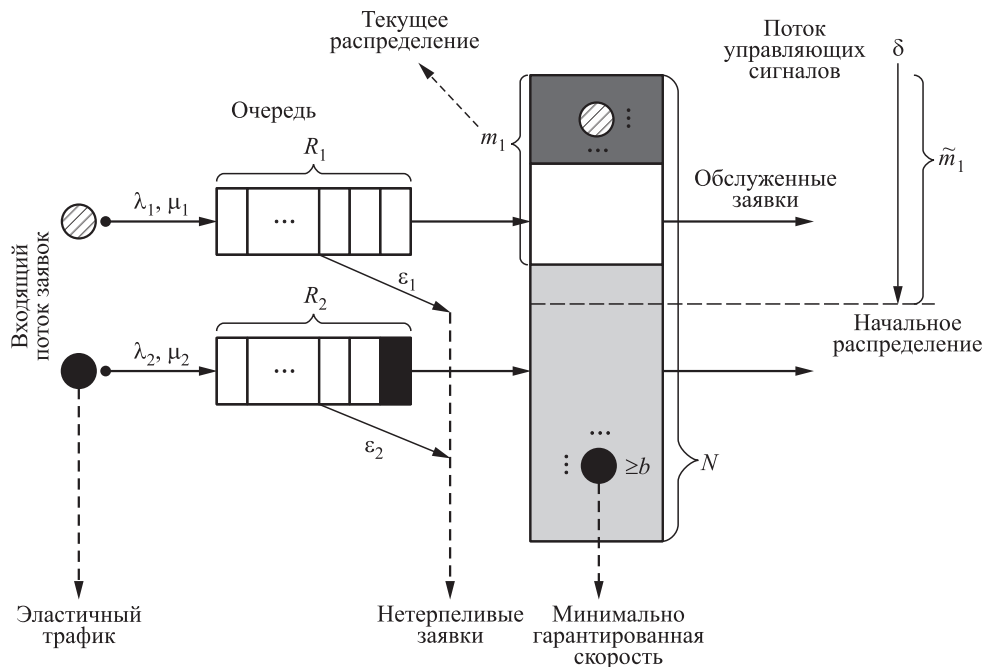


Рис. 2 Схема системы массового обслуживания

приведем к перераспределению). Если же все очереди пусты или, наоборот, полностью заняты обслуживающие ресурсы, при поступлении сигнала контроллера перераспределения ресурсов не произойдет.

Разбиение множества  $S$  на подмножества проводится в зависимости от возможности перераспределения ресурсов, а также соотношения между числом запросов в очередях и на приборах. Это состояния, в которых при поступлении сигнала произойдет  $S_\delta$  и не произойдет  $\bar{S}_\delta$  перераспределение ресурсов,  $S = S_\delta \cup \bar{S}_\delta$ . Подмножество состояний с числом приборов 1-го класса  $m_1$

$$S(m_1) = \{(i, l_1, l_2) \in S : i = m_1\},$$

$$S = \bigcup_{m_1=0}^N S(m_1),$$

тоже разбивается на те, в которых произойдет перераспределение ресурсов и в которых не произойдет:

$$S(m_1) = S_\delta(m_1) \cup \bar{S}_\delta(m_1).$$

В свою очередь,

$$S_\delta(m_1) = S_\delta^1(m_1) \cup S_\delta^2(m_1),$$

где  $S_\delta^1(m_1) = \{(i, l_1, l_2) \in S(m_1) : l_1 > m_1, l_2 < N - m_1\}$  — состояния, в которых при поступлении сигнала ресурсы 1-го класса трафика будут увеличены;  $S_\delta^2(m_1) = \{(i, l_1, l_2) \in S(m_1) : l_1 < m_1, l_2 > N - m_1\}$  — где будут увеличены ресурсы 2-го

класса. Следовательно, в случае наличия необходимого числа свободных ресурсов для обслуживания ожидающих запросов увеличение сегментов осуществляется на число ожидающих запросов

$$S_\delta^k(m_1) = S_\delta^{k1}(m_1) \cup S_\delta^{k2}(m_1),$$

$$S_\delta^{k1}(m_1) = \{(i, l_1, l_2) \in S_\delta^k(m_1) : l_1 + l_2 \leq N\}$$

или на все свободные ресурсы

$$S_\delta^{k2}(m_1) = \{(i, l_1, l_2) \in S_\delta^k(m_1) : l_1 + l_2 > N\}.$$

Задачей управления системой массового обслуживания будет выбор нового объема ресурса для 1-го класса трафика при поступлении сигнала контроллера. Множество допустимых стратегий управления зависит от состояния системы  $s \in S$  и определяется в виде:

$$A_s = \begin{cases} \{m_1, \dots, l_1\}, & s \in S_\delta^{11}; \\ \{m_1, \dots, N - l_2\}, & s \in S_\delta^{12}; \\ \{N - l_2, \dots, m_1\}, & s \in S_\delta^{21}; \\ \{l_1, \dots, m_1\}, & s \in S_\delta^{22}, \end{cases}$$

$$a(s) \in A_s, s \in S.$$

Далее зададим инфинитезимальную матрицу  $Q_a(s, s') = [q_a(s, s')]$  перехода из состояния  $s$  в состояние  $s'$  при управляющем действии  $a$ . Матрица является консервативной, т.е.  $q_a(s, s') \geq 0, s \neq s', q_a(s, s') = -q_a(s) = -\sum_{s' \neq s} q_a(s, s'), q_a(s) < \infty$ , где для  $s \neq s'$  элементы матрицы имеют вид:

$$q_a(\mathbf{s}, \mathbf{s}') = \begin{cases} \lambda_1, \mathbf{s}' = (m_1, l_1 + 1, l_2), \mathbf{s} : l_1 + 1 \leq R_1; \\ \lambda_2, \mathbf{s}' = (m_1, l_1, l_2 + 1), \mathbf{s} : l_2 + 1 \leq R_2; \\ \frac{m_1}{N} V\mu_1, \mathbf{s}' = (m_1, l_1 - 1, l_2), \\ \quad \mathbf{s} : m_1 > 0, l_1 > 0; \\ \frac{N - m_1}{N} V\mu_2, \mathbf{s}' = (m_1, l_1, l_2 - 1), \\ \quad \mathbf{s} : N - m_1 > 0, l_2 > 0; \\ (l_1 - m_1)\varepsilon_1, \mathbf{s}' = (m_1, l_1 - 1, l_2), \\ \quad \mathbf{s} : l_1 \geq m_1; \\ (l_2 - N + m_1)\varepsilon_2, \mathbf{s}' = (m_1, l_1, l_2 - 1), \\ \quad \mathbf{s} : l_2 \geq N - m_1; \\ \delta, \mathbf{s}' = (a, l_1, l_2) \text{ для } m_1 \leq a \leq l_1, \mathbf{s} \in S_\delta^{11}, \\ \quad \text{при } m_1 \leq a \leq N - l_2, \mathbf{s} \in S_\delta^{12}, \\ \quad \text{или } N - l_2 \leq a \leq m_1, \mathbf{s} \in S_\delta^{21}, \\ \quad \text{или } l_1 \leq a \leq m_1, \mathbf{s} \in S_\delta^{22}. \end{cases}$$

Наконец, система приносит вознаграждение в размере  $R(\mathbf{s})$  денежных единиц за единицу времени в течение всего периода ее пребывания в состоянии  $\mathbf{s}$ .

### 3 Функция вознаграждения

Напомним, что строится управляемая система массового обслуживания, параметром управления которой служит объем выделяемых ресурсов для 1-го класса трафика. Функция вознаграждения будет состоять из трех слагаемых, которые рассчитываются для каждого конкретного состояния системы. Каждая из трех компонент соответствует принципам оптимального распределения ресурсов, которые будут представлены ниже.

Принцип равного деления ресурсов, позволяющий определять число запросов, которые могли бы обслуживаться, но которым приходится ожидать из-за несправедливого деления ресурсов, выглядит следующим образом:

$$\chi(\mathbf{s}) = \begin{cases} \frac{N}{2} - m_1, & \mathbf{s} : m_1 < \frac{N}{2}, l_1 > m_1, \\ & l_1 - m_1 > \frac{N}{2} - m_1; \\ l_1 - m_1, & \mathbf{s} : m_1 < \frac{N}{2}, l_1 > m_1, \\ & l_1 - m_1 \leq \frac{N}{2} - m_1; \\ \frac{N}{2} - N + m_1, & \mathbf{s} : m_1 > \frac{N}{2}, l_2 > N - m_1, \\ & l_2 - N + m_1 > \frac{N}{2} - N + m_1; \\ l_2 - N + m_1, & \mathbf{s} : m_1 > \frac{N}{2}, l_2 > N - m_1, \\ & l_2 - N + m_1 \leq \frac{N}{2} - N + m_1. \end{cases}$$

Второй компонентой является принцип «успеха» перераспределения, который учитывает вероятность того, что поступит сигнал о перераспределении ресурсов, но перераспределения не произойдет из-за состояния системы, и определяется как

$$\beta_\delta(\mathbf{s}) = \delta (\delta + \lambda_1 \cdot 1 \{l_1 + 1 \leq R_1 + m_1\} + \lambda_2 \cdot 1 \{l_2 + 1 \leq R_2 + N - m_1\} + \frac{m_1}{N} V\mu_1 \cdot 1 \{m_1 > 0, l_1 > 0\} + \frac{N - m_1}{N} V\mu_2 \cdot 1 \{N - m_1 > 0, l_2 > 0\} + (l_1 - m_1)\varepsilon_1 \cdot 1 \{l_1 > m_1\} + (l_2 - N + m_1)\varepsilon_2 \cdot 1 \{l_2 > N - m_1\})^{-1}.$$

Третья компонента использования ресурсов, отражающая число запросов, которые могли бы обслуживаться, но находятся в ожидании из-за простаивания ресурсов, определяется как

$$\gamma(\mathbf{s}) = \begin{cases} N - m_1 - l_2, & \mathbf{s} \in S_\delta^{12}(m_1); \\ l_1 - m_1, & \mathbf{s} \in S_\delta^{11}(m_1); \\ m_1 - l_1, & \mathbf{s} \in S_\delta^{22}(m_1); \\ l_2 - N + m_1, & \mathbf{s} \in S_\delta^{21}(m_1). \end{cases}$$

Если в первой компоненте задержка обслуживания связана с «несправедливым» занятием всех ресурсов, то здесь, наоборот, с тем, что система выделила недостаточный объем ресурсов для обслуживания трафика.

В итоге функция вознаграждения вычисляется по формуле:

$$R(\mathbf{s}) = - \left( c_1 \chi(\mathbf{s}) \cdot 1 \{ \mathbf{s} \in \overline{S_\delta^1(m_1)} \} \left[ 1 \left\{ m_1 < \frac{N}{2}, \mathbf{s} \in S_\delta^1(m_1) \right\} + 1 \left\{ m_1 > \frac{N}{2}, \mathbf{s} \in S_\delta^2(m_1) \right\} \right] + c_2 \beta_\delta(\mathbf{s}) \cdot 1 \{ \mathbf{s} \in \overline{S_\delta} \} + c_3 \gamma(\mathbf{s}) \cdot 1 \{ \mathbf{s} \in S_\delta \} \right), \quad (1)$$

а функция среднего вознаграждения при управлении  $a$  определяется как

$$g_a = \sum_{\mathbf{s} \in S} R(\mathbf{s}) \pi_a(\mathbf{s}), \quad (2)$$

где  $\pi_a(\mathbf{s})$  — стационарное распределение вероятностей при управлении  $a$ . Используя заданную функцию среднего вознаграждения (2), будем решать оптимизационную задачу:

$$g^* = \max_a (g_a) \quad (3)$$

и получим оптимальную стратегию управления распределения ресурсов.

## 4 Оптимальная стратегия

Для решения оптимизационной задачи вида (3) в теории марковских процессов принятия решений используются итерационные алгоритмы, которые существенно сокращают вычисления. Применим итерационный алгоритм Р. Ховарда [10] для вычисления оптимальной стратегии. Система уравнений относительно среднего вознаграждения (2) и весовых коэффициентов  $v(s)$  имеет вид:

$$v(s) = \frac{1}{\sum_{s' \in S \setminus s} q_a(s, s')} \left[ R(s) + \sum_{s' \in S \setminus s} q_a(s, s') v(s') - g^a \right], \quad s \in S, \quad (4)$$

где  $q_a(s, s')$  — элементы инфинитезимальной матрицы  $Q_a(s, s')$ ;  $R(s)$  — функция вознаграждения (1). Здесь коэффициенты  $v(s)$  имеют смысл вознаграждений в случае, если система находится в состоянии  $s$ . Для поиска новой (улучшенной) стратегии задается целевая функция в виде

$$a(s) = \arg \max_{a \in A_s} \{v(a, l_1, l_2)\}, \quad s \in S_\delta. \quad (5)$$

С помощью алгоритма 1 зададим начальную стратегию перераспределения ресурсов (шаг 3), при которой ресурсы классов трафика увеличиваются либо на все свободные ресурсы, либо на число ожидающих обслуживания запросов. На первом этапе (шаг 4) вычисляются весовые коэффициенты для текущей стратегии управления. На втором этапе (шаг 5) улучшается стратегия изменением управляющих действий  $a(s)$  для каждого состояния  $s \in S_\delta$ . Этапы повторяются до тех пор, пока не будет найдена оптимальная стратегия.

## 5 Заключение

В статье построена модель обслуживания двух классов трафика с минимально гарантированной скоростью обслуживания при применении технологии нарезки сети, которая инициируется при поступлении сигнала контроллера, в виде управляемой системы массового обслуживания с эластичным трафиком и нетерпеливыми заявками. Система представлена в терминах марковского процесса принятия решений в непрерывном времени. В качестве принятия решений выступает выбор объема выделяемых ресурсов для одного из сегментов при динамическом перераспределении ресурсов. Функция вознаграждения представляет собой совокупность предложенных коэффициентов: соответствия начальному распределению ресурса, успеха перераспределения ресурса и использования ресурса.

## Литература

1. Ageev K., Sopin E., Samouylov K. Resource sharing model with minimum allocation for the performance analysis of network slicing // Information technologies and mathematical modelling. Queueing theory and applications / Eds. A. Dudin, A. Nazarov, A. Moiseev. — Communications in computer and information science ser. — Springer, 2021. Vol. 1391. P. 378–389. doi: 10.1007/978-3-030-72247-0\_28.
2. Yarkina N., Correia L. M., Moltchanov D., Gaidamaka Y., Samouylov K. Multi-tenant resource sharing with equitable-priority-based performance isolation of slices for 5G cellular systems // Comput. Commun., 2022. Vol. 188. P. 39–51. doi: 10.1016/j.comcom.2022.02.019.
3. Stepanov M. S., Stepanov S. N., Andrabi U., Petrov D., Ndayikunda J. The increasing of resource sharing efficiency in network slicing implementation // Distributed

---

### Алгоритм 1. Итерационный алгоритм вычисления оптимальной стратегии

---

- 1:  $n \leftarrow 0$
  - 2:  $v^{(0)}(s) = 0, s \in S$
  - 3:  $a^{(0)}(s) = \begin{cases} l_1, & s \in S_\delta^{11}; \\ N - l_2, & s \in S_\delta^{12}; \\ m_1 - l_2 - N + m_1, & s \in S_\delta^{21}; \\ l_1, & s \in S_\delta^{22}, \end{cases} \triangleright \text{начальная политика}$
  - 4: solve (система (4))
  - 5:  $a^{(n+1)}(s) = (\text{по формуле (5)}) \triangleright \text{улучшение политики}$
  - 6: if  $a^{(n+1)}(s) = a^{(n)}(s), s \in S$  then
  - 7:     return  $a^{(n+1)}(s), v^{(n)}(s), g^{(n)}$
  - 8: else  $n \leftarrow n + 1$ , go to step 5
  - 9: end if
-



- computer and communication networks / Eds. V. M. Vishnevskiy, K. E. Samouylov, D. V. Kozyrev. — Communications in computer and information science ser. — Springer, 2022. Vol. 1552. P. 18–35. doi: 10.1007/978-3-030-97110-6\_2.
4. Muhizi S., Ateya A. A., Muthanna A., Kirichek R., Koucheryavy A. A novel slice-oriented network model // Distributed computer and communication networks / Eds. V. Vishnevskiy, D. Kozyrev. — Communications in computer and information science ser. — Springer, 2018. Vol. 919. P. 421–431. doi: 10.1007/978-3-319-99447-5\_36.
  5. Yarkina N., Gaidamaka Y., Correia L. M., Samouylov K. An analytical model for 5G network resource sharing with flexible SLA-oriented slice isolation // Mathematics, 2020. Vol. 8. No. 7. P. 1177. 19 p. doi: 10.3390/math8071177.
  6. Shu Z., Taleb T., Song J. Resource allocation modeling for fine-granular network slicing in beyond 5G systems // Oper. Res., 2022. Vol. 70. No. 2. P. 349–363. doi: 10.1587/transcom.2021WWI0002.
  7. Kochetkova I., Vlaskina A., Burtseva S., Savich V., Hosenk J. Analyzing the effectiveness of dynamic network slicing procedure in 5G network by queuing and simulation models // 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. — Springer, 2020. Vol. 12525. P. 71–85. doi: 10.1007/978-3-030-65726-0\_7.
  8. Кочеткова И. А., Власкина А. С., Ву Н. Н., Шоргин В. С. Система массового обслуживания с управляемым по сигналам перераспределением приборов для анализа нарезки ресурсов сети 5G // Информатика и её применения, 2021. Т. 15. Вып. 3. С. 91–97. doi: 10.14357/19922264210312.
  9. Efrosinin D., Kochetkova I., Samouylov K., Stepanova N. Algorithmic analysis of a two-class multi-server heterogeneous queueing system with a controllable cross-connectivity // Analytical and stochastic modelling techniques and applications / Eds. M. Gribaudo, E. Sopin, I. Kochetkova. — Lecture notes in computer science ser. — Springer, 2020. Vol. 12023. P. 1–17. doi: 10.1007/978-3-030-62885-7\_1.
  10. Ховард Р. А. Динамическое программирование и марковские процессы / Пер. с англ. — М.: Советское радио, 1964. 192 с. (Howard R. A. Dynamic programming and Markov processes. — Cambridge, MA, USA: MIT Press, 1960. 136 p.)

Поступила в редакцию 15.07.22

## CONTROLLABLE QUEUING SYSTEM WITH ELASTIC TRAFFIC AND SIGNALS FOR ANALYZING NETWORK SLICING

A. S. Vlaskina<sup>1</sup>, S. A. Burtseva<sup>1</sup>, I. A. Kochetkova<sup>1,2</sup>, and S. Ya. Shorin<sup>2</sup>

<sup>1</sup>Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

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

**Abstract:** A controllable queuing system with impatient elastic traffic and minimum bit rate guarantee transmission is built to analyze the dynamic network slicing in wireless network. Resource management of allocation between two segments is carried out by controller signal and in the case of an idle resource of one segment and waiting users of another. Depending on the state of the system, a decision is made how to allocate the resource. The decision-making is based on the efficiency indicators of slicing taking into account the resource usage of each of the segments, deviation from the recommended "initial" slicing, and additional signal load from the controller. The corresponding objective control function is proposed and an iterative algorithm for calculating the parameters of the optimal resource allocation policy is applied.

**Keywords:** 5G; network slicing; controllable queuing system; Markov decision process; elastic traffic

**DOI:** 10.14357/19922264220312

### Acknowledgments

The paper has been supported by the RUDN University Strategic Academic Leadership Program.

### References

1. Ageev, K., E. Sopin, and K. Samouylov. 2021. Resource sharing model with minimum allocation for the performance analysis of network slicing. *Information technologies and mathematical modelling. Queueing theory and applications*. Eds. A. Dudin, A. Nazarov, and A. Moiseev. Communications in computer and information science ser. Springer. 1391:378–389. doi: 10.1007/978-3-030-72247-0\_28.

2. Yarkina, N., L. M. Correia, D. Moltchanov, Y. Gaidamaka, and K. Samouylov. 2022. Multi-tenant resource sharing with equitable-priority-based performance isolation of slices for 5G cellular systems. *Comput. Commun.* 188:39–51. doi: 10.1016/j.comcom.2022.02.019.
3. Stepanov, M. S., S. N. Stepanov, U. Andrabi, D. Petrov, and J. Ndayikunda. 2022. The increasing of resource sharing efficiency in network slicing implementation. *Distributed computer and communication networks*. Eds. V. M. Vishnevskiy, K. E. Samouylov, and D. V. Kozyrev. Communications in computer and information science ser. Springer. 1552:18–35. doi: 10.1007/978-3-030-97110-6\_2.
4. Muhizi, S., A. A. Ateya, A. Muthanna, R. Kirichek, and A. Koucheryavy. 2018. A novel slice-oriented network model. *Distributed computer and communication networks*. Eds. V. Vishnevskiy and D. Kozyrev. Communications in computer and information science ser. Springer. 919:421–431. doi: 10.1007/978-3-319-99447-5\_36.
5. Yarkina, N., Y. Gaidamaka, L. M. Correia, and K. Samouylov. 2020. An analytical model for 5G network resource sharing with flexible SLA-oriented slice isolation. *Mathematics* 8(7):1177. 19 p. doi: 10.3390/math8071177.
6. Shu, Z., T. Taleb, and J. Song. 2022. Resource allocation modeling for fine-granular network slicing in beyond 5G systems. *Oper. Res.* 70(2):349–363. doi: 10.1587/transcom.2021WWI0002.
7. Kochetkova, I., A. Vlaskina, S. Burtseva, V. Savich, and J. Hosek. 2020. Analyzing the effectiveness of dynamic network slicing procedure in 5G network by queuing and simulation models. *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. Springer. 12525:71–85. doi: 10.1007/978-3-030-65726-0\_7.
8. Kochetkova, I. A., A. S. Vlaskina, N. N. Vu, and V. S. Shorgin. 2021. Sistema massovogo obsluzhivaniya s upravlyaemym po signalam pereraspredeleniem priborov dlya analiza narezki resursov seti 5G [Queuing system with signals for dynamic resource allocation for analyzing network slicing in 5G networks]. *Informatika i ee Primeneniya — Inform. Appl.* 15(3):91–97. doi: 10.14357/19922264210312.
9. Efrosinin, D., I. Kochetkova, K. Samouylov, and N. Stepanova. 2020. Algorithmic analysis of a two-class multi-server heterogeneous queueing system with a controllable cross-connectivity. *Analytical and stochastic modelling techniques and applications*. Eds. M. Gribaudo, E. Sopin, and I. Kochetkova. Lecture notes in computer science ser. Springer. 12023:1–17. doi: 10.1007/978-3-030-62885-7\_1.
10. Howard, R. A. 1960. *Dynamic programming and Markov processes*. Cambridge, MA: MIT Press. 136 p.

Received July 15, 2022

## Contributors

**Vlaskina Anastasia S.** (b. 1995) — assistant professor, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; vlaskina-as@rudn.ru

**Burtseva Sofiya A.** (b. 1999) — student, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; sofiya.burceva@inbox.ru

**Kochetkova Irina A.** (b. 1985) — Candidate of Science (PhD) in physics and mathematics, associate professor, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; kochetkova-ia@rudn.ru

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

# МЕТАДАННЫЕ В ЗАЩИЩЕННОМ ЭЛЕКТРОННОМ ДОКУМЕНТООБОРОТЕ

А. А. Грушо<sup>1</sup>, Н. А. Грушо<sup>2</sup>, Е. Е. Тимонина<sup>3</sup>

**Аннотация:** Работа посвящена исследованию проблем информационной безопасности (ИБ) сетевых систем электронного документооборота (ЭДО) в распределенной информационной системе (РИС) предприятия. Электронный документооборот представляет собой защищенную распределенную систему информационных технологий (ИТ), каждая из которых порождается частью организационной структуры предприятия, т. е. существует защищенная связь, обеспечивающая функционал документооборота. Для обеспечения ИБ использована концепция метаданных (МД), определенная ранее в работах авторов. В отличие от ранних приемов использования МД, в ЭДО применяется множество вариантов ИТ, определяемых подмножеством участвующих в работе над документом пользователей и структурой поручений, порождаемых исполнением документов. Авторами предложена модульная конструкция МД, которая позволяет решать проблему синтеза и управления многообразием возникающих вариантов ИТ ЭДО и обеспечением их ИБ.

**Ключевые слова:** информационная безопасность; метаданные; электронный документооборот

**DOI:** 10.14357/19922264220313

## 1 Введение

Одним из распространенных классов РИС стали сетевые системы [1], поэтому проблемы контроля и обеспечения ИБ в сетевой системе особенно актуальны. Для исследования этих проблем выбрана сетевая система ЭДО РИС предприятия.

Электронный документооборот представляет собой защищенную распределенную систему ИТ, в которой каждая ИТ-структура порождается частью организационной структуры предприятия и существует защищенная связь, обеспечивающая функционал ЭДО.

Функционал ЭДО состоит из:

- защищенных функций взаимодействия с внешним окружением, не принадлежащим РИС предприятия (получение и отсылка документов);
- базы данных документов (архив документов);
- функций управления документооборотом;
- функций формирования поручений;
- функций доставки поручений исполнителям, контроля исполнения поручений и времени исполнения;
- функций идентификации и аутентификации, функций электронной подписи (ЭП), удосто-

веряющего центра, центра генерации и распределения криптографических ключей;

- функций обеспечения целостности и неотказуемости движения документов.

Предполагается следующая модель угроз.

### 1. Угрозы конфиденциальности:

- содержание документов и файлов с отчетами по поручениям, связанным с документами, считается конфиденциальной информацией, подлежащей защите;
- полный маршрут движения документа считается конфиденциальной информацией, подлежащей защите;
- содержание документов архива и текущих баз данных ЭДО считается конфиденциальной информацией, подлежащей защите.

### 2. Угрозы целостности:

- содержание документов и файлов с отчетами по поручениям, связанным с документами, считается информацией, подлежащей защите целостности;
- целостность маршрута движения документа должна быть защищена;
- должна быть обеспечена безотказность при работе с документами;

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

<sup>2</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, info@itake.ru

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

- функции идентификации и аутентификации сотрудников (пользователей компьютерных рабочих мест), работающих с документами, не должны нарушаться.

### 3. Угрозы доступности:

- доступность реализации маршрутов документов должна быть обеспечена;
- должны быть зарезервированы средства доступности к документам;
- должен быть обеспечен контроль времени работы над поручениями, связанными с каждым документом.

### 4. Угрозы потери ответственности при работе с документами:

- необходимо обеспечить идентификацию авторства любых действий при работе с документами или отсутствия определенных действий.

Для реализации ЭДО в РИС необходима дополнительная информация и инструменты защиты ЭДО как распределенной ИТ. Для обеспечения ИБ распределенных ИТ в работах [2, 3] разработан подход, основанный на МД. Суть подхода состоит в том, что управление соединениями и функции контроля выполнения ИТ реализуются централизованно на основе модели ИТ [4]. Такое управление организовано на знаниях о взаимодействиях субъектов ИТ и их расположении на хостах распределенной системы компьютеров. Эти знания формируются и хранятся в задаче  $M$ , изолированной от хостов и функций, которые на них выполняются. Такая изоляция служит основой безопасности управления выполнением ИТ, использующей даже небезопасную сеть связи. Используя данные задачи  $M$ , следующая задача  $N$  управляет соединениями и контролем развития ИТ. Обмен информацией между задачами  $M$  и  $N$  не является односторонним обменом, но это не может нарушить безопасность управления [5].

В рассматриваемом случае отличие от исследованной ранее модели ИТ состоит в том, что сама ИТ ЭДО представляет собой динамически изменяемую структуру, решающую ограниченный набор задач с постоянно меняющимся набором исполнителей, поэтому постоянно меняется модель маршрутов выполнения ИТ.

Цель работы состоит в построении архитектуры ЭДО на основе МД и обоснование безопасности предложенной модели ЭДО.

## 2 Метаданные в электронном документообороте

Рассмотрим следующую дополнительную информацию (МД) для автоматической реализации функций ЭДО.

1. Матрица  $M_1$  —  $(0, 1)$ -матрица, строки и столбцы которой помечены идентификаторами пользователей  $U_1, U_2, \dots$ , потенциально участвующих в ЭДО, где 1 на месте  $(U_k, U_m)$  означает право передавать поручение от пользователя  $U_k$  пользователю  $U_m$ . Поручением может быть реализация поручения вышестоящего начальника либо продолжение ИТ, которая утверждена в организации.
2. Далее пользователь отождествляется с рабочей станцией, на которой этот пользователь работает, рабочая станция является хостом сети организации, на которой установлен HSM (Host Security Module) [6]. В HSM установлен уникальный для пользователя ключ общего для всех алгоритма шифрования и комплекс для проверки ЭП и установки ЭП данного пользователя. При работе HSM должна проводиться идентификация и аутентификация пользователя.
3. Для каждого документа при создании поручения также устанавливается глубина распространения документа, т. е. число пользователей, которые могут далее передавать сам документ. По умолчанию глубина равна 1 или соответствует нормативу для исполняемой ИТ.
4. Для каждого документа при создании поручения также устанавливается время исполнения поручения. По умолчанию время исполнения не регламентируется или соответствует нормативу для исполняемой ИТ.
5. Первое поручение на дальнейшую работу с поступившим извне документом устанавливает пользователь, указанный в адресе поступившего документа.
6. Каждое поручение включает два изолированных информационных фрагмента:
  - (а) первый фрагмент содержит: указатель, кому дается поручение, время выдачи поручения и время его исполнения (по умолчанию время не регламентируется);
  - (б) второй фрагмент содержит документ (опционально) и сообщение с содержанием поручения или отчета.

При отсутствии содержания в сообщении по умолчанию предполагается ознакомление или априорное знание пользователем, получившим



документ, что нужно делать далее. Последнее чаще относится к выполнению утвержденной в организации ИТ. Сообщение и документ подписываются ЭП. Первый фрагмент всегда присутствует в открытом виде, а второй всегда защищен шифром и ЭП.

7. Если глубина распространения документа исчерпана или поручения выполнены до наступления предельной глубины, а других нет, то документ и все сообщения поступают в архив или документ и приложенное к нему непустое сообщение (отчет) поступает к источнику поручения.

### 3 Архитектура электронного документооборота

В сетевидной системе, по определению, существует центр управления ЭДО. Этот центр в ЭДО состоит из четырех подсистем. Первая подсистема  $Y$  управляет поручениями и движением документов; вторая подсистема решает задачу  $M$  в метаданных, т.е. управляет адресацией движения документов и временем исполнения (составляет и контролирует первый фрагмент объекта, направляемого для выполнения поручения); третья подсистема соответствует задаче  $N$  в МД и управляет соединениями с пользователями; четвертая подсистема управляет криптографией, т.е. генерирует и распределяет ключи в HSM и контролирует генерацию и функции ЭП (выполняет функции удостоверяющего центра).

Основной принцип в обеспечении ИБ ЭДО состоит в том, что любое взаимодействие с пользователем состоит из простого ориентированного цикла: задача  $N$  организует связь с пользователем и отправляет ему документы и сообщения, которые сопровождали последовательность сообщений, предшествовавших данной связи с пользователем.

Вся передаваемая информация второго фрагмента зашифрована на ключе пользователя и подписана ЭП третьей подсистемы (задача  $N$ ). Первый фрагмент фактически образует заголовок передаваемого объекта.

Использование криптографических функций относится к задаче  $N$  в управлении ЭДО и к блокам HSM у пользователей. Задача  $N$  имеет перечень ключей пользователей. Перед организацией сеанса с пользователем весь фрагмент содержания шифруется на ключе пользователя и подписывается. Реакция пользователя после расшифрования и проверки подписи состоит в том, чтобы создать сообщение и отправить его задаче  $N$  в зашифрован-

ном виде, включая хеш принятого сообщения, подписав сообщение ЭП пользователя. Этот протокол также обеспечивает аутентификацию пользователя. Задача  $N$  проверяет ЭП, расшифровывает сообщение и открывает хеш от посланного с поручением сообщения, что служит аутентификацией пользователя, приславшего сообщение. Затем задача  $N$  направляет полученное сообщение задаче  $M$ . Задача  $M$  знает, кто поручил передать пользователю информацию, и направляет ответ в подсистему  $Y$  этому пользователю — создателю поручения. Здесь возникает причинно-следственная связь между отправителем и получателем поручения.

Основной элемент архитектуры связан с созданием и исполнением поручений. Предлагается создавать эту подсистему в  $Y$  на модульном принципе. Каждый пользователь, который получает поручение, автоматически с поручением получает изолированный программный модуль с уникальным идентификатором. Вместе с поручением в этот модуль записываются исходный документ и все сообщения, которые связаны с предыдущей работой других пользователей.

Создание поручения и сообщения к нему может потребовать дополнительной информации из архива или из других источников. Информация от внешних источников обрабатывается на предмет безопасности по другим технологиям, не входящим в ЭДО (например, отражается на отдельном компьютере).

Каждый модуль имеет два входа и один выход. Вход, порождающий модуль, содержит поручение пользователю. Второй вход служит для получения реакции получателя поручения (отчетов может быть много). Выход служит для передачи информации следующему исполнителю или для передачи команды завершения работы. В последнем случае исходный документ и все последующие сообщения заносятся в архив, который сам устроен на принципах блокчейна [7]. Отработанный модуль уничтожается.

Порождение поручений в общем случае описывается корневым деревом, вершинами которого служат пользователи, а именно: пользователи, получающие поручения с метками времени, и пользователи, создающие эти поручения (идущие в соответствующие модули); ребра дерева соответствуют поручениям с уникальными идентификаторами и меткой времени, а также указателем идентификатора документа (идентификатора модуля), откуда поступило поручение.

Даже если пользователи повторяются, то циклы не могут возникнуть из-за меток времени и идентификаторов документов и сообщений, поступивших



от источников поручений. Такие графы носят временный характер и уничтожаются после окончания работы с документом и сохранением его следа в архиве. Граф ограничен возможностями матрицы  $M_1$  и отсутствием пользователей, которым должно быть направлено поручение. В этих случаях пользователь должен изменить адресацию поручения и найти новый способ дальнейшей работы над поручениями предыдущих пользователей.

## 4 Информационная безопасность электронного документооборота

Конфиденциальность содержания документов обеспечивается ответственностью пользователей и выполнением правил доступа. В частности, ответственность за разглашение конфиденциальной информации следующему участнику цепочки поручений несет пользователь, дающий этому участнику поручение. Когда ответственность определена, можно опираться на правила политик ограничения доступа, пользуясь достоверно определяемой ответственностью за нарушение этих правил. Индивидуальные ключи каждого пользователя обеспечивают конфиденциальность при сетевом взаимодействии. Работа в индивидуальном модуле изолирует пользователя от других пользователей, возможно заинтересованных в ознакомлении с конфиденциальной информацией. Защита конфиденциальной информации в графах поручений других полностью автоматических компонентах зависит от проверенности программного обеспечения в подсистеме  $Y$  и на рабочих станциях пользователей, физической защиты подсистемы  $Y$  и HSM. Открытый текст документов и сообщений к ним появляется у нескольких пользователей. Кража конфиденциальной информации одним из них создает условия отсутствия ответственности за кражу. Эта проблема решается ответственностью пользователя, давшего поручение. Однако поиск и выявление истинного виновника кражи требуют дополнительной информации, выходящей за пределы ЭДО.

Полный маршрут исполнения поручений доступен только на уровне  $Y$ , и опять считается, что этот уровень безопасен. Защита маршрута не всегда служит качественной защитой от разрушения персональной ответственности, так как пользователю, совершившему кражу, достоверно не известно, кто еще имел доступ к конфиденциальной информации в цепочке его уровня и далее. Однако постороннему пользователю маршрут движения документа может дать частичную информацию о содержании конфиденциальной информации.

Конфиденциальность баз данных и архива обеспечивается средствами, не входящими в ЭДО.

Защита целостности обеспечивается сохранением всех документов и их ЭП в ходе выполнения ИТ и при их хранении в архиве. Блокчейн-технологии обеспечивают невозможность изъятия документов или их частей. Наличие ЭП и стандартных протоколов обеспечивает безотказность всех компонентов технологии ЭДО.

Целостность графов поручений обеспечивается безопасностью подсистемы  $Y$  и надежностью автоматических действий управления ЭДО. Целостность и надежность идентификации и аутентификации определяется надежностью задач  $M$  и  $N$ , а также работой HSM.

Доступность обеспечивается надежностью всех компонент ЭДО. Выявление сбоев доступности на сетевом и уровне рабочих станций возможно за счет модульности взаимодействий пользователя, дающего поручения, и пользователей-исполнителей поручений. В условиях стихийных или технологических катастроф всегда возможен переход к ручному исполнению документооборота.

Нарушения ответственности частично рассматривались выше. Модульный характер причинно-следственных отношений дающего поручения и исполняющего поручения позволяют создавать данные, которые в памяти блокчейна однозначно позволяют выявить нарушения правил работы с документами, в частности пропуски выполнения необходимых действий, и определить виновного.

Приведенный анализ показывает, что перечисленные штатные средства защиты информации и особенности построенной архитектуры позволяют обеспечить ИБ от заданного перечня угроз. Использование концепции МД, специальной архитектуры ЭДО, криптографии, модульной конструкции управления позволило обосновать безопасность модели ЭДО.

## 5 Заключение

В работе построена модель ЭДО в виде множества защищенных ИТ в РИС. Для обеспечения ИБ использована концепция МД, определенная ранее в работах авторов. В отличие от ранних приемов использования МД, в ЭДО используется множество вариантов ИТ, определяемых подмножеством задействованных пользователей и структурой поручений, порождаемых исполнением документов.

Авторами предложена модульная конструкция МД, которая позволяет решать проблему синтеза и управления многообразием возникающих вариантов ИТ ЭДО и обеспечением их ИБ. Пользовате-

ли через соответствующие модули создают поручения другим пользователям. Отображение взаимодействий этих пользователей формирует корневые деревья. Меняющийся состав модулей и деревьев позволяет реализовать различные варианты работы с документами. Ребра деревьев соответствуют поручениям и отчетам об их выполнении. При этом отчеты могут носить комплексный характер, т. е. собираться из сообщений, составленных всеми пользователями, выполнявшими поручения. Иными словами, сборка каждого отчета возможна только по цепочкам путей от вершин к корню дерева.

## Литература

1. Володин П. С., Золотарева Е. С., Мешков А. М. Сетевое управление: понятие и сущность // Ж. У. Экономика. Управление. Финансы, 2018. Вып. 2. С. 51–64.
2. Grusho A., Grusho N., Zabezhalo M., Zatsarinny A., Timonina E. Information security of SDN on the basis of metadata // Computer network security / Eds. J. Rak, J. Bay, I. V. Kotenko, *et al.* — Lecture notes in computer science ser. — Springer, 2017. Vol. 10446. P. 339–347.
3. Grusho A., Grusho N., Timonina E. Information flow control on the basis of meta data // Distributed computer and communication networks / Eds. V. M. Vishnevskiy, K. E. Samouylov, D. V. Kozhyrev. — Lecture notes in computer science ser. — Springer, 2019. Vol. 11965. P. 548–562.
4. Grusho A. A., Timonina E. E., Shorgin S. Ya. Modelling for ensuring information security of the distributed information systems // 31st European Conference on Modelling and Simulation Proceedings. — Dudweiler, Germany: Digitaldruck Pirrot GmbH, 2017. P. 656–660.
5. Грушо А. А., Применко Э. А., Тимонина Е. Е. Теоретические основы компьютерной безопасности. — М.: Академия, 2009. 272 с.
6. Meyer C. H., Matyas S. M. Cryptography: A new dimension in computer data security — a guide for the design and implementation of secure systems. — Hoboken, NJ, USA: Wiley, 1982. 755 p.
7. Grusho A., Piskovsky V., Zabezhalo M. System to track access in digital economy systems // Law Digital Technologies, 2021. Vol. 1. Iss. 1. P. 10–18.

Поступила в редакцию 15.07.22

---



---

## METADATA IN SECURE ELECTRONIC DOCUMENT MANAGEMENT

A. A. Grusho, N. A. Grusho, and E. E. Timonina

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

**Abstract:** The paper is devoted to the study of information security problems of network-centric electronic document management systems in the distributed information system of the enterprise. Electronic document management is a secure distributed system of information technologies, each of which is generated by part of the organizational structure of the enterprise, i. e., there is a secure connection that provides the functionality of document management. To ensure information security, the concept of metadata defined earlier in the works of the authors was used. Unlike earlier works that used the concept of metadata, electronic document management uses many variants of information technologies determined by a subset of users participating in the work on the document and the structure of orders generated by the execution of documents. New in the work is the modular design of metadata which allows solving the problem of synthesis and management of a variety of emerging variants of information technologies of electronic document management and ensuring their information security.

**Keywords:** information security; metadata; electronic document management

**DOI:** 10.14357/19922264220313

## References

1. Volodin, R. S., E. S. Zolotareva, and A. M. Meshkov. 2018. Setetsentricheskoe upravlenie: ponyatie i sushchnost' [Network-centric control: Concept and essence]. *Zh. U. Ekonomika. Upravlenie. Finansy* [J. U. Economy. Management. Finance] 2:51–64.
2. Grusho, A., N. Grusho, M. Zabezhalo, A. Zatsarinny, and E. Timonina. 2017. Information security of SDN on the basis of meta data. *Computer network security*. Eds. J. Rak, J. Bay, I. V. Kotenko, *et al.* Lecture notes in computer science ser. Springer. 10446:339–347.
3. Grusho, A. A., N. A. Grusho, and E. E. Timonina. 2019. Information flow control on the basis of meta data. *Distributed computer and communication networks*. Eds. V. M. Vishnevskiy, K. E. Samouylov, and D. V. Kozhyrev. Lecture notes in computer science ser. Springer. 11965:548–562.
4. Grusho, A. A., E. E. Timonina, and S. Ya. Shorgin. 2017. Modelling for ensuring information security of the

- distributed information systems. *31st European Conference on Modelling and Simulation Proceedings*. Dudweiler, Germany: Digitaldruck Pirrot GmbH. 656–660.
5. Grusho, A. A., E. A. Primenko, and E. E. Timonina. 2009. *Teoreticheskie osnovy komp'yuternoy bezopasnosti* [Theoretical foundations of computer security]. Moscow: Akademiya. 272 p.
  6. Meyer, C. H., and S. M. Matyas. 1982. *Cryptography: A new dimension in computer data security — a guide for the design and implementation of secure systems*. Hoboken, NJ: Wiley. 755 p.
  7. Grusho, A., V. Piskovsky, and M. Zabezhailo. 2021. System to track access in digital economy systems. *Law Digital Technologies* 1(1):10–18.

Received July 15, 2022

## Contributors

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

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

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

# КИНЕМАТИЧЕСКИЕ МОДЕЛИ ЗАДАЧ ПРЕСЛЕДОВАНИЯ НА ПЛОСКОСТИ МЕТОДАМИ ПАРАЛЛЕЛЬНОГО СБЛИЖЕНИЯ И ПОГОНИ

А. А. Дубанов<sup>1</sup>, В. А. Нефедова<sup>2</sup>

**Аннотация:** Рассматриваются модели задачи преследования на плоскости методами параллельного сближения и погони. Цель работы — модификация методов параллельного сближения и погони на случаи, когда в момент начала преследования вектор скорости преследователя направлен не в точку на окружности Аполлония в методе параллельного сближения и не на цель в методе погони. В рассматриваемых моделях преследователь не может мгновенно изменить направление движения. Было наложено условие, что радиус кривизны траектории движения преследователя не может быть меньше определенной величины. Предлагаемый метод основан на том, что преследователь, выбирая шаг на этапе итераций, будет стараться следовать прогнозируемым траекториям. По материалам статьи написаны тестовые программы, которые рассчитывают траектории преследователя, учитывая изложенные условия. Выполненные анимированные изображения визуализируют изменение координат преследователя, цели и прогнозируемых траекторий во времени.

**Ключевые слова:** цель; преследователь; траектория; сближение; моделирование

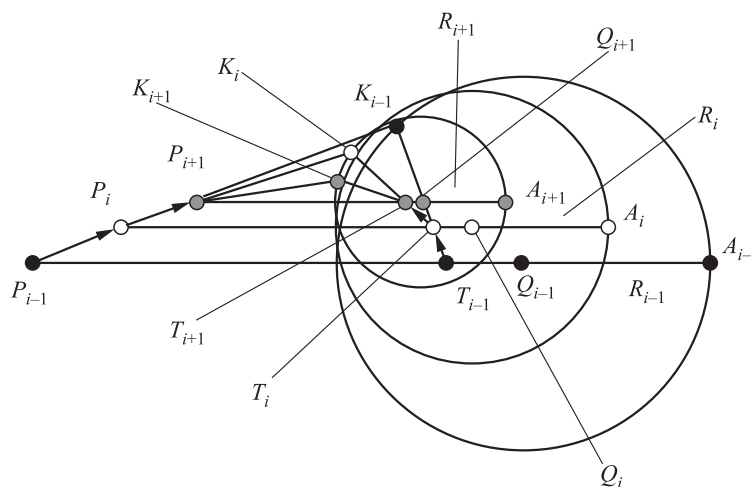
**DOI:** 10.14357/19922264220314

## 1 Введение

При расчете траекторий преследователя в задачах преследования в пространстве и на плоскости используются методы параллельного сближения и погони. В описании задачи преследования методом параллельного сближения в трудах Л. О. Петросяна [1, 2] направление вектора скорости в точке нахождения преследователя  $P_i$  и направление вектора скорости в точке нахождения цели  $T_i$  пере-

секаются в точке  $K_i$ , принадлежащей окружности Аполлония (рис. 1), соответствующей данному моменту времени.

Для точек  $P$  и  $T$  точка  $K$  окружности Аполлония характерна тем, что отношение длин  $|PK|/|QK| = |V_P|/|V_T|$  есть отношение модулей скоростей преследователя и цели. При дискретном моделировании точек траектории преследователя  $\{P_i\}$  можно предложить следующую итерационную схему (рис. 2):



**Рис. 1** Итерационная схема метода параллельного сближения

<sup>1</sup>Бурятский государственный университет, alandubanov@mail.ru

<sup>2</sup>Бурятский государственный университет, emelyanovatorik@gmail.com

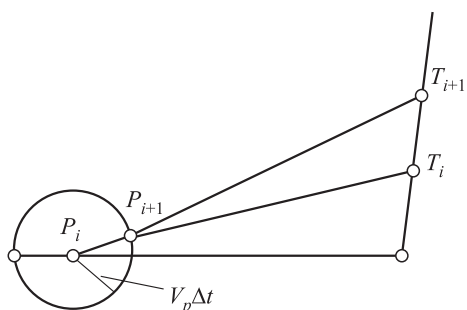


Рис. 2 Итерационная схема метода погони

$$P_i = P_{i-1} + V_P \Delta T \frac{K_{i-1} - P_{i-1}}{|K_{i-1} - P_{i-1}|}.$$

Радиусы окружностей Аполлония определяются выражением

$$R_i = \frac{V_T^2}{V_P^2 - V_T^2} |T_i - P_i|.$$

Центры окружностей Аполлония рассчитываются по формуле:

$$Q_i = T_i + \frac{V_i^2}{V_P^2 - V_T^2} (T_i - P_i).$$

Координаты точки  $K_i$  получаются в результате решения системы уравнений относительно непрерывного параметра  $t$ :

$$(K_i - Q_i)^2 = R_i^2;$$

$$K_i = T_i + V_T \frac{T_{i+1} - T_i}{|T_{i+1} - T_i|} t.$$

Такова одна из моделей построения траектории преследователя в методе параллельного сближения на плоскости. При этом требуется, чтобы направления векторов движения преследователя и цели пересекались в точках, принадлежащих окружностям Аполлония.

Для метода погони в задаче преследования на плоскости с постоянной скоростью характерно то, что вектор скорости преследователя точно совпадает с направлением на цель. В такой постановке задача имеет как непрерывную модель для решения, так и квазидискретную. В непрерывной модели решение основывается на численном или аналитическом решении системы дифференциальных уравнений:

$$(T_x(t) - P_x(t)) \frac{dP_y(t)}{dt} = (T_y(t) - P_y(t)) \frac{dP_x(t)}{dt}; \quad (1)$$

$$\left(\frac{dP_x(t)}{dt}\right)^2 + \left(\frac{dP_y(t)}{dt}\right)^2 = V^2, \quad (2)$$

где  $\begin{bmatrix} P_x \\ P_y \end{bmatrix}$  — координаты преследователя;  $\begin{bmatrix} T_x \\ T_y \end{bmatrix}$  — координаты цели;  $V$  — скорость цели.

Уравнение (1) системы говорит о том, что вектор скорости преследователя сонаправлен с вектором линии, соединяющей преследователя и цель. Уравнение (2) системы говорит о постоянстве модуля скорости.

В качестве простого примера квазидискретной модели можно привести одну из итерационных схем, которая вычисляет координаты точек следующего положения преследователя (см. рис. 2):

$$P_{i-1} = P_i + V_P \Delta t (T_i - P_i),$$

где  $V_P$  — скорость преследователя;  $\Delta t$  — период дискретизации. Как видно из постановки задачи преследования методом погони с постоянной скоростью, вектор скорости преследователя должен быть всегда направлен на цель, даже в момент начала преследования. В обеих описанных моделях задается начальное направление вектора скорости преследователя. Цель данной статьи — представить модифицированные модели преследования, в которых направление вектора скорости преследователя произвольно.

## 2 Кинематическая модель метода параллельного сближения

Итерационную схему, представленную на рис. 1, можно интерпретировать иначе. На рис. 3 представлен итерационный процесс, определяющий координаты точки  $P_i$  при известных координатах точек  $P_{i-1}$ ,  $T_{i-1}$  и  $T_i$  и скоростей преследователя и цели  $V_P$  и  $V_T$  соответственно.

Сначала определяется единичный вектор

$$\vec{r} = \frac{T_{i-1} - P_{i-1}}{|T_{i-1} - P_{i-1}|}.$$

Координаты точки  $T_i = T_{i-1} + \vec{V}_T \Delta T$ , где  $\Delta T$  — период дискретизации, предопределены поведением цели. Значит, прямую линию, которая будет соединять точки  $P_i$  и  $T_i$ , можно представить в виде

$$L(\mu) = T_i + \mu \vec{r}.$$

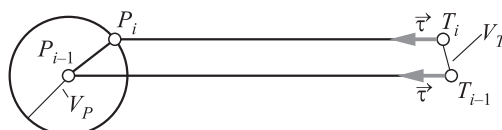


Рис. 3 Интерпретация итерационной схемы метода параллельного сближения



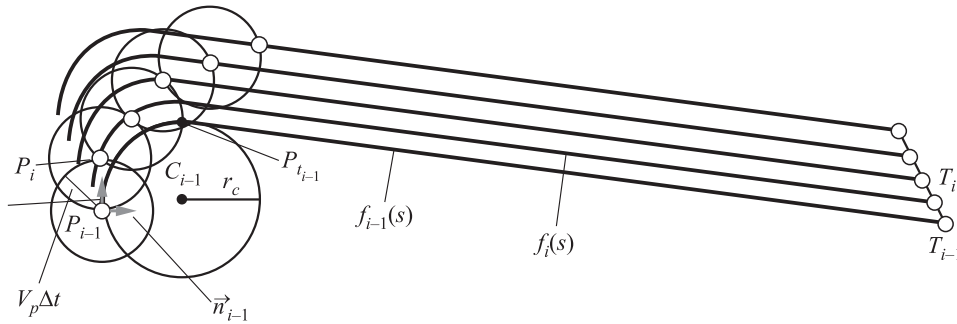


Рис. 4 Квазидискретная модель параллельного сближения

Тогда координаты  $P_i$  следующего шага итераций траектории преследователя можно интерпретировать как точку пересечения окружности радиуса  $V_P \Delta T$  с центром в точке  $P_{i-1}$  и прямой линии  $L(\mu)$ :

$$\begin{aligned} (L(\mu) - P_{i-1})^2 &= (V_P \Delta T)^2; \\ L(\mu) &= T_i + \mu \vec{\tau}. \end{aligned}$$

Решение вышеприведенной системы уравнений относительно параметра  $\mu$  даст такое значение параметра, при котором будет выполняться следующее:

$$P_i = T_i + \mu \vec{\tau}.$$

Итерационную схему параллельного сближения предлагается модифицировать следующим образом. Пусть в момент начала сближения вектор скорости преследователя  $P_{i-1}$  направлен произвольным образом, но не в точку на окружности Аполлония (рис. 4). В силу инертности преследователя минимальный радиус кривизны траектории не может быть меньше определенного значения  $r_c$ . Точке преследователя  $P_{i-1}$  соответствуют вектор скорости  $\vec{V}_{P_{i-1}}$  и вектор единичной нормали  $\vec{n}_{i-1}$ ,  $\vec{V}_{P_{i-1}} \vec{n}_{i-1} = 0$ . Далее находится центр окружности радиуса  $r_c$ :

$$C_{i-1} = P_{i-1} + V_P \Delta T \vec{n}_{i-1}.$$

К построенной окружности строится касательная из точки  $T_{i-1}$  для нахождения точки  $P_{t_{i-1}}$  сопряжения прямой и окружности (см. рис. 4). Дугу окружности  $P_{i-1} P_{t_{i-1}}$  и отрезок  $[P_{t_{i-1}} T_{i-1}]$  будем считать одной составной кривой линией  $f_{i-1}(s)$ , где параметром  $s$  служит длина дуги данной параметрической кривой.

В предлагаемой тестовой программе отсчет длины дуги начинается от точки  $T_{i-1}$ :  $f_{i-1}(0) = T_{i-1}$ . Проведем параллельный перенос линии  $f_{i-1}(s)$  на вектор  $T_i - T_{i-1}$ . Положение точки  $T_i$  известно и полностью определяется поведением цели. В рамках решения данной задачи будем считать поведе-

ние цели полностью детерминированным. Уравнение параллельной линии  $f_i(s) = f_{i-1}(s) + T_i - T_{i-1}$  будем считать известным, и для нахождения точки  $P_i$  следующего шага преследователя необходимо решение следующей системы уравнений и неравенств относительно параметра  $s$  ( $0 \leq s \leq s_{i-1}$ ):

$$\begin{aligned} (f_i(s) - P_{i-1})^2 &= (V_P \Delta T)^2; \\ f_i(s) &= f_{i-1}(s) + T_i - T_{i-1}, \end{aligned}$$

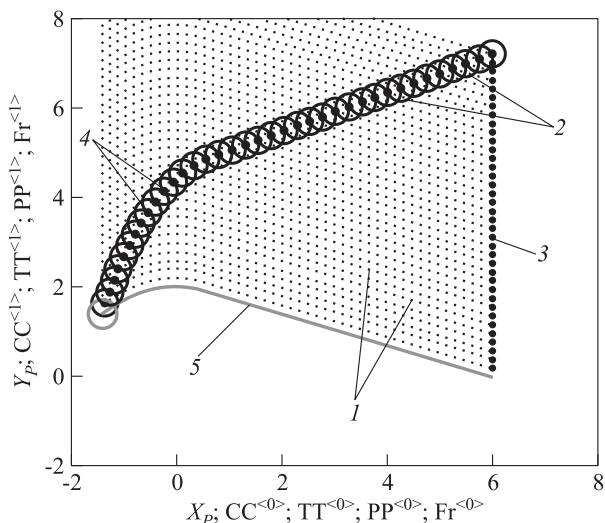
где  $s_{i-1}$  — это значение параметра  $s$ , соответствующее точке  $P_{i-1}$ . Первое, что сделано в предлагаемой программе, — для составной кривой в момент начала преследования, состоящей из дуги и отрезка, выполнена параметризация от длины дуги. Для этого было необходимо получить упорядоченный набор точек  $\{x_i, y_i\}$ . По каждой координате встроенными средствами MathCAD выполнена кубическая сплайн-интерполяция от формального параметра  $\delta$  и получены функции  $X(\delta)$ ,  $Y(\delta)$ ,  $i \in [0, N - 1]$ ,  $\delta_i \leq \delta \leq \delta_{i+1}$ , где  $\delta_i = i$ ;  $N$  — число элементов массивов  $\{x_i, y_i\}$ .

Далее был составлен якобиан для передачи во встроенные решатели системы MathCAD:

$$D(s, \delta) = \frac{1}{\sqrt{dX^2/d\delta + dY^2/d\delta}}.$$

Полученное решение выражает зависимость массивов  $\{x_i, y_i\}$  от параметра длины дуги  $s$ . Таким образом, можно считать, что уравнение базовой кривой, с которой будет совершаться параллельный перенос, получено (рис. 5). Далее предстоит создать вычислительный цикл, в котором решается система уравнений и неравенств. Данная задача сводится к численному решению уравнения поиска нулей функции методом секущей в заданном диапазоне.

Встроенные средства численного решения уравнений системы MathCAD позволяют решить уравнение



**Рис. 5** Кинематическая модель параллельного сближения: 1 — линии параллельного переноса  $Y_p(X_p)$ ; 2 — окружности числа  $CC^{(1)}(CC^{(0)})$ ; 3 — траектория цепи  $TT^{(1)}(TT^{(0)})$ ; 4 — траектория преследователя  $PP^{(1)}(PP^{(0)})$ ; 5 — базовая линия предполагаемых траекторий  $Fr^{(1)}(Fr^{(0)})$

$$(f_{i-1}(s) + T_i - T_{i-1} - P_{i-1})^2 - (V_P \Delta T)^2 = 0$$

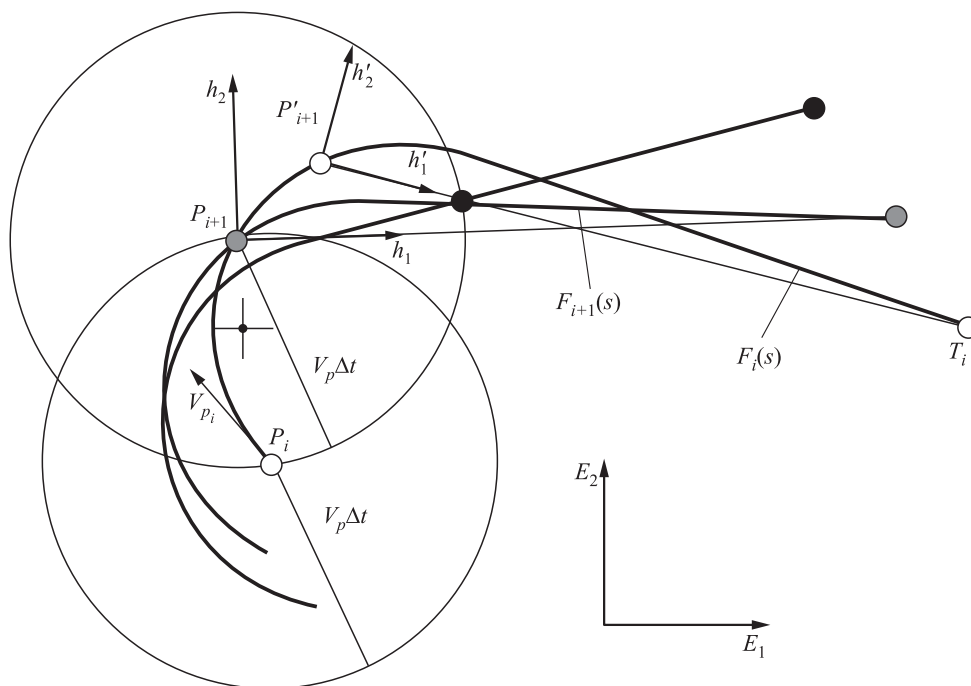
в диапазоне  $s \in [0, s_{i-1}]$ . На анимированном изображении рис. 5 представлены результаты моделирования тестовой программы [3].

### 3 Кинематическая модель метода погони

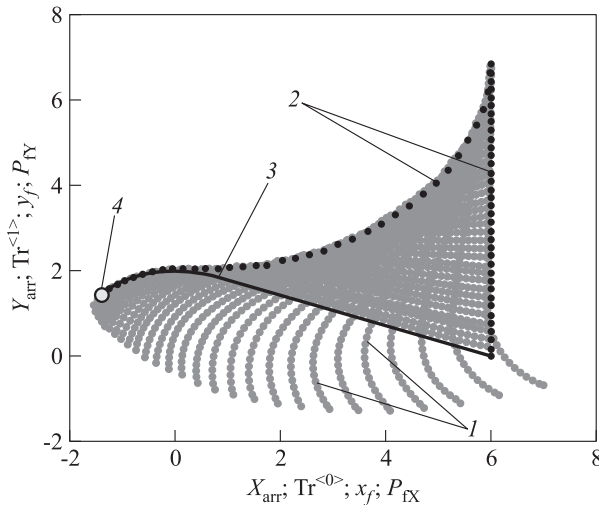
Рассмотрим следующую итерационную схему. Будем считать, что в момент времени  $t_i$  известно положение цели  $T_i$ , положение преследователя  $P_i$  и векторное уравнение  $F_i(s)$  прогнозируемой на данный момент времени траектории движения преследователя (рис. 6). В таком итерационном процессе ставится задача рассчитать координаты  $P_{i+1}$  следующего шага преследователя и выполнить аффинные преобразования векторной функции  $F_i(s)$ , чтобы найти выражения для функции  $F_{i+1}(s)$ . Чтобы найти координаты точки  $P_{i+1}$ , необходимо решить уравнение

$$|F_i(s_{i+1}) - F_i(s_i)| = V_P \Delta t$$

относительно параметра  $s_{i+1}$ . Когда разрабатывалась модель итерационного процесса, задавались начальные координаты  $T$  и  $P$  и начальный вектор скорости движения преследователя  $V_P$ . Также определялась векторная функция  $F(s)$  в момент начала преследования. На рис. 1 это составная кривая из дуги  $\widetilde{PP}_t$  и прямолинейного сегмента  $[P_t T]$ , где параметром служит длина дуги этой кривой. На  $i$ -м итерационном шаге происходит следующее. Строится окружность с центром в точке  $P_i$  радиуса  $V_P \Delta t$ . Точка пересечения этой окружности



**Рис. 6** Моделирование траектории преследователя



**Рис. 7** Кинематическая модель метода погони: 1 — точки предполагаемых траекторий  $Y_{arr}(X_{arr})$ ; 2 — точки траекторий преследователя и цели  $(T_2)^{(1)}((T_2)^{(0)})$ ; 3 — динамическая предполагаемая траектория  $y_f(x_f)$ ; 4 — динамическая точка преследователя  $P_{fY}(P_{fX})$

и траектории  $F_i(s)$  будет точкой следующего шага преследователя  $P_{i+1}$ . Потом находится точка пересечения  $P'_{i+1}$  параметрической функции  $F_i(s)$  с окружностью с центром в точке  $T_i$  и радиуса  $|T_{i+1} - P_{i+1}|$ . Затем формируется локальный базис  $(h'_1, h'_2)$  с центром координат в точке  $P'_{i+1}$ . Пересчитывается функция  $F_i(s)$ . В базисе  $(h'_1, h'_2)$  она будет выглядеть как  $F'_i(s)$ . Компоненты базиса  $(h'_1, h'_2)$  таковы:

$$h'_1 = \frac{T_i - P'_{i+1}}{|T_i - P'_{i+1}|}, \quad h'_2 = \begin{bmatrix} -h'_{1y} \\ h'_{1x} \end{bmatrix}.$$

Функция  $F'_i(s)$  будет иметь вид:

$$F'_i(s) = \begin{bmatrix} (F_i(s) - P'_{i+1}) h'_1 \\ (F_i(s) - P'_{i+1}) h'_2 \end{bmatrix}.$$

Сформируем базис  $(h_1, h_2)$  с центром координат в точке  $P_{i+1}$ . Компоненты базиса  $(h_1, h_2)$  будут выглядеть как

$$h_1 = \frac{T_{i+1} - P_{i+1}}{|T_{i+1} - P_{i+1}|}; \quad h_2 = \begin{bmatrix} -h_{1y} \\ h_{1x} \end{bmatrix}.$$

Отметим очень важный момент, что  $|T_{i+1} - P_{i+1}| = |T_i - P'_{i+1}|$ . Отсюда можно утверждать, что локальное представление в локальном базисе  $(h_1, h_2)$  с центром в точке  $P_{i+1}$  кривой  $F_{i+1}(s)$  будет совпадать с локальным представлением базиса  $(h'_1, h'_2)$ :  $F'_{i+1}(s) = F'_i(s)$ . Базис  $(E_1, E_2)$  в базисе  $(h_1, h_2)$  выглядит как

$$e_1 = \begin{bmatrix} E_1 h_1 \\ E_1 h_2 \end{bmatrix}; \quad e_2 = \begin{bmatrix} E_2 h_1 \\ E_2 h_2 \end{bmatrix}.$$

Уравнение линии  $F_{i+1}(s)$  будет иметь вид:

$$F_{i+1}(s) = \begin{bmatrix} F'_{i+1}(s) e_1 \\ F'_{i+1}(s) e_2 \end{bmatrix} + P_{i+1}.$$

Итак, на  $i$ -м шаге итерации имеем следующее: шаг  $T_{i+1}$  выбирается целью, шаг преследователя  $P_{i+1}$  рассчитывается как точка пересечения окружности  $(P_i, V_P \Delta t)$  и ранее рассчитанной линии  $F_i(s)$ , а на основе уже имеющихся данных рассчитывается новая прогнозируемая траектория преследователя  $F_i(s)$ . Анимированное изображение рис. 7 как раз демонстрирует результаты работы программы [4].

## 4 Заключение

В настоящей статье рассматриваются кинематические модели задачи преследования на плоскости методами параллельного сближения и погони. В момент преследования скорость преследователя направлена произвольно. Данные методики возможны для использования при разработке геометрической модели группового преследования с одновременным достижением цели или целей. По предложенным моделям и алгоритмам написаны тестовые программы расчета траекторий в системе компьютерной математики MathCAD. Тексты программ доступны на ресурсе [5]. При написании статьи за основу приняты теоретические результаты, полученные в [1, 2, 6]. Также приняты во внимание результаты работ [7–13].

## Литература

1. Петросян Л. А., Рихсиев Б. Б. Преследование на плоскости. — М.: Наука, 1991. 96 с.
2. Петросян Л. А. Дифференциальные игры преследования // Соросовский образовательный ж., 1995. № 1. С. 88–91.
3. Метод параллельного сближения на плоскости с ограничениями на кривизну. <https://www.youtube.com/watch?v=qNXdykK21Z8>.
4. Метод погони на плоскости с ограничениями на кривизну. <https://www.youtube.com/watch?v=UQ5bVKjVqZ4>.
5. Программный код в системе MathCAD. <http://dubanov.exponenta.ru/books.htm>.
6. Айзекс Р. Дифференциальные игры / Пер. с англ. — М.: Мир, 1967. 480 с. (Isaacs R. Differential games: A mathematical theory with applications to warfare and pursuit, control and optimization. — Dover books on mathematics ser. — Wiley, 1965. 384 p.)

7. Ibragimov G., Hussin N. A. A Pursuit-evasion differential game with many pursuers and one evader // *Malaysian J. Mathematical Sciences*, 2010. Vol. 4. Iss. 2. P. 183–194.
8. Кузьмина Л. И., Осипов Ю. В. Расчет длины траектории для задачи преследования // *Вестник МГСУ*, 2013. № 12. С. 20–26.
9. Саматов Б. Т. Задача преследования убегания при интегрально-геометрических ограничениях на управления преследователя // *Автоматика и телемеханика*, 2013. Вып. 7. С. 17–28.
10. Ibragimov G., Norshakila A. R., Kuchkarov A., Ismail F. Multi pursuer differential game of optimal approach with integral constraints on controls of players // *Taiwan. J. Math.*, 2015. Vol. 19. Iss. 3. P. 963–976.
11. Petrov N. N., Solov'eva N. A. Group pursuit with phase constraints in recurrent Pontryagin's example // *Int. J. Pure Applied Mathematics*, 2015. Vol. 100. Iss. 2. P. 263–278.
12. Романников Д. О. Пример решения минимаксной задачи преследования с использованием нейронных сетей // *Сборник научных трудов НГТУ*, 2018. № 2(92). С. 108–116. doi: 10.17212/2307-6879-2018-2-108-116.
13. Ахметжанов А. Р. Динамические игры преследования на поверхностях: Автореф. дис. . . . канд. физ.-мат. наук. — М.: МФТИ, 2019. 28 с.

*Поступила в редакцию 18.08.20*

## KINEMATIC MODELS OF PURSUIT PROBLEMS ON THE PLANE BY THE METHODS OF PARALLEL APPROACH AND PURSUIT

A. A. Dubanov and V. A. Nefedova

Banzarov Buryat State University, 6a Ranzhurov Str., Ulan-Ude 670000, Russian Federation

**Abstract:** This article provides accurate pursuit models based on the parallel approach and chase methods. This article is a modification of the methods of parallel rapprochement and chasing what happens when the pursuit begins. It cannot instantly change the direction of movement. This should be a less affordable option. The proposed method is based on the fact that the pursuer chooses a step at the iteration stage and will try to follow the predicted trajectories. Based on the materials of the article, test programs have been written that calculate the trajectories of the pursuer taking into account the stated conditions. Execution of animated images visualizes the change in the coordinates of the pursuer, target, and predicted time trajectories.

**Keywords:** target; pursuer; trajectory; convergence; modeling

**DOI:** 10.14357/19922264220314

### References

1. Petrosyan, L. A., and B. B. Rihsiev. 1961. *Presledovanie na ploskosti* [Flat pursuit]. Moscow: Nauka. 96 p.
2. Petrosyan, L. A. 1995. *Differentsial'nye igry presledovaniya* [Differential pursuit games]. *Sorosovskiy obrazovatel'nyy zh.* [Soros Educational J.] 1:88–91.
3. Metod paralel'nogo sblizheniya na ploskosti s ogranicheniyami na kriviznu [Method of parallel approach on a plane with restrictions on curvature]. Available at: <https://www.youtube.com/watch?v=qNXdykK21Z8> (accessed May 25, 2022).
4. Metod pogoni na ploskosti s ogranicheniyami na kriviznu [Plane chase method with curvature constraints]. Available at: <https://www.youtube.com/watch?v=UQ5bVKjVqZ4> (accessed May 25, 2022).
5. Programmnyy kod v sisteme MathCAD [Program code in the MathCAD system]. Available at: <http://dubanov.exponenta.ru/books.htm> (accessed May 25, 2022).
6. Isaacs, R. 1965. *Differential games: A mathematical theory with applications to warfare and pursuit, control and optimization*. Dover books on mathematics ser. Wiley. 384 p.
7. Ibragimov, G. A., and N. A. Hussin. 2010. Pursuit-evasion differential game with many pursuers and one evader. *Malaysian J. Mathematical Sciences* 4(2):183–194.
8. Kuzmina, L. I., and Y. Y. Osipov. 2013. Raschet dliny traektorii dlya zadachi presledovaniya [Calculation of the path length in the pursuit problem]. *Vestnik MGSU* [Bulletin of MGSU] 12:20–26.
9. Samatov, B. T. 2013. The pursuit-evasion problem under integral-geometric constraints on pursuer controls. *Automat. Rem. Contr.* 74:1072–1081.
10. Ibragimov, G., A. R. Norshakila, A. Kuchkarov, and F. Ismail. 2015. Multi pursuer differential game of optimal approach with integral constraints on controls of players. *Taiwan. J. Math.* 19 (3):963–976.
11. Petrov, N. N., and N. A. Solov'eva. 2015. Group pursuit with phase constraints in recurrent Pontryagin's example. *Int. J. Pure Applied Mathematics* 100(2):263–278.

12. Romannikov, D. O. 2018. Primer resheniya minimaksnoy zadachi presledovaniya s ispol'zovaniem neyronnykh setey [An example of solving a minimax pursuit problem using neural networks]. *Sbornik nauchnykh trudov NGTU* [Transactions of scientific papers of the Novosibirsk State Technical University] 2(92):108–116. doi: 10.17212/2307-6879-2018-2-108-116.
13. Ahmetzhanov, A. R. 2019. Dinamicheskie igry presledovaniya na poverkhnostyakh [Dynamic pursuit games on surfaces]. Moscow: MIPT. PhD Thesis. 28 p.

*Received August 18, 2020*

## Contributors

**Dubanov Alexander A.** (b. 1967) — Candidate of Science (PhD) in technology, associate professor, Department of Geometry and Methods of Teaching Mathematics, Institute of Mathematics and Informatics, Banzarov Buryat State University, 6a Ranzhurov Str., Ulan-Ude 670000, Russian Federation; alandubanov@mail.ru

**Nefedova Victoria A.** (b. 2000) — student, Institute of Mathematics and Informatics, Banzarov Buryat State University, 6a Ranzhurov Str., Ulan-Ude 670000, Russian Federation; emelyanovatorik@gmail.com



**Алию Башир** (р. 1975) — аспирант кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

**Босов Алексей Вячеславович** (р. 1969) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Бурцева София Артемовна** (р. 1999) — студент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

**Васильев Николай Семенович** (р. 1952) — доктор физико-математических наук, профессор Московского государственного технического университета имени Н. Э. Баумана

**Власкина Анастасия Сергеевна** (р. 1995) — ассистент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

**Горшенин Андрей Константинович** (р. 1986) — доктор физико-математических наук, доцент, руководитель отдела, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Грушо Александр Александрович** (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Грушо Николай Александрович** (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Гусейнова Екатерина Игоревна** (р. 1999) — студент магистратуры экономического факультета Московского государственного университета имени М. В. Ломоносова

**Дубанов Александр Анатольевич** (р. 1967) — кандидат технических наук, доцент кафедры геометрии и методики преподавания математики Института

математики и информатики Бурятского государственного университета

**Зацман Игорь Моисеевич** (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Зейфман Александр Израилевич** (р. 1954) — доктор физико-математических наук, профессор, заведующий кафедрой прикладной математики Вологодского государственного университета; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; главный научный сотрудник Вологодского научно-го центра Российской академии наук; старший научный сотрудник Московского центра фундаментальной и прикладной математики Московского государственного университета имени М. В. Ломоносова

**Ковалёв Иван Александрович** (р. 1996) — аспирант Вологодского государственного университета; научный сотрудник Московского центра фундаментальной и прикладной математики Московского государственного университета имени М. В. Ломоносова

**Кочеткова Ирина Андреевна** (р. 1985) — кандидат физико-математических наук, доцент Российского университета дружбы народов; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Кривенко Михаил Петрович** (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Малашенко Юрий Евгеньевич** (р. 1946) — доктор физико-математических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Мачнев Егор Александрович** (р. 1996) — аспирант кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

**Мигуля Максим Алексеевич** (р. 2001) — бакалавр Национального исследовательского университета «Высшая школа экономики»

**Мокров Евгений Владимирович** (р. 1988) — кандидат физико-математических наук, ассистент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

**Нефедова Виктория Андреевна** (р. 2000) — студент Института математики и информатики Бурятского государственного университета

**Нуриев Виталий Александрович** (р. 1980) — доктор филологических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Сатин Яков Александрович** (р. 1978) — кандидат физико-математических наук, доцент кафедры прикладной математики Вологодского государственного университета

**Синицина Анна Владимировна** (р. 1988) — кандидат физико-математических наук, научный сотрудник Ярославского государственного университета им. П. Г. Демидова

**Тимонина Елена Евгеньевна** (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Шведов Алексей Сергеевич** (р. 1956) — доктор физико-математических наук, профессор, профессор факультета экономических наук Национального исследовательского университета «Высшая школа экономики»

**Шнурков Петр Викторович** (р. 1953) — кандидат физико-математических наук, доцент Национального исследовательского университета «Высшая школа экономики»

**Шоргин Сергей Яковлевич** (р. 1952) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

---

## Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

---

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.
5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или  $\LaTeX$  (.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 — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 10 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на *русском и английском языках*:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:  
[http://www.ipiran.ru/journal/issues/2013\\_07\\_01/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp) и  
[http://www.ipiran.ru/journal/issues/2013\\_07\\_01\\_eng/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp);
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/bgn/> бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

**Описание статьи из журнала:**

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaistvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

**Описание статьи из электронного журнала:**

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

**Описание статьи из продолжающегося издания (сборника трудов):**

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

**Описание материалов конференций:**

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

**Описание книги (монографии, сборники):**

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

**Описание переводной книги** (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — New York, NY, USA: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. New York: Wiley. 521 p.

**Описание неопубликованного документа:**

Laturov, A. R., M. M. Khasanov, and V. A. Baikov. 2004 (unpubl.). *Geologiya i dobycha (NGT GiD)* [Geology and production (NGT GiD)]. Certificate on official registration of the computer program No. 2004611198.

**Описание интернет-ресурса:**

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

**Описание диссертации или автореферата диссертации:**

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. Moscow. D.Sc. Diss. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 23 p.

**Описание ГОСТа:**

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

**Описание патента:**

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.

11. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- использовать attach (присоединение);
- в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.

12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

**Адрес редакции журнала «Информатика и её применения»:**

Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05

e-mail: [ieep@frccsc.ru](mailto:ieep@frccsc.ru) (Стригина Светлана Николаевна)

<http://www.ipiran.ru/journal/issues/>



---

## Requirements for manuscripts submitted to Journal “Informatics and Applications”

---

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
- mathematical methods for studying complex systems and processes;
- information systems and networks;
- information technologies; and
- architecture and software of computational complexes and networks.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

***Agreement on the transfer of rights to publish:***

*“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.*

*Author(s) signature(s): (name(s), address(es), date).*

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or L<sup>A</sup>T<sub>E</sub>X (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 10 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site: <http://www.ipiran.ru/journal/template.doc>.

8. The articles should enclose data both in **Russian and English**:

- title;
- author’s name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format: (see site)

[http://www.ipiran.ru/journal/issues/2013\\_07\\_01/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp) and

[http://www.ipiran.ru/journal/issues/2013\\_07\\_01\\_eng/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp);

- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;
  - indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.  
Important! Keywords must not be sentences;
  - Acknowledgments.
9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.net/ru/bgn/>).
- Please take into account the following examples of Russian references appearance:
- Article in journal:**  
Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.
- Journal article in electronic format:**  
Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).
- Article from the continuing publication (collection of works, proceedings):**  
Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy "stal'-kompozit" [Experimental study of the strength of joints "steel-composite"]. *Trudy MGTU "Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem"* [Bauman MSTU "Mathematical Modeling of Complex Technical Systems" Proceedings]. 593:125–130.
- Conference proceedings:**  
Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursoberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazoidachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.
- Books and other monographs:**  
Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.
- Dissertation and Thesis:**  
Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo 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;
  - an article and additional materials have to be attached using the "attach" function;
  - an electronic version of the article should contain the file with the text and a separate file with figures.
12. "Informatics and Applications" journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

**Editorial Board address:**

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia  
Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05  
e-mail: [iiep@frccsc.ru](mailto:iiep@frccsc.ru) (to Svetlana Strigina)  
<http://www.ipiran.ru/english/journal.asp>