

# Информатика и её применения

Том 14 Выпуск 3 Год 2020

## СОДЕРЖАНИЕ

Статистическое оценивание распределений случайных коэффициентов стохастического дифференциального уравнения Ланжевена <b>А. К. Горшенин, В. Ю. Королев, А. А. Щербинина</b>	<b>3</b>
О марковских и рациональных потоках случайных событий. I <b>В. А. Наумов, К. Е. Самуйлов</b>	<b>13</b>
Аппроксимация множества решений систем нелинейных неравенств с использованием графических ускорителей <b>М. В. Попов, М. А. Посыпкин</b>	<b>20</b>
Однолинейная система массового обслуживания с инверсионным порядком обслуживания с вероятностным приоритетом, групповым пуассоновским потоком и фоновыми заявками <b>Т. А. Милованова, Р. В. Разумчик</b>	<b>26</b>
О распределении отношения суммы элементов выборки, превосходящих некоторый порог, к сумме всех элементов выборки. I <b>В. Ю. Королев</b>	<b>35</b>
О статистических свойствах оценки риска в задаче обращения преобразования Радона при случайном объеме проекционных данных <b>О. В. Шестаков</b>	<b>44</b>
Метод логарифмических моментов для оценивания параметров гамма-экспоненциального распределения <b>А. А. Кудрявцев, О. В. Шестаков</b>	<b>49</b>
Основные понятия программирования в изложении для дошкольников <b>В. Б. Бетелин, А. Г. Кушниренко, А. Г. Леонов</b>	<b>55</b>
Вычислительные аспекты применения CC-VaR на совокупности рынков <b>Г. А. Агасандян</b>	<b>62</b>
Методы математической статистики в задаче поиска инсайдера <b>А. А. Грушо, М. И. Забежайло, Д. В. Смирнов, Е. Е. Тимонина, С. Я. Шоргин</b>	<b>71</b>
Выявление аномалий с помощью метаданных <b>А. А. Грушо, Е. Е. Тимонина, Н. А. Грушо, И. Ю. Терехина</b>	<b>76</b>
Аппроксимация множества достижимых потоков многопользовательской сети <b>Ю. Е. Малашенко, И. А. Назарова</b>	<b>81</b>
Байесовский подход к построению индивидуальной траектории пользователя в системе дистанционного обучения <b>А. В. Босов, Я. Г. Мартюшова, А. В. Наумов, А. П. Сапунова</b>	<b>86</b>

# Информатика и её применения

Том 14 Выпуск 3 Год 2020

## СОДЕРЖАНИЕ

Анализ механизмов нарезки сети с учетом гарантий для различных типов трафика	
<b><i>К. А. Агеев, Э. С. Сопин, Н. В. Яркина, К. Е. Самуйлов, С. Я. Шоргин</i></b>	<b>94</b>
О концепции стохастической модели с управлением в моменты выхода процесса на границу заданного подмножества множества состояний	
<b><i>П. В. Шнурков, Д. А. Новиков</i></b>	<b>101</b>
Извлечение оптимизационных моделей из данных	
<b><i>В. И. Донской</i></b>	<b>109</b>
Проблемно-ориентированная верификация полноты темпоральных онтологий и заполнение понятийных лакун	
<b><i>И. М. Зацман</i></b>	<b>119</b>
Использование тематических моделей для парного сравнения коллекций научных статей	
<b><i>Ф. В. Краснов, А. В. Диментов, М. Е. Шварцман</i></b>	<b>129</b>
Об авторах	<b>136</b>
Правила подготовки рукописей	<b>138</b>
Requirements for manuscripts	<b>141</b>

# СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ КОЭФФИЦИЕНТОВ СТОХАСТИЧЕСКОГО ДИФФЕРЕНЦИАЛЬНОГО УРАВНЕНИЯ ЛАНЖЕВЕНА\*

А. К. Горшенин<sup>1</sup>, В. Ю. Королев<sup>2</sup>, А. А. Щербинина<sup>3</sup>

**Аннотация:** Разработан метод статистического оценивания распределений случайных коэффициентов стохастических дифференциальных уравнений (СДУ) типа Ланжевена с помощью техники скользящего разделения смесей (СРС). Предложены дискретные аппроксимации для оценок указанных распределений. С целью изучения изменчивости распределений коэффициентов сдвига (дрейфа) и диффузии СДУ во времени предложен алгоритм последовательной идентификации (определения локальной связности) компонент получаемых смесей. В его основу положена комбинация жадного алгоритма для поиска числа компонент и одного из методов кластеризации ( $k$ - или  $c$ -средних). Применение метода иллюстрируется конкретными примерами анализа процесса теплообмена между атмосферой и океаном для Гольфстрима и тропиков.

**Ключевые слова:** стохастические дифференциальные уравнения; смешанные распределения; локальная связность; жадный алгоритм; кластеризация

**DOI:** 10.14357/19922264200301

## 1 Введение

В физике СДУ Ланжевена принято называть следующее соотношение:

$$dX(t) = a(t) dt + b(t) dW, \quad (1)$$

определяющее случайный процесс  $X(t)$ , где  $W(t)$  — винеровский процесс, а коэффициенты  $a(t)$  и  $b(t)$  случайны. Стохастические дифференциальные уравнения вида (1) широко используются, например, в задаче ассимиляции данных при анализе разномасштабной изменчивости геофизических переменных [1]. В финансовой математике известны специальные версии уравнения (1). В частности, весьма популярна модель геометрического броуновского движения вида:

$$dX(t) = aX(t) dt + bX(t) dW, \quad (2)$$

где  $a \in \mathbb{R}$ ,  $b > 0$ . Известно много обобщений модели (2) с конкретными видами зависимости  $a$  и  $b$  от  $X(t)$  и других случайных процессов, например модели Леланда [2], Барлса—Сонера [3],

Хестона [4], Кокса—Ингерсолла—Росса [5], Халла—Уайта [6] и другие так называемые модели стохастической волатильности (см. также [7–9]).

При отсутствии априорной информации о «физической» структуре процесса  $X(t)$  для успешного изучения и прогнозирования его эволюции перво-степенную важность приобретает задача статистического оценивания функциональных коэффициентов  $a(t)$  и  $b(t)$ . В силу их случайности данная задача допускает как минимум две принципиально разные формулировки. Во-первых, можно пытаться найти (случайные же) оценки самих функций  $a(t)$  и  $b(t)$ , т. е. найти их точечные аппроксимации, и, во-вторых, пытаться статистически оценить распределения случайных величин  $a(t)$  и  $b(t)$ . Во втором случае, зная какие-либо свойства этих коэффициентов, например структуру их функциональной зависимости от исходного процесса  $X(t)$  (как в упомянутых выше моделях Леланда, Барлса—Сонера, Хестона, Кокса—Ингерсолла—Росса или Беляева), можно найти оценки числовых параметров этих моделей.

В данной статье рассматривается вторая задача. Пусть  $n \geq 1$  и  $t_0 = 0 < t_1 < \dots < t_n$  — моменты

\* Исследования выполнены при частичной поддержке РФФИ (проект 19-07-00914), Стипендии Президента Российской Федерации молодым ученым и аспирантам (СП-538.2018.5) и в соответствии с программой Московского центра фундаментальной и прикладной математики.

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

<sup>2</sup>Факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vkorolev@cs.msu.ru

<sup>3</sup>Факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова, shcherbinina.aa.97@gmail.com

времени, в которые наблюдается процесс  $X(t)$ . Для простоты предположим, что  $t_i - t_{i-1} = 1$  для любого  $i \geq 1$ . Из уравнения (1) видно, что распределение приращения  $X(t_i) - X(t_{i-1})$  процесса  $X(t)$  можно аппроксимировать распределением вида:

$$\mathbb{P}(X(t_i) - X(t_{i-1}) < x) \approx \mathbb{E}\Phi\left(\frac{x - A_i}{B_i}\right), \quad (3)$$

где  $\Phi(x)$  — стандартная нормальная функция распределения;  $A_i \in \mathbb{R}$  и  $B_i > 0$  — случайные величины. В свою очередь, для распределений случайных величин  $A_i$  и  $B_i$ , по отношению к которым берется математическое ожидание в формуле (3), можно использовать дискретную аппроксимацию. Тогда вместо выражения (3) для распределения приращения  $X(t_i) - X(t_{i-1})$  можно применить приближение вида:

$$\mathbb{P}(X(t_i) - X(t_{i-1}) < x) \approx \sum_{k=1}^K p_k \Phi\left(\frac{x - a_k}{b_k}\right), \quad (4)$$

где  $K \in \mathbb{N}$ ,  $p_k \geq 0$ ,  $k = \overline{1, K}$ ,  $\sum_k p_k = 1$ . Очевидно, параметры  $p_k$ ,  $a_k$  и  $b_k$ , формирующие так называемые динамические и диффузионные компоненты [10], зависят также от  $i$  и изменяются при переходе от  $t_i$  к  $t_{i+1}$ .

Для статистического оценивания параметров  $p_k$ ,  $a_k$  и  $b_k$  можно использовать СРС-метод, описанный в [10]. Статистические закономерности поведения рассматриваемых процессов  $X(t)$ ,  $a(t)$  и  $b(t)$  изменяются во времени, вообще говоря, нерегулярным образом, результатом чего является отсутствие универсального смешивающего закона. Поэтому изучение динамики изменения статистических закономерностей в поведении исследуемого процесса проводится на последовательных интервалах времени. Тем самым параметры смесей (сдвига (дрейфа)  $a_k$ , масштаба (диффузии)  $b_k$  и веса компонент  $p_k$ ) оцениваются как функции времени. С этой целью при каждом положении скользящего окна используется ЕМ (expectation-maximization) алгоритм или какие-либо из его модификаций [10–12].

Для всеобъемлющего изучения закономерностей эволюции процесса  $X(t)$ , задаваемого СДУ (1), необходимо знать, как ведут себя случайные коэффициенты  $a(t)$  и  $b(t)$  во времени. Для этого, в свою очередь, необходимо восстановить эволюцию во времени их распределений. В рамках аппроксимации (4) последняя задача сводится к последовательной идентификации компонент смеси (4), т. е. к определению соответствия оценок параметров  $p_k$ ,  $a_k$  и  $b_k$ , полученных на разных (например, соседних) положениях скользящего окна. Другими словами, чтобы определить эволюцию распределения случайных функций  $a(t)$  и  $b(t)$ , надо восстановить

компоненты смеси (4) как функции времени. Таким образом, решается задача приближенного восстановления двумерного распределения

$$F_t(x, y) \equiv \mathbb{P}(a(t) < x, b(t) < y).$$

С этой целью при каждом  $i = \overline{1, n}$  ищется дискретная аппроксимация распределения  $F_{t_i}(x, y)$ , а затем осуществляется интерполяция этой функции для остальных  $t$ . Для интерполяции необходимо установить соответствие между множествами возможных значений  $\{a_k, b_k : k = \overline{1, K}\}$  случайных коэффициентов  $a(t_i)$  и  $b(t_i)$  при разных  $i$ , т. е. восстановить эволюцию компонент аппроксимирующей смеси (4) во времени.

В процессе шагов СРС-метода выделяются компоненты смеси (4), которые эволюционируют при сдвиге скользящего окна. При этом достаточно сложно судить о том, какая из оценок параметров соответствует тому или иному значению на предыдущем шаге. Обычно подобная взаимосвязь определяется визуально и, очевидно, является достаточно субъективной. В статье предложен подход к автоматизации решения данной задачи на основе комбинации жадного алгоритма и методов кластеризации  $k$ - или  $c$ -средних [13–16]. При этом на первом этапе определяется число кластеров для методов машинного обучения, которые используются непосредственно для выявления связанных компонент. Данная процедура будет использована для получения статистических оценок коэффициентов уравнения (1) на примере анализа описываемых им процессов переноса тепла между океаном и атмосферой [1] в ряде географических точек.

## 2 Метод определения связности локальных компонент

Обозначим через  $I^{(n)}$  набор индексов (номеров) компонент для шага с номером  $n$  СРС-метода, т. е.  $I^{(n)} = \{1, 2, \dots, k^{(n)}\}$ , а через  $J^{(n+1)} = \{1, 2, \dots, k^{(n+1)}\}$  — аналогичный набор для шага  $(n+1)$ . Через  $I_0$  и  $J_0$  обозначим множество индексов из первого и второго наборов соответственно, для которых удалось найти ближайшую компоненту. Первоначально полагаем, что  $I_0 = \emptyset$  и  $J_0 = \emptyset$ . Для каждого фиксированного  $J \in J^{(n+1)} \setminus J_0$  найдем наиболее близкий номер  $I$  в смысле решения следующей оптимизационной задачи:

$$I = \arg \min_{i \in I^{(n)} \setminus I_0} \left( \left| a_i^{(n)} - a_J^{(n+1)} \right|^p + \left| \sigma_i^{(n)} - \sigma_J^{(n+1)} \right|^p \right)^{1/p}. \quad (5)$$

В этом случае минимизируемое в правой части выражение представляет собой  $\ell^p$ -норму ( $p = 1, 2, \dots$ ) соответствующего вектора разностей координат в пространстве точек  $(a, \sigma)$ .

Полагая после этого  $I_0 = I_0 \cup I$  и  $J_0 = J_0 \cup J$ , необходимо повторить процедуру заново. Возможны следующие случаи.

1.  $|I^{(n)}| = |J^{(n+1)}|$ , т.е.  $k^{(n)} = k^{(n+1)}$ . В этом случае оба набора будут исчерпаны одновременно.
2.  $|I^{(n)}| < |J^{(n+1)}|$ , т.е.  $k^{(n)} < k^{(n+1)}$ . В этом случае процедура останавливается, когда исчерпан набор  $I^{(n)} \setminus I_0$ . Оставшиеся в  $J^{(n+1)} \setminus J_0$  элементы формируют новые компоненты, которые появились только на  $(n + 1)$ -м шаге.

Случай  $|I^{(n)}| > |J^{(n+1)}|$ , т.е.  $k^{(n)} > k^{(n+1)}$ , не рассматривается, поскольку основная цель данной процедуры — определение числа компонент за весь период эволюции процесса, поэтому уменьшаться оно не может, даже если на каком-то шаге произошло сокращение локального значения для числа компонент. Отметим, что указанная процедура, очевидно, является жадной. При этом, поскольку ее конечная цель состоит в определении числа кластеров для следующего шага, данная особенность не представляется критической.

Для точного определения числа компонент необходимо задавать некоторый допустимый порог близости  $\varepsilon(\mathbf{a}, \sigma)$ :

$$\left( \left| a_I^{(n)} - a_J^{(n+1)} \right|^p + \left| \sigma_I^{(n)} - \sigma_J^{(n+1)} \right|^p + \left| p_I^{(n)} - p_J^{(n+1)} \right|^p \right)^{1/p} < \varepsilon(\mathbf{a}, \sigma). \quad (6)$$

Такая проверка нужна для того, чтобы корректно определять ситуацию, при которой компоненты с номерами  $I$  и  $J$  на  $n$ -м и  $(n + 1)$ -м шагах считались одинаковыми, и не было необходимости создавать новую в рамках жадного алгоритма. Реализация метода определения числа компонент приведена в алгоритме 1.

Предполагается, что данный алгоритм применяется для компонент с положительными весами (нулевые значения соответствуют случаю уменьшения их числа на каком-либо шаге). Кроме того, для всех допустимых значений  $i \neq j$  и  $n$  должны существовать такие  $\delta_a > 0$  и  $\delta_\sigma > 0$ , что выполнено хотя бы одно из условий:

$$\left| a_i^{(n)} - a_j^{(n)} \right| > \delta_a; \quad \left| \sigma_i^{(n)} - \sigma_j^{(n)} \right| > \delta_\sigma.$$

Если же они оба нарушены, то необходимо объединять эти компоненты в одну с соответствующей корректировкой (суммированием) весов, т.е. предполагается, что все компоненты различны, — это гарантирует корректность применения жадного алгоритма.

Алгоритм 1 используется в качестве важной составной части метода формирования матрицы связности. Она представляет собой вспомогательную структуру, в которой на каждом шаге скользящего окна сохраняется актуальное состояние всех выделенных к текущему моменту компонент. Сначала ко всему ряду применяется метод EM-типа для определения числа компонент на каждом шаге (как было отмечено выше, оно не может убывать). Затем в двухмерном пространстве  $(\mathbf{a}, \sigma)$  используется один из методов кластеризации с полученным жадным алгоритмом числом локальных компонент-кластеров. Веса не учитываются, так как вклад компоненты в смесь может изменяться, при этом параметры — математическое ожидание и дисперсия — варьируются не слишком сильно, и тогда считается, что это та же самая компонента. Соответствующая процедура представлена в алгоритме 2.

Очевидно, что данная процедура может быть использована и в случае смесей многомерных распределений. При этом в формулу оптимизации (5) должны быть добавлены все параметры соответствующего распределения (с соответствующей модификацией условия (6)), а затем может быть про-

**Алгоритм 1** Динамическое определение числа локальных компонент

```

function NUMGREEDY (Params,  $I^{(n)}$ ,  $J^{(n+1)}$ )
   $I_0 \leftarrow \emptyset$ ,  $J_0 \leftarrow \emptyset$ , Comps  $\leftarrow \emptyset$ ; // Инициализация
  repeat // Продолжающиеся или новые компоненты
    // Оптимизация выражения (5) с учетом условия (6)
    [ $I$ ,  $J$ ]  $\leftarrow$  FINDI(Params,  $J^{(n+1)} \setminus J_0$ ,  $I^{(n)} \setminus I_0$ );
    if  $I \neq \emptyset$  then // Найдена предшествующая  $J$  компонента
       $I_0 \leftarrow I_0 \cup I$ ,  $J_0 \leftarrow J_0 \cup J$ ;
    else // Добавление новой компоненты
       $J_0 \leftarrow J_0 \cup J$ ;
      Comps  $\leftarrow$  ADDNEWCOMP(Params,  $J$ );
  until ( $J^{(n+1)} \setminus J_0 \neq \emptyset$ )
  return Comps;
  
```

**Алгоритм 2** Определение компонент связности в СРС-методе

```

function MSMCOMPONENTS(Data, options)
  Params ← EMS(Data, options.EM) // СРС-метод
  // Инициализация числом ненулевых компонент на первом шаге
  Comps(1) ← Params.k(1);
  for (n = 1:LENGTH(Params)-1) do
    Comps(n+1) ← NUMGREEDY(Params, I(n), J(n+1));
  // Метки для каждого набора параметров, кластеризация
  Labels ← CLUST(Params, MAX(Comps), options.ClustAlg);
  // Матрица связности для компонент СРС-метода
  HistMatrConnect ← CONNECTIVITY(Params, Labels);
  PLOTCOMP(HistMatrConnect); // Визуализация результатов
  return HistMatrConnect;
  
```

ведена кластеризация в пространстве переменных новой размерности.

### 3 Оценивание коэффициентов уравнения Ланжевена на примере турбулентных потоков тепла между океаном и атмосферой

Для статистической оценки коэффициентов в уравнении Ланжевена воспользуемся алгоритмом 2, а также СРС-оценками, полученными при аппроксимации распределений приращений потоков тепла между океаном и атмосферой в нескольких географических точках. Отметим, что СРС-метод ранее использовался для статистического моделирования закономерностей в подобного рода данных [17, 18].

На рис. 1–4 верхние графики демонстрируют статистическую структуру процесса теплообмена между океаном и атмосферой в указанных географических точках для двух классов потоков [19] — скрытых и явных, — являющихся составляющими теплового баланса.

Благодаря структурам, возвращаемым функцией `MSMComponents`, можно точно отследить, когда те или иные из компонент существовали, прерывались и возобновлялись, и проанализировать их взаимосвязь с реальными физическими процессами. На нижних графиках, содержащих СРС-оценки для динамической и диффузионной компонент, продемонстрирована эволюция весов, т. е. вклад соответствующей структурной составляющей в общее развитие процесса во времени.

Для аппроксимации были использованы четырехкомпонентные нормальные смеси, однако при выбранных настройках жадного алгоритма 1 для всех рядов (за исключением явных потоков в тро-

пиках) были получены пять локальных компонент связности.

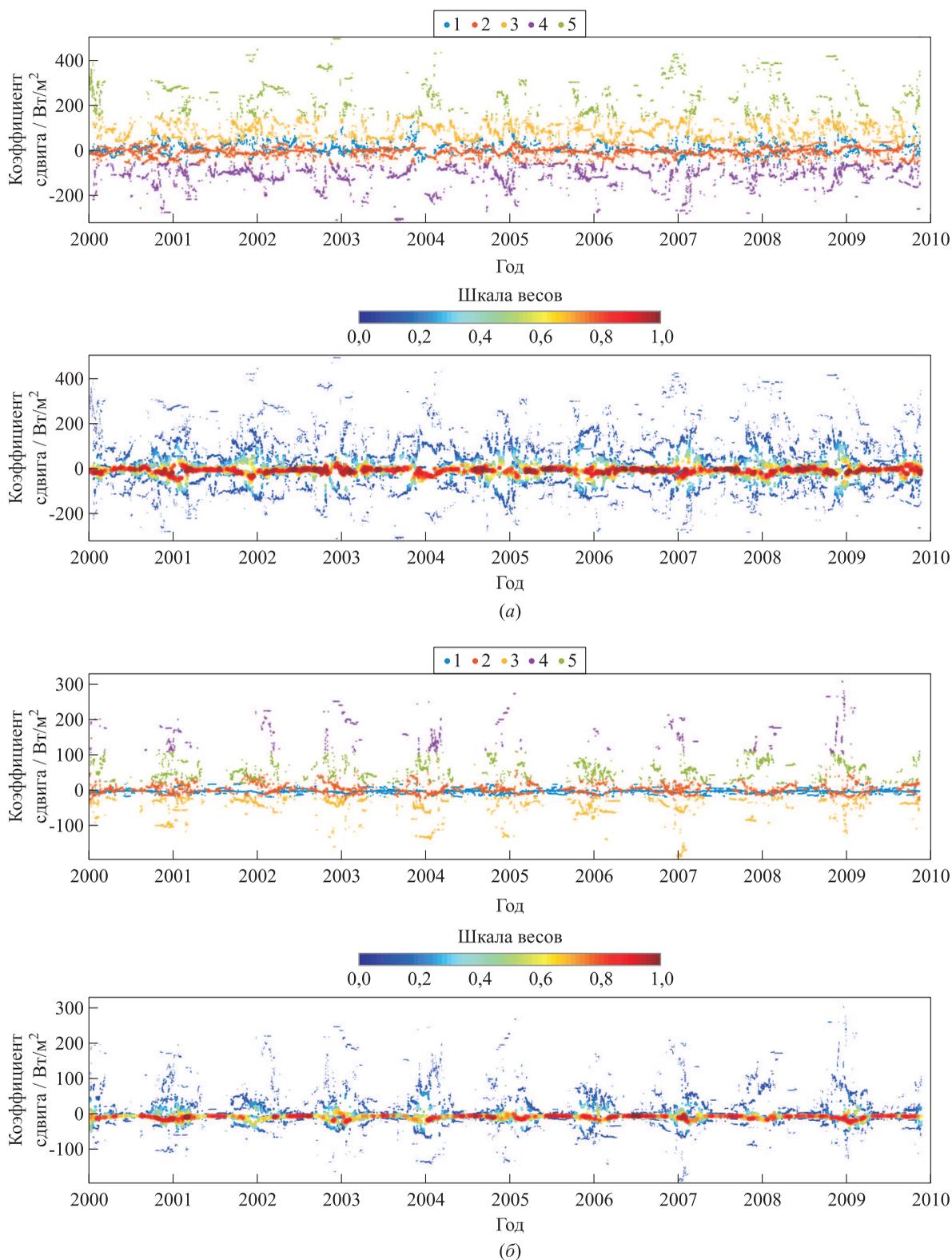
Видно, что общее число компонент не слишком сильно изменяется, поэтому результаты автоматического определения с помощью жадного алгоритма 1 варьируются от ряда к ряду не очень существенно. Однако для лучшего учета локальных процессов полученное число компонент (4–5) может быть расширено за счет повышения чувствительности процедуры путем выбора меньшего порогового значения в формуле (6).

### 4 Заключение

В работе описан метод статистического оценивания распределений случайных параметров СДУ типа Ланжевена с помощью СРС-техники. Предложены дискретные аппроксимации для оценок указанных распределений. С целью изучения изменчивости распределений коэффициентов СДУ во времени предложен алгоритм последовательной идентификации (определения локальной связности) компонент получаемых смесей. В его основу положена комбинация жадного алгоритма для поиска числа компонент и одного из методов кластеризации ( $k$ - или  $c$ -средних). Функциональные параметры (компоненты распределения (4) как функции времени), полученные в результате описываемых статистических процедур, могут быть использованы при обучении интеллектуальных алгоритмов прогнозирования процессов, удовлетворяющих уравнениям типа (1). Применение метода иллюстрируется конкретными примерами анализа процесса теплообмена между атмосферой и океаном.

### Литература

1. Belyaev K., Kuleshov A., Tuchkova N., Tanajura C. A. S. An optimal data assimilation method and its application to the numerical simulation of the ocean dynamics // Math. Comp. Model. Dyn., 2018. Vol. 1. Iss. 24. P. 12–25.



**Рис. 1** Оценки распределения сдвига (Гольфстрим): (а) явные потоки; (б) скрытые потоки; 1–5 — структурные компоненты

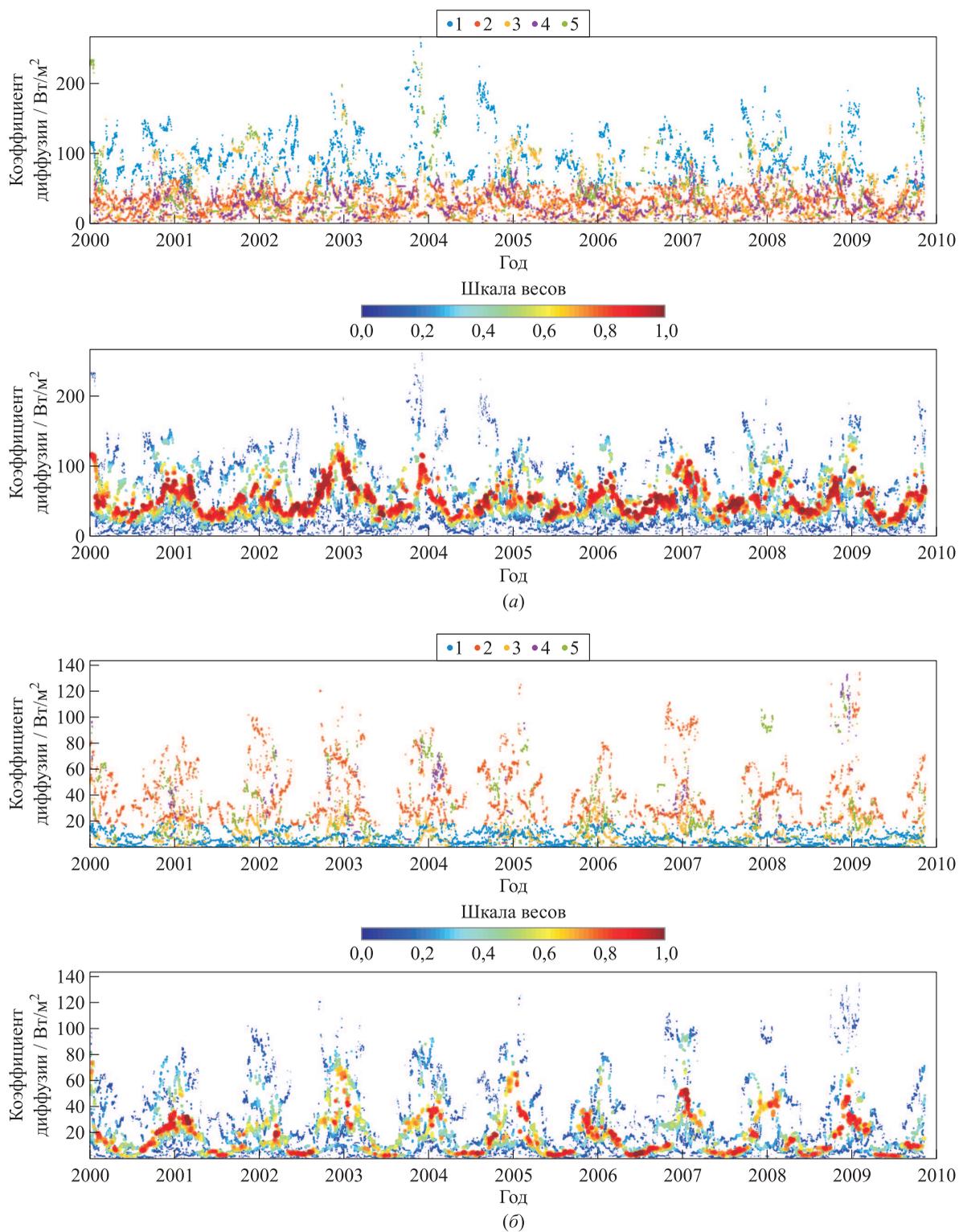
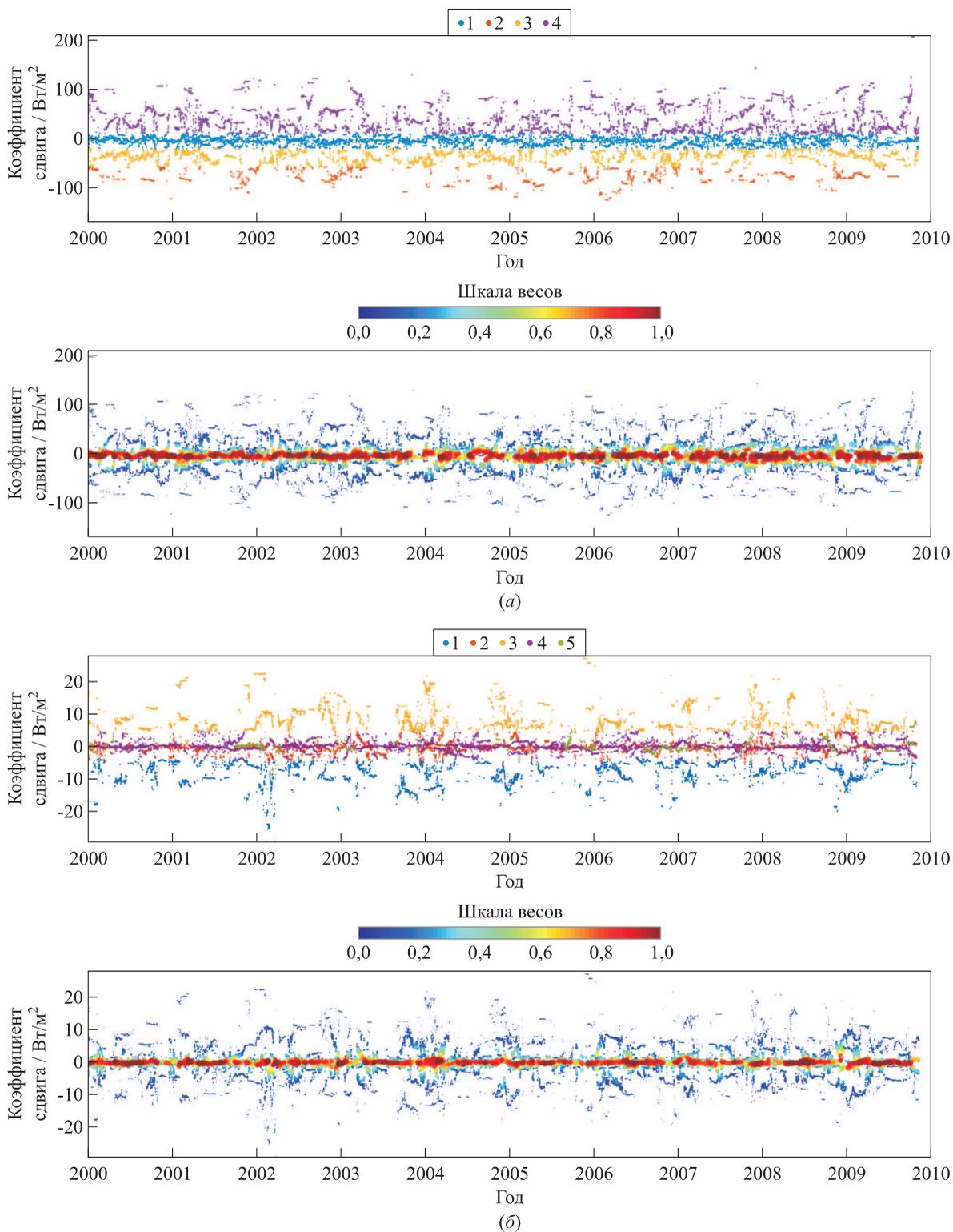
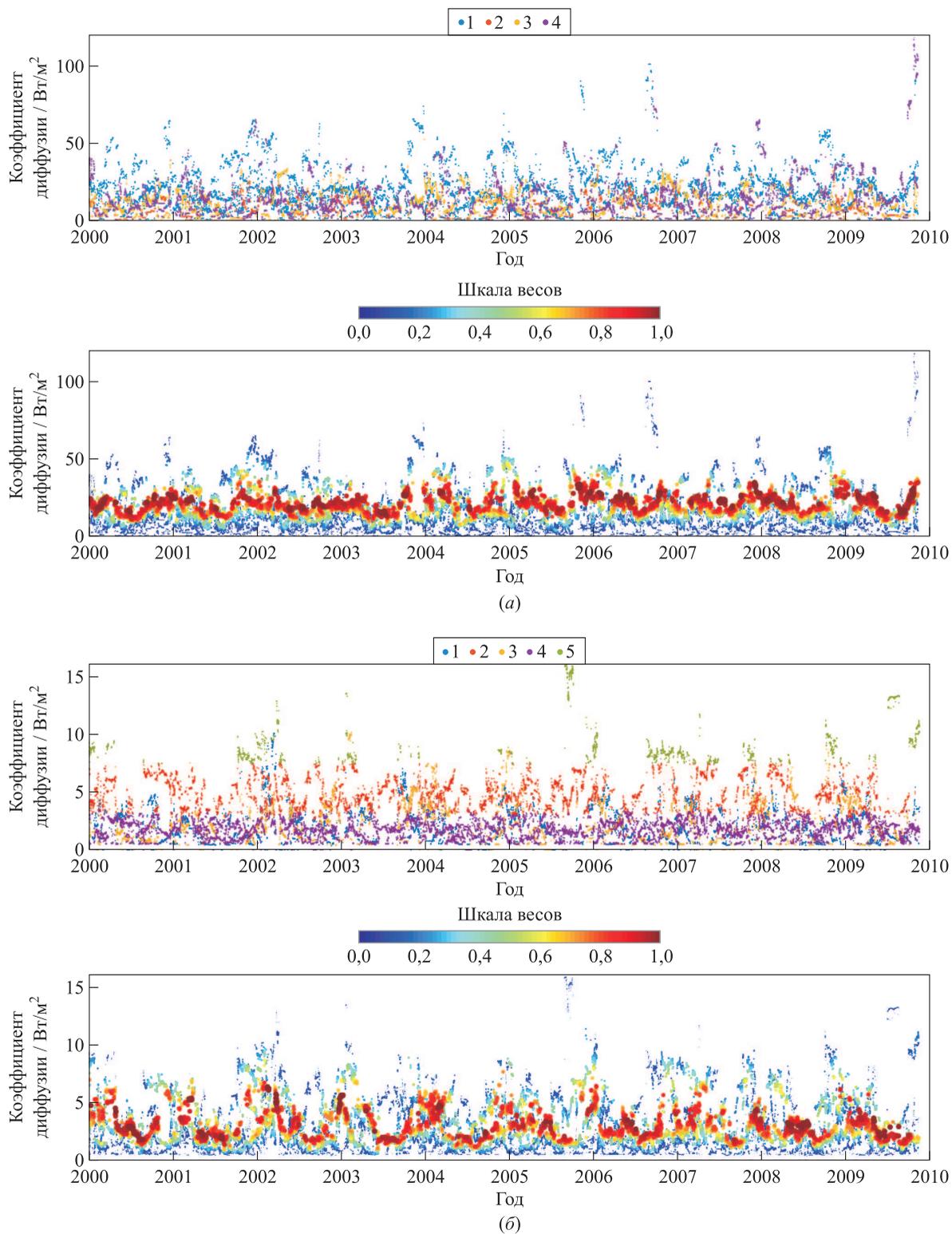


Рис. 2 Оценки распределения коэффициента диффузии (Гольфстрим): (а) явные потоки; (б) скрытые потоки; 1–5 — структурные компоненты



**Рис. 3** Оценки распределения сдвига (тропики): (а) явные потоки; (б) скрытые потоки; 1–5 — структурные компоненты



**Рис. 4** Оценки распределения коэффициента диффузии (тропики): (а) явные потоки; (б) скрытые потоки; 1–5 — структурные компоненты

2. *Leland H. E.* Option pricing and replication with transactions costs // *J. Financ.*, 1985. Vol. 40. P. 1283–1301.
3. *Barles G., Soner H. M.* Option pricing with transaction costs and a nonlinear Black–Scholes equation // *Financ. Stoch.*, 1998. Vol. 2. P. 369–397.
4. *Heston S. L.* A closed-form solution for options with stochastic volatility, with application to bond and currency options // *Rev. Financ. Stud.*, 1993. Vol. 6. P. 327–343.
5. *Cox J. C., Ingersoll J. E., Ross S. A.* A theory of the term structure of interest rates // *Econometrica*, 1985. Vol. 53. P. 385–407.
6. *Hull J., White A.* The pricing of options on assets with stochastic volatilities // *J. Financ.*, 1987. Vol. 42. P. 281–308.
7. *Dupire B.* Pricing with a smile // *Risk*, 1994. Vol. 7. P. 18–20.
8. *Derman E., Kani J.* Riding on a smile // *Risk*, 1994. Vol. 7. P. 32–39.
9. *Шуряев А. Н.* Основы стохастической финансовой математики. Т. 1. Факты. Модели. — М.: МЦНМО, 2016. 440 с.
10. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Московского ун-та, 2011. 512 с.
11. *Горшенин А. К., Королев В. Ю., Турсунбаев А. М.* Медианные модификации EM-алгоритма для разделения смесей вероятностных распределений и их применение к декомпозиции волатильности финансовых индексов // *Статистические методы оценивания и проверки гипотез*, 2008. Т. 21. С. 169–195.
12. *Gorshenin A. K.* On implementation of EM-type algorithms in the stochastic models for a matrix computing on GPU // *AIP Conf. Proc.*, 2015. Vol. 1648. Art. No. 250008. 4 p.
13. *Steinhaus H.* Sur la division des corps materiels en parties // *B. Acad. Pol. Sci.*, 1956. Vol. 4. Iss. 12. P. 801–804.
14. *MacQueen J.* Some methods for classification and analysis of multivariate observations // *5th Berkeley Symposium on Mathematical Statistics and Probability Proceedings*, 1967. P. 281–297.
15. *Dunn J. C.* A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters // *J. Cybernetics*, 1973. Vol. 3. Iss. 3. P. 32–57.
16. *David A., Vassilvitskii S.* K-means++: The advantages of careful seeding // *18th Annual ACM-SIAM Symposium on Discrete Algorithms Proceedings*. — ACM, 2007. P. 1027–1035.
17. *Korolev V. Yu., Gorshenin A. K., Gulev S. K., Belyaev K. P.* Statistical modeling of air–sea turbulent heat fluxes by finite mixtures of Gaussian distributions // *Comm. Com. Inf. Sc.*, 2015. Vol. 564. P. 152–162.
18. *Горшенин А. К.* Концепция онлайн-комплекса для стохастического моделирования реальных процессов // *Информатика и её применения*, 2016. Т. 10. Вып. 1. С. 72–81.
19. *Perry A. H., Walker J. M.* Ocean–atmosphere system. — Upper Saddle River, NJ, USA: Prentice Hall Press, 1977. 180 p.

*Поступила в редакцию 15.07.20*

---



---

## STATISTICAL ESTIMATION OF DISTRIBUTIONS OF RANDOM COEFFICIENTS IN THE LANGEVIN STOCHASTIC DIFFERENTIAL EQUATION

A. K. Gorshenin<sup>1</sup>, V. Yu. Korolev<sup>1,2</sup>, and A. A. Shcherbinina<sup>2</sup>

<sup>1</sup>Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

<sup>2</sup>Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation

**Abstract:** A method is described for statistical estimation of the distributions of random coefficients of the Langevin stochastic differential equation (SDE) by the technique of moving separation of mixtures. Discrete approximations are proposed for these distributions. For the purpose of study of variability of the distributions of the SDE coefficients in time, an algorithm is proposed for sequential identification (determination of local connectivity) of the components of the resulting mixture distributions. This algorithm is based on combining a greedy algorithm for the determination of the number of components with a clustering method (*k*- or *c*-means). The application of the proposed method is illustrated by particular examples of the analysis of processes of heat transfer between atmosphere and ocean.

**Keywords:** mixture distribution; local connectivity; greedy algorithm; clustering

**DOI:** 10.14357/19922264200301

## Acknowledgments

The research was partially supported by the Russian Foundation for Basic Research (project 19-07-00914) and the RF Presidential scholarship program (project No. 538.2018.5). The research was conducted in accordance with the Program of Moscow Center for Fundamental and Applied Mathematics.

## References

1. Belyaev, K., A. Kuleshov, N. Tuchkova, and C. A. S. Tanajura. 2018. An optimal data assimilation method and its application to the numerical simulation of the ocean dynamics. *Math. Comp. Model. Dyn.* 1(24):12–25.
2. Leland, H. E. 1985. Option pricing and replication with transactions costs. *J. Financ.* 40:1283–1301.
3. Barles, G., and H. M. Soner. 1998. Option pricing with transaction costs and a nonlinear Black–Scholes equation. *Financ. Stoch.* 2:369–397.
4. Heston, S. L. 1993. A closed-form solution for options with stochastic volatility, with application to bond and currency options. *Rev. Financ. Stud.* 6:327–343.
5. Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53:385–407.
6. Hull, J., and A. White. 1987. The pricing of options on assets with stochastic volatilities. *J. Financ.* 42:281–308.
7. Dupire, B. 1994. Pricing with a smile. *Risk* 7:18–20.
8. Derman, E., and J. Kani. 1994. Riding on a smile. *Risk* 7:32–39.
9. Shiryayev, A. N. 2016. *Osnovy stokhasticheskoy finansovoy matematiki. T. 1. Fakty. Modeli* [Foundations of stochastic financial mathematics. Vol. 1. Facts. Models]. Moscow: MCCME. 440 p.
10. Korolev, V. Yu. 2011. *Veroyatnostno-statisticheskie metody dekompozitsii volatil'nosti khaoticheskikh protsessov* [Probabilistic and statistical methods of decomposition of volatility of chaotic processes]. Moscow: Izd-vo Moskovskogo un-ta. 512 p.
11. Gorshenin, A. K., V. Yu. Korolev, and A. M. Tursunbaev. 2017. Median modifications of the EM-algorithm for separation of mixtures of probability distributions and their applications to the decomposition of volatility of financial indexes. *J. Math. Sci.* 227(2):176–195.
12. Gorshenin, A. K. 2015. On implementation of EM-type algorithms in the stochastic models for a matrix computing on GPU. *AIP Conf. Proc.* 1648:250008. 4 p.
13. Steinhaus, H. 1956. Sur la division des corps materiels en parties. *B. Acad. Pol. Sci.* 4(12):801–804.
14. MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. *5th Berkeley Symposium on Mathematical Statistics and Probability Proceedings.* 281–297.
15. Dunn, J. C. 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybernetics* 3(3):32–57.
16. David, A., and S. Vassilvitskii. 2007. K-means++: The advantages of careful seeding. *18th Annual ACM-SIAM Symposium on Discrete Algorithms Proceedings.* ACM. 1027–1035.
17. Korolev, V. Yu., A. K. Gorshenin, S. K. Gulev, and K. P. Belyaev. 2015. Statistical modeling of air–sea turbulent heat fluxes by finite mixtures of Gaussian distributions. *Comm. Com. Inf. Sc.* 564:152–162.
18. Gorshenin, A. K. 2016. Kontseptsiya onlayn-kompleksa dlya stokhasticheskogo modelirovaniya real'nykh protsessov [Concept of online service for stochastic modeling of real processes]. *Informatika i ee Primeneniya — Inform. Appl.* 10(1):72–81.
19. Perry, A. H., and J. M. Walker. 1977. *Ocean–atmosphere system.* Upper Saddle River, NJ: Prentice Hall Press. 180 p.

Received July 15, 2020

## Contributors

**Gorshenin Andrey K.** (b. 1986) — Candidate of Science (PhD) in physics and mathematics, associate professor, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; leading scientist, Faculty of Computational Mathematics and Cybernetics and Moscow Center for Fundamental and Applied Mathematics, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; agorshenin@frcsc.ru

**Korolev Victor Yu.** (b. 1954) — Doctor of Science in physics and mathematics, professor, head of department, Faculty of Computational Mathematics and Cybernetics, and principal scientist, Moscow Center for Fundamental and Applied Mathematics, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vkorolev@cs.msu.ru

**Shcherbinina Anastasia A.** (b. 1997) — Master student, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; shcherbinina.aa.97@gmail.com

# О МАРКОВСКИХ И РАЦИОНАЛЬНЫХ ПОТОКАХ СЛУЧАЙНЫХ СОБЫТИЙ. I\*

В. А. Наумов<sup>1</sup>, К. Е. Самуйлов<sup>2</sup>

**Аннотация:** Статья представляет собой первую часть обзора, призванного ознакомить заинтересованных читателей с основами теории марковских потоков событий для более подробного изучения и облегчения применения этих моделей на практике. В первой части приведены свойства общих марковских потоков событий и показана их связь с марковскими аддитивными процессами и процессами марковского восстановления. Во второй части обзора будут рассмотрены частные случаи таких потоков — подклассы марковских потоков событий, а именно: простые и групповые потоки однородных и неоднородных событий, важные для приложений. Далее будет показано, как свойства марковских потоков событий связаны с мультипликативностью стационарных распределений марковских систем. В завершение обзора будут обсуждены матрично-экспоненциальные распределения и рациональные потоки событий, расширяющие возможности марковских потоков для моделирования сложных систем и при этом сохраняющие удобство их анализа с помощью вычислительной техники.

**Ключевые слова:** марковские процессы; марковские аддитивные процессы; потоки без последействия; МС-потоки

**DOI:** 10.14357/19922264200302

## 1 Введение

При анализе стохастических систем применяются различные модели потоков случайных событий. Марковские потоки событий служат обобщением марковски-модулированных пуассоновских потоков [1–3], образованных моментами изменения состояний цепей Маркова [4, 5], и процессов восстановления фазового типа [6]. Эти модели появились, когда стало ясным, что будущие методы анализа стохастических систем будут основаны на использовании вычислительной техники. Марковские потоки задаются с помощью нескольких матриц, что делает удобным использование вычислительной техники для их анализа.

Свойства марковских потоков вытекают из общей теории двухкомпонентных марковских процессов, однородных во времени и по одной из компонент, сейчас известных под именем марковские аддитивные процессы (Markov Additive Processes). Считающий процесс марковских потоков — это процесс с независимыми приращениями, заданный на цепи Маркова. Потоки событий, для которых считающий процесс имеет независимые приращения, были названы А. Я. Хинчиным потоками без последействия [7]. Поэтому марковские потоки назывались потоками без последействия, заданными

на цепи Маркова [8], до тех пор пока на одном из семинаров в МГУ Г. П. Климов не предложил для этого класса потоков сокращенное название *МС-потоки* (от Markov Chain).

После того как М. F. Neuts предложил свою модель марковских потоков [9, 10], такие модели стали называть N-процессами [11–14]. В 1990 г. для марковских потоков было придумано новое название Markovian Arrival Process и стало использоваться неудачное сокращение MAP [15]. Неудачное потому, что начиная с опубликованной в 1972 г. работы E. Cinlar [16] это сокращение уже использовалось для обозначения марковских аддитивных процессов. Чтобы избежать путаницы в 2003 г. S. Asmussen в своей книге [17] предложил использовать для Markovian Arrival Process сокращение MAP, но по-прежнему как для Markov Additive Processes, так и для Markovian Arrival Process чаще используется сокращение MAP.

Цель настоящего обзора, состоящего из двух частей, — ознакомление читателей с основами теории марковских потоков, с тем чтобы облегчить применение этих моделей на практике и, если потребуется, их более подробное изучение. На сегодня опубликовано большое число работ, посвященных марковским и рациональным потокам событий, и в настоящем обзоре представлены лишь

\* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-17-50126.

<sup>1</sup> Исследовательский институт инноваций, г. Хельсинки, Финляндия, valeriy.naumov@pfu.fi

<sup>2</sup> Российский университет дружбы народов; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, samouylov-ke@rudn.ru

основные сведения о них. Многие результаты были взяты непосредственно из ссылок, и их доказательства намеренно опущены. Их можно найти в публикациях, перечисленных в списке литературы. Начинается обзор с рассмотрения случайных величин фазового типа в разд. 2. В разд. 3 определяются марковские потоки общего вида и показана их связь с марковскими аддитивными процессами и процессами марковского восстановления. Во второй части обзора рассматриваются важные для приложений подклассы марковских потоков однородных и неоднородных событий. Далее там же обсуждаются матрично-экспоненциальные распределения и рациональные потоки событий, которые расширяют возможности марковских потоков для моделирования сложных систем и при этом сохраняют удобство их анализа.

В этой работе жирные строчные буквы обозначают векторы, а жирные прописные буквы — матрицы. Неравенства  $\mathbf{x} \leq \mathbf{y}$  и  $\mathbf{x} \geq \mathbf{y}$  понимаются поэлементно;  $\delta(i, j) = 1$ , если  $i = j$ , и  $\delta(i, j) = 0$  в противном случае;  $\mathbf{e}_i$   $i$ -я координата равна единице, а остальные равны нулю. Кроме того, используются следующие обозначения:  $\mathbf{I} = [\delta(i, j)]$  — единичная матрица;  $\mathbf{u}$  — вектор-столбец из единиц;  $\text{Ind}(A)$  — индикатор события  $A$ ;  $\mathcal{N}^K$  — множество неотрицательных целочисленных векторов длины  $K$ ,  $\mathcal{N}_0^K = \mathcal{N}^K \setminus \{0\}$ . Для краткости вместо «наступило  $n_1$  событий типа 1,  $n_2$  событий типа 2, ...,  $n_K$  событий типа  $K$ » будем писать «наступило  $\mathbf{n}$  событий», где  $\mathbf{n} = (n_1, n_2, \dots, n_K)$ .

## 2 Случайные величины фазового типа

Распределения фазового типа представляют собой обобщение хорошо известного распределения Эрланга, появившегося в [18] и позднее названного по имени автора. Переводы этой и других работ А. К. Эрланга на английский язык опубликованы в специальном издании [19], посвященном его 70-летию. Распределения фазового типа — это распределения времени до попадания конечной цепи Маркова в поглощающее состояние [6]. Пусть  $X(t)$ ,  $t \geq 0$ , — случайный процесс с конечным множеством состояний  $\mathcal{X} = \{1, 2, \dots, L\}$  и  $\tau$  — неотрицательная случайная величина,  $P(\tau > 0) > 0$ . Случайный процесс  $X_0(t) = X(t)\text{Ind}(\tau > t)$  принимает значения  $0, 1, 2, \dots, L$ , причем его нулевое состояние — поглощающее. Случайная величина  $\tau$  называется случайной величиной фазового типа с фазовым процессом  $X(t)$ , если процесс  $X_0(t)$  — однородный марковский процесс, вероятности переходов которого

$$p(i, j, t) = P(X(t) = j, \tau > t | X(0) = i, \tau > 0), \quad i, j \in \mathcal{X},$$

удовлетворяют условиям:

$$\begin{aligned} \lim_{t \rightarrow 0} p(i, j, t) &= \delta(i, j), \quad i, j \in \mathcal{X}; \\ \lim_{t \rightarrow \infty} p(i, j, t) &= 0, \quad i, j \in \mathcal{X}. \end{aligned} \quad (1)$$

В этом случае существует матрица интенсивностей переходов  $\mathbf{S} = [s(i, j)]$  такая, что

$$p(i, j, \Delta t) = \delta(i, j) + s(i, j)\Delta t + o(\Delta t), \quad i, j \in \mathcal{X},$$

и для матрицы  $\mathbf{P}(t) = [p(i, j, t)]$  справедливо выражение [20]:

$$\mathbf{P}(t) = \exp(t\mathbf{S}). \quad (2)$$

Из равенств (1) и (2) следует, что все собственные числа матрицы  $\mathbf{S}$  имеют отрицательные действительные части [21]. Матрица  $\mathbf{S}$  невырождена, и матрица  $-\mathbf{S}^{-1}$  неотрицательна. В силу того что 0 является поглощающим состоянием процесса  $X_0(t)$ , его матрица интенсивностей переходов  $\mathbf{S}_0$  и матрица вероятностей переходов  $\mathbf{P}_0(t)$  имеют следующий вид:

$$\mathbf{S}_0 = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{s} & \mathbf{S} \end{bmatrix}; \quad \mathbf{P}_0(t) = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{p}(t) & \mathbf{P}(t) \end{bmatrix},$$

где  $\mathbf{s} = -\mathbf{S}\mathbf{u}$  и  $\mathbf{p}(t) = \mathbf{u} - \mathbf{P}(t)\mathbf{u}$  — векторы-столбцы длины  $L$ .

Для функции распределения случайной величины фазового типа  $F(t) = P(\tau \leq t)$  справедливы следующие выражения:

$$F(t) = 1 - \mathbf{q} \exp(t\mathbf{S})\mathbf{u}; \quad (3)$$

$$F(0) = 1 - \mathbf{q}\mathbf{u}; \quad \frac{d}{dt} F(t) = \mathbf{q} \exp(t\mathbf{S}), \quad t > 0,$$

где  $\mathbf{q} = [q(i)]$  — вектор-строка начальных вероятностей  $q(i) = P(X(0) = i, \tau > 0)$ . По определению выражение вида (3) представляет собой функцию распределения фазового типа, если вектор  $\mathbf{q}$  и матрица  $\mathbf{S}$  удовлетворяют неравенствам:

$$0 < \sum_{j \in \mathcal{X}} q(j) \leq 1, \quad q(i) \geq 0, \quad i \in \mathcal{X};$$

$$\sum_{j \in \mathcal{X}} s(i, j) \leq 0, \quad s(i, j) \geq 0, \quad i \neq j, \quad i, j \in \mathcal{X}.$$

Кроме того, необходимо, чтобы из любого состояния марковского процесса  $X_0(t)$  с матрицей интенсивностей переходов  $\mathbf{S}_0$  можно было попасть в поглощающее состояние 0. Для этого достаточно, чтобы матрица  $\mathbf{S} + \mathbf{s}\mathbf{q}$  была неразложимой, а вектор  $\mathbf{s}$  ненулевым. Не будем здесь останавливаться на свойствах функций распределения фазового типа, поскольку они являются матрично-экспоненциальными функциями распределения, рассматриваемыми во второй части обзора.

### 3 Марковские потоки событий

Случайные величины фазового типа — это суммы некоторого числа случайных величин, имеющих экспоненциальные функции распределения. Это позволяет применять при анализе стохастических систем с такими случайными величинами метод этапов Эрланга. Использование марковских потоков событий позволяет распространить метод этапов Эрланга на потоки случайных событий.

Рассмотрим некоторый поток групп неоднородных событий  $(t_l, \sigma_l)$ ,  $l = 1, 2, \dots$ , где  $0 < t_1 < t_2 < \dots$  — моменты наступления событий, также называемые вызывающими моментами [22], а  $\sigma_l$  — это вектор  $\sigma_l = (\sigma_{l,1}, \dots, \sigma_{l,K})$ , в котором  $\sigma_{l,k}$  есть размер группы событий типа  $k$ , наступивших в момент  $t_l$ ,  $k = 1, \dots, K$ . Обозначим через  $N_k(t) = \sum_{t_l \leq t} \sigma_{l,k}$  число событий типа  $k$ , наступивших за время  $t$ ,  $\mathbf{N}(t) = (N_1(t), N_2(t), \dots, N_K(t))$ . Будем называть поток  $(t_l, \sigma_l)$  марковским, если для некоторого случайного процесса  $X(t)$  с конечным множеством состояний  $\mathcal{X} = \{1, 2, \dots, L\}$  процесс  $\xi(t) = (X(t), \mathbf{N}(t))$  является однородным марковским процессом, причем для любых моментов времени  $0 \leq h < t$  вектор числа событий  $\mathbf{N}(t) - \mathbf{N}(h)$ , наступивших в интервале  $(h, t]$ , при заданном  $X(h)$  не зависит от вектора числа событий  $\mathbf{N}(h)$ , наступивших до момента  $h$  [23]. Другими словами, для марковского потока процесс  $\xi(t) = (X(t), \mathbf{N}(t))$  однороден во времени и по второй компоненте [24]. Будем предполагать траектории процесса  $\xi(t) = (X(t), \mathbf{N}(t))$  непрерывными справа. Его вероятности переходов

$$p_{\mathbf{k},\mathbf{n}}(i, j, t) = P(X(h+t) = j, \mathbf{N}(h+t) = \mathbf{n} | X(h) = i, \mathbf{N}(h) = \mathbf{k}) \quad (4)$$

удовлетворяют условию:

$$p_{\mathbf{k},\mathbf{n}}(i, j, t) = p_{\mathbf{0},\mathbf{n}-\mathbf{k}}(i, j, t), \quad i, j \in \mathcal{X}, \mathbf{k}, \mathbf{n} \in \mathcal{N}^K, \mathbf{k} \leq \mathbf{n}, \quad (5)$$

поэтому они однозначно определяются функциями

$$p_{\mathbf{n}}(i, j, t) = p_{\mathbf{0},\mathbf{n}}(i, j, t), \quad i, j \in \mathcal{X}, \mathbf{n} \in \mathcal{N}^K.$$

Из (4)–(5) следует, что интенсивности переходов процесса  $\xi(t)$

$$a_{\mathbf{k},\mathbf{n}}(i, j) = \lim_{t \rightarrow 0} \frac{1}{t} (p_{\mathbf{k},\mathbf{n}}(i, j, t) - \delta(\mathbf{k}, \mathbf{n})\delta(i, j)), \quad i, j \in \mathcal{X}, \mathbf{k}, \mathbf{n} \in \mathcal{N}^K, \mathbf{k} \leq \mathbf{n},$$

имеют следующий вид:

$$a_{\mathbf{k},\mathbf{n}}(i, j) = \begin{cases} a_{\mathbf{n}-\mathbf{k}}(i, j), & \text{если } \mathbf{k} \leq \mathbf{n}; \\ 0 & \text{в остальных случаях,} \end{cases}$$

где

$$a_{\mathbf{0}}(i, j) = \lim_{t \rightarrow 0} \frac{1}{t} (p_{\mathbf{0}}(i, j, t) - \delta(i, j)), \quad i, j \in \mathcal{X};$$

$$a_{\mathbf{n}}(i, j) = \lim_{t \rightarrow 0} \frac{1}{t} p_{\mathbf{n}}(i, j, t), \quad i, j \in \mathcal{X}, \mathbf{n} \in \mathcal{N}_0^K.$$

Матрицы вероятностей переходов  $\mathbf{P}_{\mathbf{n}}(t) = [p_{\mathbf{n}}(i, j, t)]$  однозначно определяются матрицами интенсивностей переходов  $\mathbf{A}_{\mathbf{n}} = [a_{\mathbf{n}}(i, j)]$ ,  $\mathbf{n} \geq \mathbf{0}$ . Действительно, матрицы  $\mathbf{P}_{\mathbf{n}}(t)$  образуют единственное решение системы уравнений Колмогорова

$$\frac{d}{dt} \mathbf{P}_{\mathbf{n}}(t) = \sum_{\substack{\mathbf{i} \in \mathcal{N}^K, \\ \mathbf{i} \leq \mathbf{n}}} \mathbf{A}_{\mathbf{i}} \mathbf{P}_{\mathbf{n}-\mathbf{i}}(t), \quad \mathbf{n} \in \mathcal{N}^K, \quad (6)$$

удовлетворяющее начальным условиям:

$$\mathbf{P}_{\mathbf{0}}(0) = \mathbf{I}, \quad \mathbf{P}_{\mathbf{n}}(0) = \mathbf{0}, \quad \mathbf{n} \in \mathcal{N}_0^K. \quad (7)$$

Решение системы уравнений (6)–(7) можно найти с помощью рекуррентной формулы

$$\mathbf{P}_{\mathbf{0}}(t) = \exp(t\mathbf{A}_{\mathbf{0}}),$$

$$\mathbf{P}_{\mathbf{n}}(t) = \sum_{\substack{\mathbf{i} \in \mathcal{N}_0^K, \\ \mathbf{i} \leq \mathbf{n}}} \int_0^t \mathbf{P}_{\mathbf{0}}(t-h) \mathbf{A}_{\mathbf{i}} \mathbf{P}_{\mathbf{n}-\mathbf{i}}(h) dh, \quad \mathbf{n} \in \mathcal{N}_0^K.$$

Из (6) следует, что матрицы

$$\mathbf{P}(\mathbf{z}, t) = \sum_{\mathbf{n} \in \mathcal{N}^K} \mathbf{P}_{\mathbf{n}}(t) z_1^{n_1} \dots z_K^{n_K};$$

$$\mathbf{A}(\mathbf{z}) = \sum_{\mathbf{n} \in \mathcal{N}^K} \mathbf{A}_{\mathbf{n}} z_1^{n_1} \dots z_K^{n_K},$$

где  $\mathbf{n} = (n_1, \dots, n_K)$ ,  $\mathbf{z} = (z_1, \dots, z_K)$ ,  $|z_k| \leq 1$ ,  $k = 1, \dots, K$ , связаны простым равенством:

$$\mathbf{P}(\mathbf{z}, t) = \exp(t\mathbf{A}(\mathbf{z})).$$

Первая компонента  $X(t)$  однородного во времени и по второй компоненте процесса  $\xi(t) = (X(t), \mathbf{N}(t))$  сама является однородным марковским процессом [24]. Будем называть  $X(t)$  фазовым процессом марковского потока. Матрица вероятностей переходов  $\mathbf{P}(t)$  этого процесса дается формулой:

$$\mathbf{P}(t) = \sum_{\mathbf{n} \in \mathcal{N}^K} \mathbf{P}_{\mathbf{n}}(t).$$

Матрицы интенсивностей переходов  $\mathbf{A}_{\mathbf{n}}$  обладают следующими свойствами:

- (1) внедиагональные элементы матрицы  $\mathbf{A}_0$  неотрицательны;
- (2) матрицы  $\mathbf{A}_n, n \in \mathcal{N}_0^K$ , неотрицательны;
- (3) матрица

$$\mathbf{A} = \sum_{n \in \mathcal{N}^K} \mathbf{A}_n$$

является матрицей интенсивностей переходов фазового процесса  $X(t)$ .

В дальнейшем будем предполагать, что матрица интенсивностей переходов  $\mathbf{A}$  неразложима,  $\mathbf{\Lambda} = \mathbf{A} - \mathbf{A}_0 \neq \mathbf{0}$ , и обозначать через  $\mathbf{p}$  вектор-строку стационарных вероятностей фазового процесса. Этот вектор является единственным решением системы линейных уравнений

$$\mathbf{p}\mathbf{A} = \mathbf{0}; \quad \mathbf{p}\mathbf{u} = 1.$$

Существуют два основных способа задания марковского потока. Обычно он задается набором матриц  $\mathbf{A}_n, n \in \mathcal{N}^K$ . При другом способе, называемом конструктивным, поток задается с помощью набора чисел  $\gamma(i) > 0, i \in \mathcal{X}$ , и неотрицательных матриц  $\Psi_n = [\psi_n(i, j)], n \in \mathcal{N}^K$ , таких что

$$\sum_{j \in \mathcal{X}} \sum_{n \in \mathcal{N}^K} \psi_n(i, j) = 1, \quad i \in \mathcal{X}.$$

Эти параметры описывают однородный марковский процесс, у которого время пребывания в каждом состоянии  $i$  имеет экспоненциальное распределение с параметром  $\gamma(i)$ . В момент его окончания с вероятностью  $\psi_n(i, j)$  наступает  $\mathbf{n}$  событий потока и процесс переходит из состояния  $i$  в состояние  $j$ . Так определенный марковский поток имеет матрицы  $\mathbf{A}_n$  с элементами:

$$a_n(i, j) = \gamma(i)(\psi_n(i, j) - \delta(\mathbf{n}, \mathbf{0})\delta(i, j)), \\ i, j \in \mathcal{X}, \quad \mathbf{n} \in \mathcal{N}^K.$$

При заданных матрицах  $\mathbf{A}_n$  параметры конструктивного способа задания потока можно получить следующим образом. Сначала надо выбрать величины  $\gamma(i)$  так, чтобы выполнялись неравенства  $\gamma(i) \geq -a_0(i, i)$  для всех  $i$ . После этого вероятности переходов  $\psi_n(i, j)$  найти по формуле:

$$\psi_n(i, j) = \frac{a_n(i, j)}{\gamma(i)} + \delta(\mathbf{n}, \mathbf{0})\delta(i, j), \\ i, j \in \mathcal{X}, \quad \mathbf{n} \in \mathcal{N}^K.$$

Ясно, что параметры конструктивного способа задания потока не единственны и зависят от выбранных величин  $\gamma(i)$ . Их, например, можно взять равными  $\gamma(i) = \gamma, i \in \mathcal{X}$ , где  $\gamma \geq \max_i |a_0(i, i)|$ . Тогда моменты времени, в которые возможно наступление событий заданного таким способом мар-

ковского потока, образуют пуассоновский поток интенсивности  $\gamma$ .

Зная матрицы вероятностей переходов  $\mathbf{P}_n(t)$ , можно найти распределение числа событий, наступивших в непересекающиеся интервалы времени:

$$p_{\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m}(x_0, x_1, \dots, x_m) = \\ = \mathbf{P} \left( \mathbf{N} \left( \sum_{j=0}^r x_j \right) - \mathbf{N} \left( \sum_{j=0}^{r-1} x_j \right) = \mathbf{k}_r, \right. \\ \left. r = 1, 2, \dots, m \right) = \\ = \boldsymbol{\alpha} \mathbf{P}(x_0) \mathbf{P}_{\mathbf{k}_1}(x_1) \mathbf{P}_{\mathbf{k}_2}(x_2) \cdots \mathbf{P}_{\mathbf{k}_m}(x_m) \mathbf{u}, \\ \mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m \in \mathcal{N}^K, \quad x_0, x_1, \dots, x_m > 0, \\ m = 1, 2, \dots, \quad (8)$$

где  $\boldsymbol{\alpha} = [\alpha(i)]$  — начальное распределение фазового процесса,  $\alpha(i) = \mathbf{P}(X(0) = i)$ . Разным начальным распределениям фазового процесса соответствуют разные марковские потоки. В случае  $\boldsymbol{\alpha} = \mathbf{p}$  говорят о стационарной версии марковского потока [23]. При начальном распределении фазового процесса  $\boldsymbol{\alpha} = \mathbf{p}$  считающий процесс  $\mathbf{N}(t)$  марковского потока является процессом со стационарными приращениями, поскольку распределение (8) числа событий, наступивших в интервалах длины  $x_1, \dots, x_m$ , не зависит от начального положения  $x_0$  этих интервалов на оси времени. Параметр  $\lambda$  стационарной версии марковского потока и вектор математических ожиданий числа одновременно наступивших событий  $\mathbf{m} = \mathbf{M}(\boldsymbol{\sigma}_l)$  вычисляются по формулам:

$$\lambda = \mathbf{p}\mathbf{\Lambda}\mathbf{u}; \quad \mathbf{m} = \frac{1}{\lambda} \sum_{n \in \mathcal{N}_0^K} (\mathbf{p}\mathbf{a}_n)\mathbf{n},$$

где  $\mathbf{a}_n = \mathbf{A}_n\mathbf{u}$ .

Полагая  $t_0 = 0$ , обозначим через  $\tau_l = t_l - t_{l-1}, l = 1, 2, \dots$ , длины интервалов между моментами наступления событий и через  $X_l = X(t_l), l = 1, 2, \dots$ , состояния фазового процесса в моменты после наступления событий потока. Марковский поток групп событий является полумарковским потоком, поскольку последовательность  $(X_l, \boldsymbol{\sigma}_l, \tau_l), l = 1, 2, \dots$ , — процесс марковского восстановления [25]. Условное распределение последующего состояния этого процесса зависит только от текущего состояния фазового процесса:

$$\mathbf{P}(X_l = j, \boldsymbol{\sigma}_l = \mathbf{n}, \tau_l < x | X_{l-1} = i, \boldsymbol{\sigma}_{l-1} = \mathbf{k}, \\ \tau_{l-1} < y) = \\ = \mathbf{P}(X_l = j, \boldsymbol{\sigma}_l = \mathbf{n}, \tau_l < x | X_{l-1} = i) = G_n(i, j, x).$$

Матрицы  $\mathbf{G}_n(x) = [G_n(i, j, x)]$ , описывающие связанный с марковским потоком процесс марков-

ского восстановления  $(X_l, \sigma_l, \tau_l)$ , и их преобразования Лапласа-Стилтьеса имеют следующий вид:

$$\begin{aligned} \mathbf{G}_n(x) &= \int_0^x \exp(z\mathbf{A}_0)\mathbf{A}_n dz = \\ &= (\exp(x\mathbf{A}_0) - \mathbf{I})\mathbf{A}_0^{-1}\mathbf{A}_n, \quad \mathbf{n} \in \mathcal{N}_0^K; \quad (9) \\ \int_0^\infty e^{-vx} d\mathbf{G}_n(x) &= (v\mathbf{I} - \mathbf{A}_0)^{-1}\mathbf{A}_n, \quad \mathbf{n} \in \mathcal{N}_0^K. \end{aligned}$$

Используя матрицы  $\mathbf{G}_n(x)$ , можно найти совместное распределение числа  $\sigma_l$  наступивших событий и длин  $\tau_l$  интервалов между вызывающими моментами, а также плотность этого распределения:

$$\begin{aligned} F_{\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m}(x_1, x_2, \dots, x_m) &= \\ &= P(\sigma_l = \mathbf{k}_l, \tau_l < x_l, l = 1, 2, \dots, m) = \\ &= \alpha \mathbf{G}_{\mathbf{k}_1}(x_1) \mathbf{G}_{\mathbf{k}_2}(x_2) \cdots \mathbf{G}_{\mathbf{k}_m}(x_m) \mathbf{u}; \quad (10) \\ f_{\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m}(x_1, x_2, \dots, x_m) &= \\ &= \alpha \exp(x_1 \mathbf{A}_0) \mathbf{A}_{\mathbf{k}_1} \exp(x_2 \mathbf{A}_0) \mathbf{A}_{\mathbf{k}_2} \cdots \\ &\quad \cdots \exp(x_m \mathbf{A}_0) \mathbf{A}_{\mathbf{k}_m} \mathbf{u}, \end{aligned}$$

где

$$\begin{aligned} \mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m &\in \mathcal{N}_0^K; \\ x_0, x_1, \dots, x_m &> 0, \quad m = 1, 2, \dots \end{aligned}$$

Вероятности переходов вложенной цепи Маркова  $(X_l, \sigma_l)$  зависят лишь от состояния фазового процесса после предыдущего вызывающего момента:

$$\begin{aligned} P(X_l = j, \sigma_l = \mathbf{n} | X_{l-1} = i, \sigma_{l-1} = \mathbf{k}) &= \\ &= P(X_l = j, \sigma_l = \mathbf{n} | X_{l-1} = i) = Q_n(i, j), \end{aligned}$$

где  $Q_n(i, j) = \lim_{x \rightarrow \infty} G_n(i, j, x)$ . Из (9) следует, что матрица вероятностей переходов  $\mathbf{Q}_n = [Q_n(i, j)]$  вложенной цепи Маркова  $(X_l, \sigma_l)$  и матрица вероятностей переходов  $\mathbf{Q}$  цепи Маркова  $X_l$  имеют следующий вид:

$$\mathbf{Q} = -\mathbf{A}_0^{-1}\mathbf{A}, \quad \mathbf{Q}_n = -\mathbf{A}_0^{-1}\mathbf{A}_n, \quad \mathbf{n} \in \mathcal{N}_0^K.$$

Стационарные распределения  $\mathbf{q} = [q(i)]$  и  $\mathbf{q}_n = [q_n(i)]$ ,  $\mathbf{n} \in \mathcal{N}_0^K$ , вложенных цепей Маркова  $X_l$  и  $(X_l, \sigma_l)$  связаны со стационарным распределением  $\mathbf{p}$  процесса  $X(t)$  следующими равенствами:

$$\begin{aligned} \mathbf{q} &= \frac{1}{\lambda} \mathbf{p} \mathbf{A}; \quad \mathbf{p} = -\lambda \mathbf{q} \mathbf{A}_0^{-1}, \\ \mathbf{q} &= \sum_{\mathbf{n} \in \mathcal{N}_0^K} \mathbf{q}_n; \quad \mathbf{q}_n = \frac{1}{\lambda} \mathbf{p} \mathbf{A}_n, \quad \mathbf{n} \in \mathcal{N}_0^K. \end{aligned}$$

Если вектор из единиц  $\mathbf{u}$  является правым собственным вектором каждой из матриц  $\mathbf{A}_n$  и выполняются равенства

$$\mathbf{A}_n \mathbf{u} = \lambda_n \mathbf{u}, \quad \mathbf{n} \in \mathcal{N}_0^K,$$

то из (10) следует, что при любом начальном распределении  $\alpha$  марковский поток будет стационарным потоком без последствия. Аналогично, если вектор стационарных вероятностей  $\mathbf{p}$  является левым собственным вектором матриц  $\mathbf{A}_n$  и выполняются равенства

$$\mathbf{p} \mathbf{A}_n = \lambda_n \mathbf{p}, \quad \mathbf{n} \in \mathcal{N}_0^K,$$

то при начальном распределении  $\alpha = \mathbf{p}$  марковский поток будет стационарным потоком без последствия. В обоих случаях вызывающие моменты марковского потока образуют пуассоновский поток интенсивности

$$\lambda = \sum_{\mathbf{n} \in \mathcal{N}_0^K} \lambda_n,$$

а векторы  $\sigma_l$  числа наступивших в эти моменты событий независимы и имеют распределение вероятностей

$$P(\sigma_l = \mathbf{n}) = \frac{\lambda_n}{\lambda}, \quad \mathbf{n} \in \mathcal{N}_0^K.$$

Различные операции над марковскими потоками, такие как суперпозиция, прореживание, линейные преобразования и др., рассмотрены в [23].

## 4 Заключение

В первой части обзора показано, как марковские модели потоков событий могут быть построены на базе теории двухкомпонентных марковских процессов, однородных по второй компоненте. Здесь также представлены основные свойства и характеристики марковских потоков событий. Вторая часть обзора посвящена описанию важных для приложений подклассов марковских потоков как однородных, так и неоднородных событий. Кроме того, в ней будут рассмотрены современные модели матрично-экспоненциальных распределений и рациональных потоков событий.

## Литература

1. Башарин Г. П. О вычислении моментов избыточной нагрузки сложной системы // Электросвязь, 1970. № 4. С. 68–77.
2. Башарин Г. П., Кокотушкин В. А. Об условиях усиленного статистического равновесия для сложных сис-

- тем массового обслуживания // Проблемы передачи информации, 1971. Т. 7. № 3. С. 67–75.
3. Neuts M. F. A queue subject to extraneous phase changes // Adv. Appl. Probab., 1971. Vol. 3. P. 78–119.
  4. Башарин Г. П. О вычислении моментов обслуженной и избыточной нагрузок сложной системы // Техническая кибернетика, 1972. № 1. С. 42–51.
  5. Rudemo M. Point processes generated by transitions of Markov chains // Adv. Appl. Probab., 1973. Vol. 5. Iss. 2. P. 262–286.
  6. Neuts M. F. Probability distribution of phase type // Liber amicorum Professor emeritus Dr. H. Florin / Ed. R. Holvoet. — Louvain-la-Neuve, Belgium: Department of Mathematics, University of Louvain, 1975. P. 173–206.
  7. Хинчин А. Я. Потoki случайных событий без последнего действия // Теория вероятностей и ее применения, 1956. Т. 1. № 1. С. 3–18.
  8. Наумов В. А. Исследование некоторых многофазных систем массового обслуживания: Дис. ... канд. физ.-мат. наук. — М.: УДН, 1978. 98 с.
  9. Neuts M. F. A versatile Markovian point process. — Newark, DE, USA: Department of Statistics and Computer Science, University of Delaware, 1977. Technical Report 77/13. 29 p.
  10. Neuts M. F. A versatile Markovian point process // J. Appl. Probab., 1979. Vol. 16. Iss. 4. P. 764–779.
  11. Ramaswami V. The  $N/G/1$  queue and its detailed analysis // Adv. Appl. Probab., 1980. Vol. 12. Iss. 1. P. 222–261.
  12. Blondia C. The  $N/G/1$  finite capacity queue // Commun. Stat. Stochastic Models, 1989. Vol. 5. Iss. 2. P. 273–294.
  13. Saito H. The departure process of an  $N/G/1$  queue // Perform. Evaluation, 1990. Vol. 11. Iss. 4. P. 241–251.
  14. Lee G., Jeon J. A new approach to an  $N/G/1$  queue // Queueing Syst., 2000. Vol. 35. Iss. 1-4. P. 317–322.
  15. Lucantoni D. M., Meier-Hellstern K., Neuts M. F. A single-server queue with server vacations and a class of non-renewal arrival processes // Adv. Appl. Probab., 1990. Vol. 22. Iss. 3. P. 676–705.
  16. Çinlar E. Markov additive processes // Z. Wahrscheinlichkeit., 1972. Iss. 24. P. 85–93; 95–121.
  17. Asmussen S. Applied probability and queues. — New York, NY, USA: Springer, 2003. 438 p.
  18. Erlang A. K. Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler // Elektroteknikerens, 1917. Iss. 13. P. 5–13.
  19. Erlang A. K. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges // T. Danish Acad. Technical Sci., 1948. Iss. 2. P. 138–155.
  20. Булинский А. В., Ширяев А. Н. Теория случайных процессов. — М.: Физматлит, 2005, 402 с.
  21. Гантмахер Ф. Р. Теория матриц. — М.: Наука, 1966. 576 с.
  22. Хинчин А. Я. Математические методы теории массового обслуживания. — М.: АН СССР, 1955. 124 с.
  23. Pacheco A., Prabhu N. U. Markov-additive processes of arrivals // Advances in queueing theory, methods, and open problems / Ed. J. H. Dshalalow. — Boca Raton, FL, USA: CRC Press, 1995. P. 167–194.
  24. Ежов И. И., Скороход А. В. Марковские процессы, однородные по второй компоненте // Теория вероятностей и ее применения, 1969. Т. 14. № 1. С. 3–14; № 4. С. 679–692.
  25. Королюк В. С., Турбин А. Ф. Полумарковские процессы и их приложения. — Киев: Наукова Думка, 1976. 184 с.

Поступила в редакцию 02.07.20

## ON MARKOVIAN AND RATIONAL ARRIVAL PROCESSES. I

V. A. Naumov<sup>1</sup> and K. E. Samouylov<sup>2,3</sup>

<sup>1</sup>Service Innovation Research Institute, 8A Annankatu, Helsinki 00120, Finland

<sup>2</sup>Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

<sup>3</sup>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

**Abstract:** This article is the first part of a review carried out within the framework of the RFBR project No. 19-17-50126. The purpose of this review is to get the interested readers familiar with the basics of the theory of Markovian arrival processes to facilitate the application of these models in practice and, if necessary, to study them in detail. In the first part of the review, the properties of general Markovian arrival processes are presented and their relationship with Markov additive processes and Markov renewal processes is shown. In the second part of the review, the important for applications subclasses of Markovian arrival processes, i. e., simple and batch arrival processes of homogeneous and heterogeneous arrivals, are considered. After that, it is shown how the properties of Markovian arrival processes are associated with the product form of stationary distributions of Markov systems. In

conclusion, matrix-exponential distributions and rational arrival processes are discussed that expand the capabilities of Markovian arrival processes for modeling complex systems, while preserving the convenience of analyzing them using computations.

**Keywords:** Markov chain; Markovian arrival process; Markov additive process; MAP; MArP

**DOI:** 10.14357/19922264200302

## Acknowledgments

The reported study was funded by RFBR, project number 19-17-50126.

## References

1. Basharin, G. P. 1970. O vychislenii momentov izbytochnoy nagruzki slozhnoy sistemy [On calculating the moments of excessive load of a complex system]. *Elektrosvyaz'* [Electrosvyaz Magazine] 4:68–77.
2. Basharin, G. P., and V. A. Kokotushkin. 1971. Conditions for strong statistical equilibrium of complex mass servicing systems. *Probl. Inform. Transm.* 7(3):242–248.
3. Neuts, M. F. 1971. A queue subject to extraneous phase changes. *Adv. Appl. Probab.* 3:78–119.
4. Basharin, G. P. 1972. O vychislenii momentov obsluzhennoy i izbytochnoy nagruzok slozhnoy sistemy [On calculating the moments of serviced and redundant loads of a complex system]. *Tekhnicheskaya kibernetika* [Technical Cybernetics] 1:42–51.
5. Rudemo, M. 1973. Point processes generated by transitions of Markov chains. *Adv. Appl. Probab.* 5(2):262–286.
6. Neuts, M. F. 1975. Probability distribution of phase type. *Liber amicorum Professor emeritus Dr. H. Florin*. Ed. R. Holvoet. Louvain-la-Neuve, Belgium: Department of Mathematics, University of Louvain. 173–206.
7. Khinchin, A. Ya. 1956. Sequences of chance events without after-effects. *Theor. Probab. Appl.* 1(1):1–15.
8. Naumov, V. A. 1978. Issledovanie nekotorykh mnogofaznykh sistem massovogo obsluzhivaniya [Research of some multiphase queuing systems]. Moscow: UDN. PhD Thesis. 98 p.
9. Neuts, M. F. 1977. A versatile Markovian point process. Newark, DE: Department of Statistics and Computer Science, University of Delaware. Technical Report 77/13. 29 p.
10. Neuts, M. F. 1979. A versatile Markovian point process. *J. Appl. Probab.* 16(4):764–779.
11. Ramaswami, V. 1980. The  $N/G/1$  queue and its detailed analysis. *Adv. Appl. Probab.* 12(1):222–261.
12. Blondia, C. 1989. The  $N/G/1$  finite capacity queue. *Commun. Stat. Stochastic Models* 5(2):273–294.
13. Saito, H. 1990. The departure process of an  $N/G/1$  queue. *Perform. Evaluation* 11(4):241–251.
14. Lee, G., and J. Jeon. 2000. A new approach to an  $N/G/1$  queue. *Queueing Syst.* 35(1-4):317–322.
15. Lucantoni, D. M., K. Meier-Hellstern, and M. F. Neuts. 1990. A single-server queue with server vacations and a class of non-renewal arrival processes. *Adv. Appl. Probab.* 22(3):676–705.
16. Çinlar, E. 1972. Markov additive processes. *Z. Wahrscheinlichkeit.* 24:85–93; 95–121.
17. Asmussen, S. 2003. *Applied probability and queues*. New York, NY: Springer. 438 p.
18. Erlang, A. K. 1917. Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler. *Elektroteknikerne* 13:5–13.
19. Erlang, A. K. 1948. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *T. Danish Acad. Technical Sci.* 2:138–155.
20. Bulinskiy, A. V., and A. N. Shiryaev. 2005. *Teoriya sluchaynykh protsessov* [Theory of random processes]. Moscow: Fizmatlit. 402 p.
21. Gantmakher, F. R. 1966. *Teoriya matrits* [Matrix theory]. Moscow: Nauka. 576 p.
22. Khinchin, A. Ya. 1955. *Matematicheskie metody teorii massovogo obsluzhivaniya* [Mathematical methods of queuing theory]. Moscow: AS USSR. 124 p.
23. Pacheco, A., and N. U. Prabhu. 1995. Markov-additive processes of arrivals. *Advances in queuing theory, methods, and open problems*. Ed. J. H. Dshalalov. Boca Raton, FL: CRC Press. 167–194.
24. Ezhov, I. I., and A. V. Skorokhod. 1969. Markov processes with homogeneous second component. *Theor. Probab. Appl.* 14(1):1–13; 14(4):652–667.
25. Korolyuk, V. S., and A. F. Turbin. 1976. *Polumarkovskie protsessy i ikh prilozheniya* [Semi-Markov processes and their applications]. Kiev: Naukova Dumka. 184 p.

Received July 2, 2020

## Contributors

**Naumov Valeriy A.** (b. 1950) — Candidate of Science (PhD) in physics and mathematics, scientific director, Service Innovation Research Institute, 8A Annankatu, Helsinki 00120, Finland; valeriy.naumov@pfu.fi

**Samouylov Konstantin E.** (b. 1955) — Doctor of Science in technology, professor, Head of Department, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya St., Moscow 117198, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation; samouylov\_ke@rudn.university

# АППРОКСИМАЦИЯ МНОЖЕСТВА РЕШЕНИЙ СИСТЕМ НЕЛИНЕЙНЫХ НЕРАВЕНСТВ С ИСПОЛЬЗОВАНИЕМ ГРАФИЧЕСКИХ УСКОРИТЕЛЕЙ\*

М. В. Попов<sup>1</sup>, М. А. Посыпкин<sup>2</sup>

**Аннотация:** Существует множество задач, сводящихся к решению системы неравенств. Точное получение множества решений в подобных задачах не всегда возможно, из-за чего прибегают к различным методам аппроксимации данного множества. При повышении точности аппроксимации искомого множества увеличивается объем необходимых вычислений и, соответственно, время работы алгоритмов. В работе для увеличения быстродействия алгоритмов поиска аппроксимируемого множества применяются параллельные вычисления на графических ускорителях. Приводится описание и реализация последовательного метода аппроксимации системы неравенств и предлагается параллельный гибридный алгоритм, сочетающий перебор на равномерной сетке и идеи метода ветвей и границ. Этот алгоритм хорошо подходит для реализации на графических ускорителях и не приводит к избыточному перебору. Приведено сравнение эффективности работы последовательного и двух вариантов параллельного алгоритмов на примере прикладной задачи аппроксимации рабочей области робота. Рабочая область состоит из множества возможных положений инструмента и служит одной из ключевых характеристик робота.

**Ключевые слова:** оптимизация; параллельные вычисления; графический ускоритель, GPU; CUDA; нелинейные неравенства

**DOI:** 10.14357/19922264200303

## 1 Введение

В статье рассматривается задача построения внутренней и внешней аппроксимаций множества, задаваемого системой нелинейных неравенств. Для решения этой задачи разработаны подходы [1], основной недостаток которых заключается в высокой трудоемкости. В данной работе показано, что упомянутая проблема может быть преодолена с помощью методов высокопроизводительных вычислений.

В качестве платформы для высокопроизводительных вычислений рассматриваются современные графические ускорители (GPU, graphics processing units) [2]. Данный тип устройств получил широкое распространение в последнее время благодаря наилучшему соотношению цена/производительность. В то же время разработка программ для этой платформы является нетривиальной задачей в силу особенностей ее архитектуры. Современные графические ускорители обладают большим числом ядер, но возможности взаимодействия между потоками существенно ограничены по отношению к центральному процессору.

В работе предлагается гибридный алгоритм, сочетающий перебор на равномерной сетке и идеи

метода ветвей и границ. Предложенный алгоритм хорошо подходит для реализации на графических ускорителях и не приводит к избыточному перебору.

## 2 Последовательный алгоритм

Рассматривается задача нахождения множества решений системы неравенств вида:

$$\left. \begin{aligned} f_j(x) &\leq 0, \quad j \in [1, m]; \\ a_i &\leq x_i \leq b_i, \quad i \in [1, n]. \end{aligned} \right\} \quad (1)$$

Для поиска решения необходимо задать начальные границы, внутри которых будет находиться искомая область —  $n$ -мерный параллелепипед  $X \in R^n$ . Далее задается точность аппроксимации  $d$  получаемого решения, представляющая собой минимальный размер аппроксимирующего параллелепипеда. Под размером будем понимать длину максимального измерения. После этого параллелепипед  $X$  разбивается на меньшие параллелепипеды  $P_k$  равномерной сеткой. Число точек сетки по измерению  $j$  задается формулой:

$$k_j = \left\lceil \frac{b_j - a_j}{d} \right\rceil.$$

\* Работа выполнена при частичной поддержке РНФ (проект 16-19-00148).

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

<sup>2</sup>Федеральный исследовательский центр «Информатика и управление» Российской академии наук, mposypkin@gmail.com

Для каждого из полученных параллелепипедов проверяется выполнение следующих условий:

$$\max_{j \in [1, m]} \max_{x \in P_k} f_j(x) < 0; \quad (2)$$

$$\max_{j \in [1, m]} \min_{x \in P_k} f_{x \in P_k}(x) > 0. \quad (3)$$

Можно выделить три основных случая:

- (1) выполнено условие (2) — параллелепипеды полностью лежат внутри искомой области;
- (2) выполнено условие (3) — параллелепипед лежит вне искомой области;
- (3) оба условия не выполнены — параллелепипед классифицируется как «граничный».

Метод перебора на равномерной сетке удобен тем, что можно заранее рассчитать число операций, необходимых для получения аппроксимирующего покрытия фигуры, образованной функциями из (1).

Для последовательного алгоритма перебора на равномерной сетке число рассматриваемых параллелепипедов определяется по формуле  $\prod_{j=1}^m k_j$ . Для каждого параллелепипеда  $P_k \in X$  проводится проверка выполнения условий (2) и (3) для каждой функции  $f_j$ . Время проверки  $t_c$  не отличается для разных параллелепипедов. Таким образом, общее время составляет  $t_c \prod_{j=1}^m k_j$ .

Недостаток данного алгоритма состоит в необходимости полного перебора всех параллелепипедов, число которых существенно возрастает с уменьшением значения точности аппроксимации. Для преодоления указанного недостатка может применяться метод «ветвей и границ», адаптированный для данной задачи. Отличие заключается в том, что сетка не задается сразу, а параллелепипеды формируются рекурсивно. Приведем описание алгоритма.

- I. В текущий список помещается  $m$ -мерный параллелепипед  $X \in R^m$ , который полностью содержит в себе искомую область.
- II. Из текущего списка извлекается параллелепипед и делится на две части по наибольшему из измерений.
- III. Выполняется проверка условий (2) и (3) для полученных делением параллелепипедов и в зависимости от результата проверки проводится одно из следующих действий:
  - (1) если выполнено условие (2), то параллелепипед помещается в список полученных прямоугольников для последующего отображения;
  - (2) если выполнено условие (3), то параллелепипед далее не рассматривается;

- (3) если параллелепипед оказывается граничным, но его максимальное изменение превосходит величину точности аппроксимации, то он помещается в текущий список и выполняется переход к шагу 2. В противном случае параллелепипед добавляется к множеству граничных параллелепипедов.

Данный метод не позволяет заранее рассчитать число получаемых параллелепипедов для покрытия искомой области, но, как показывает эксперимент, оно будет много меньше, чем при методе равномерной сетки. Минус данного метода — его рекурсивность, из-за чего данный алгоритм сложен для распараллеливания.

### 3 Параллельный алгоритм

Реализация метода равномерной сетки для параллельных вычислений на графическом ускорителе позволяет сократить число выполняемых операций и время вычислений в десятки раз. Главное отличие параллельного алгоритма от последовательного состоит в том, что обработка параллелепипедов выполняется одновременно разными потоками графической карты.

Алгоритм равномерной сетки (рис. 1) для параллельной аппроксимации множества решений системы неравенств состоит в разбиении параллелепипеда, содержащего исследуемую область, на параллелепипеды меньшего размера с их последующей параллельной обработкой. Параллельная

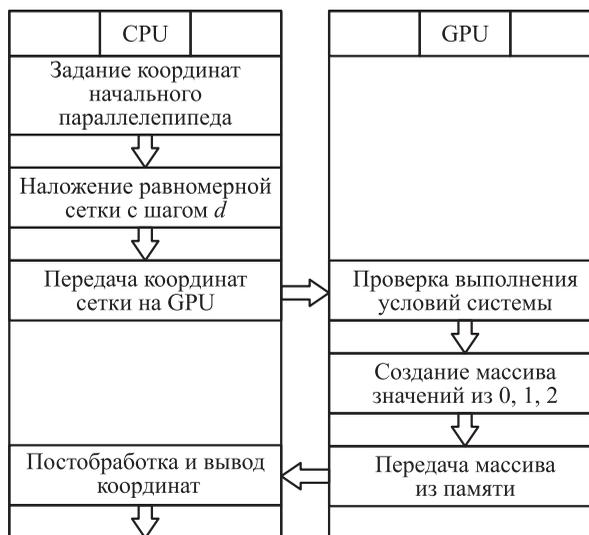


Рис. 1 Алгоритм равномерной сетки

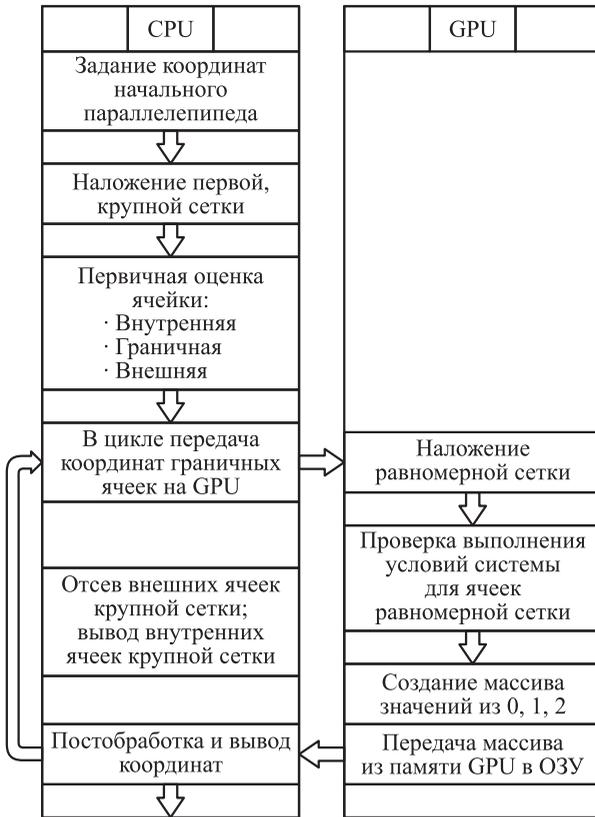


Рис. 2 Алгоритм двойной сетки

обработка выполняется на графическом ускорителе — один поток обрабатывает несколько ячеек сетки. При этом координаты ячеек не пересылаются на GPU, а генерируются исходя из номера потока.

Алгоритм равномерной сетки идеально подходит для реализации на графических ускорителях, так как обработка ячеек равномерной сетки может выполняться абсолютно независимо. Независимость потоков создает предпосылки для массивного параллелизма без синхронизации. Недостаток данного алгоритма — высокая ресурсоемкость: при увеличении точности аппроксимации количество обрабатываемых параллелепипедов существенно возрастает и превышает пределы графической памяти.

Преодолеть указанный недостаток позволяет алгоритм двойной сетки (рис. 2), сочетающий элементы метода ветвей и границ и перебора на равномерной сетке. Сначала на центральном процессоре (CPU, central processing unit) проводится предварительная обработка ячеек крупной равномерной сетки. В результате часть таких параллелепипедов классифицируется как внешние или внутренние, сохраняется и далее не рассматривается. Граничные параллелепипеды передаются графическому ускорителю по одному, где они обрабатываются так же,

как исходный параллелепипед в алгоритме равномерной сетки.

Результатом обработки равномерной сетки на GPU служит массив, содержащий значения 0, 1, 2 в зависимости от принадлежности параллелепипеда к искомому множеству аппроксимации. Далее в зависимости от индекса элемента массива для него вычисляются координаты параллелепипеда.

Для графического отображения полученного множества используется язык Python и библиотека Matplotlib. При большом числе отображаемых параллелепипедов сильно возрастает время вывода изображения и задействуется большой объем оперативной памяти. Для оптимизации процесса было принято решение объединять параллелепипеды, лежащие внутри искомого множества, в крупные. Таким образом удалось существенно уменьшить затраты времени и оперативной памяти.

#### 4 Реализация и экспериментальные исследования

Предложенные методы были реализованы с помощью языка C и технологии CUDA (Compute Unified Device Architecture). В качестве примера прикладного применения рассмотренных методов выбрана задача аппроксимации рабочей области (множества всех возможных положений рабочего инструмента) плоского параллельного робота. Роботы параллельной структуры характеризуются замкнутой кинематической цепью, сочетают высокую точность позиционирования и жесткость конструкции [3]. Для экспериментов выбран планарный робот с двумя степенями свободы (2-RPR). Схематично данное устройство представлено на рис. 3. Рабочий инструмент робота, закрепленный в точке X, приводится в движение двумя призматическими двигателями, которые изменяют длины штанг AX и BX.

Рабочая область данного робота определяется следующей системой неравенств:

$$\begin{aligned}
 f_1(x_1, x_2) &= x_1^2 + x_2^2 - (l_1^{\max})^2 \leq 0; \\
 f_2(x_1, x_2) &= (l_1^{\min})^2 - x_1^2 - x_2^2 \leq 0; \\
 f_3(x_1, x_2) &= (x_1 - l_0)^2 + x_2^2 - (l_2^{\max})^2 \leq 0; \\
 f_4(x_1, x_2) &= (l_2^{\min})^2 - (x_1 - l_0)^2 - x_2^2 \leq 0,
 \end{aligned}$$

где

$$-l_1^{\max} \leq x_1 \leq l_0 + l_2^{\max}; 0 \leq x_2 \leq \min(l_1^{\max}, l_2^{\max}).$$

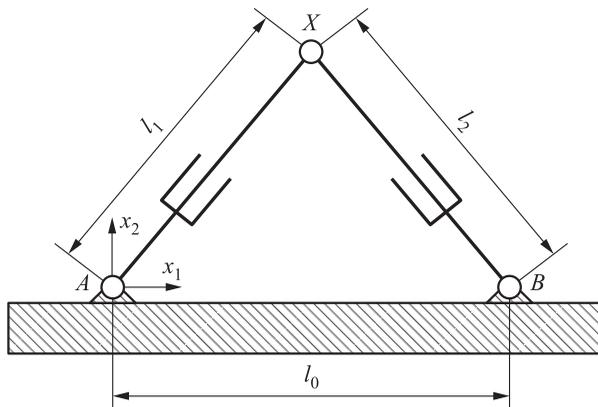


Рис. 3 Схема плоского робота 2-RPR

Значения  $l_{1,2}$  изменяются в диапазоне  $8 \leq l_{1,2} \leq 12$ , т. е.  $l_1^{\min} = 8$ ;  $l_1^{\max} = 12$ . Расстояние между точками закрепления штанг составляет  $l_0 = 5$ . Требуется аппроксимировать область, в которой, исходя из конструктивных особенностей длин звеньев, может находиться точка  $X(x_1, x_2)$  закрепления рабочего инструмента. Для получения гарантированных оценок минимума и максимума функций в левых частях неравенств применялись методы интервального анализа [4, 5].

Реализованные методы тестировались на персональном компьютере с Intel Core I7 с базовой тактовой частотой 2,8 ГГц, 16 ГБ ОЗУ, видеокартой Nvidia GeForce 1060: 3 ГБ оперативной памяти,

1280 ядер с базовой тактовой частотой от 1506 МГц. Тестирование проводилось для поставленной задачи со значениями точности аппроксимации, равными 0,1; 0,05; 0,01; 0,005; 0,001 и 0,0001.

Аппроксимации рабочей области для разных значений точности, полученные методом двойной сетки, приведены на рис. 4.

На рис. 4, б убраны линии, выделяющие получаемые параллелепипеды, поскольку при высокой точности аппроксимации толщина отображаемых линий становится соизмерима с высотой параллелепипедов, из-за чего линии сливаются. Таким образом, для точности  $d = 0,01$  высота параллелепипедов, которые не объединены друг с другом, слишком мала для графического отображения на целой фигуре.

Время работы алгоритмов в миллисекундах представлено в таблице.

Проведенные эксперименты позволяют сделать следующие выводы.

1. При низкой точности аппроксимации (0,1 или 0,05) параллельные варианты работают с невысокой эффективностью, что означает недостаточную загруженность вычислительных мощностей видеокарты. Применение GPU при таком объеме вычислений нецелесообразно.
2. При повышении точности возрастает объем вычислений и загруженность ядер графической карты, что приводит к увеличению эффективности расчетов. С помощью параллельного ал-

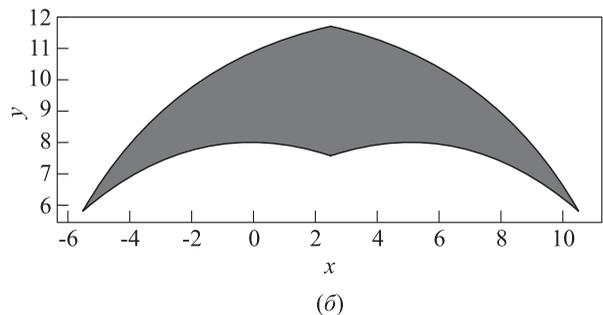
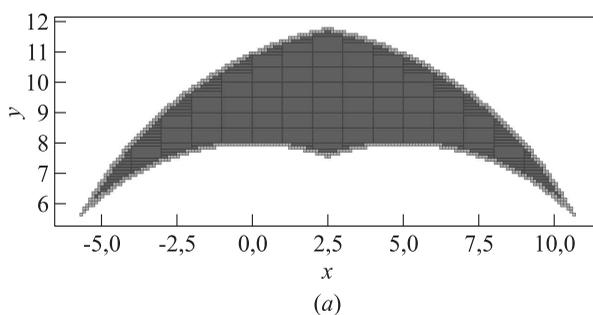


Рис. 4 Искомые множества: (а) точность  $d = 0,1$ ; (б) точность  $d = 0,01$

Время выполнения методов в зависимости от точности аппроксимации (мс)

Точность аппроксимации	Последовательный метод равномерной сетки	Метод ветвей и границ	Параллельный метод равномерной сетки	Параллельный метод двойной сетки
0,1	13	8	47	54
0,05	41	27	47	54
0,01	945	889	92	54
0,005	4 003	3 950	140	54
0,001	99 100	132 300	510	102
0,0001	>2,5 ч	>2,5 ч	—	4 400

горитма равномерной сетки не удается получить решения при высокой точности аппроксимации в силу превышения максимально допустимых значений числа обрабатываемых потоков или пределов локальной памяти.

3. Принцип дополнительного разбиения исходной области, заложенный в параллельном алгоритме двойной сетки, позволяет снизить объем вычислений и данных, обрабатываемых на графической карте, благодаря чему возможно получение более высокой точности аппроксимации.

## 5 Заключение

В данной работе разработаны два параллельных алгоритма для численной аппроксимации решения системы нелинейных неравенств, ориентированные на графические ускорители: алгоритм равномерной сетки и алгоритм двойной сетки. Результаты экспериментов показали, что параллельные методы с вычислениями на графическом ускорителе оказались существенно эффективнее последовательных однопоточных методов на центральном процессоре. Параллельный алгоритм двойной сетки оказался наиболее эффективным по скорости работы и наилучшим по точности аппроксимации. Этот метод сочетает в себе аппроксимацию на равномерной сетке, которая хорошо распараллеливается, и метод ветвей и границ, позволяющий сократить число операций за счет эффективного отсева «неперспективных» областей на начальном этапе.

В дальнейшем планируется рассмотреть роботов более сложной конструкции [6], что неизбежно

приведет к необходимости привлечения большей вычислительной мощности для расчетов. Предполагается использование серверов, оснащенных несколькими графическими ускорителями, установленных в центре обработки данных ФИЦ ИУ РАН [7].

## Литература

1. *Evtushenko Yu., Posypkin M., Rybak L., Turkin A.* Approximating a solution set of nonlinear inequalities // *J. Global Optim.*, 2018. Vol. 71. Iss. 1. P. 129–145.
2. *Cheng J., Grossman M., McKercher T.* Professional Cuda C programming. — New York, NY, USA: John Wiley & Sons, 2014. 499 p.
3. *Merlet J. P.* Parallel robots. — Solid mechanics and its applications ser. — Springer Science & Business Media, 2006. Vol. 128. 402 p.
4. *Moore R. E., Bierbaum F.* Methods and applications of interval analysis. — SIAM studies in applied and numerical mathematics ser. — Soc. for Industrial & Applied Math., 1979. 201 p.
5. *Hansen E., Walster G. W.* Global optimization using interval analysis: Revised and expanded. — Pure and applied mathematics ser. — CRC Press, 2003. Book 264. 728 p.
6. *Malyshev D., Rybak L., Behera L., Mohan S.* Workspace modelling of a parallel robot with relative manipulation mechanisms based on optimization methods // *Robotics and mechatronics* / Eds. C. H. Kuo, P. C. Lin, T. Essomba, G. C. Chen. — Cham: Springer, 2019. Vol. 78. P. 151–163.
7. *Zatsarinny A. A., Gorshenin A. K., Kondrashev V. A., Volovich K. I., Denisov S. A.* Toward high performance solutions as services of research digital platform // *Procedia Comput. Sci.*, 2019. Vol. 150. P. 622–627.

Поступила в редакцию 08.10.19

---

---

# APPROXIMATION OF THE SET OF SOLUTIONS OF SYSTEMS OF NONLINEAR INEQUALITIES USING GRAPHIC ACCELERATORS

M. V. Popov and M. A. Posypkin

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

**Abstract:** Solutions of certain problems can be reduced to the solution of some systems of inequalities. But the computation of the set of exact solutions may not be feasible. Thus, various methods for approximation of the solution set have been developed. The more accurate approximation is required, the bigger number of calculations must be performed and, consequently, the runtime of the algorithms increases. Nowadays, it is common to speed up algorithms by paralleling computations on graphics accelerators. The paper describes the serial method for approximation of the solution of systems of inequalities and proposes the parallel hybrid algorithm that combines iterations on the uniform grid and the branch and bound method. This algorithm is suited for direct implementation on graphics accelerators and does not suffer from the excessive enumeration of possible solution candidates. The

sequential algorithm and the two versions of the parallel algorithm are compared through one example: the problem of approximation of the working area of the robot which consists of the set of robot's tool positions and is the key robot's characteristic.

**Keywords:** optimization; parallel computing; graphics accelerator, GPU; CUDA; nonlinear inequalities

**DOI:** 10.14357/19922264200303

## Acknowledgments

The paper was partially supported by the Russian Science Foundation (project 16-19-00148).

## References

1. Evtushenko, Yu. G., M. A. Posypkin, L. A. Rybak, and A. V. Turkin. 2018. Approximating a solution set of nonlinear inequalities. *J. Global Optim.* 71(1):129–145.
2. Cheng, J., M. Grossman, and T. McKercher. 2014. *Professional Cuda C programming*. New York, NY: John Wiley & Sons. 499 p.
3. Merlet, J. P. 2006. *Parallel robots*. Solid mechanics and its applications ser. Springer Science & Business Media. Vol. 128. 402 p.
4. Moore, R. E., and F. Bierbaum. 1979. *Methods and applications of interval analysis*. SIAM studies in applied and numerical mathematics ser. Soc. for Industrial & Applied Math. 201 p.
5. Hansen, E., and G. W. Walster. 2003. *Global optimization using interval analysis: Revised and expanded*. Pure and applied mathematics ser. CRC Press. Book 264. 728 p.
6. Malyshev, D., L. A. Rybak, L. Behera, and S. Mohan. 2019. Workspace modelling of a parallel robot with relative manipulation mechanisms based on optimization methods. *Robotics and mechatronics*. Eds. C. H. Kuo, P. C. Lin, T. Essomba, and G. C. Chen. Cham: Springer. 78:151–163.
7. Zatsarinny, A. A., A. K. Gorshenin, V. A. Kondrashev, K. I. Volovich, and S. A. Denisov. 2019. Toward high performance solutions as services of research digital platform. *Procedia Comput. Sci.* 150:622–627.

Received October 8, 2019

## Contributors

**Popov Mikhail V.** (b. 1995) — PhD student, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; alvopim@gmail.com

**Posypkin Mikhail A.** (b. 1974) — Doctor of Science in physics and mathematics, associate professor, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; mposypkin@gmail.com

# ОДНОЛИНЕЙНАЯ СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С ИНВЕРСИОННЫМ ПОРЯДКОМ ОБСЛУЖИВАНИЯ С ВЕРОЯТНОСТНЫМ ПРИОРИТЕТОМ, ГРУППОВЫМ ПУАССОНОВСКИМ ПОТОКОМ И ФОНОВЫМИ ЗАЯВКАМИ\*

Т. А. Милованова<sup>1</sup>, Р. В. Разумчик<sup>2</sup>

**Аннотация:** Рассматривается система массового обслуживания (СМО) с одним прибором, групповым пуассоновским потоком основных заявок и насыщенным потоком фоновых заявок, запас которых не ограничен. Основные заявки имеют относительный приоритет перед заявками фонового потока. Таким образом, прерывание обслуживания заявки второго типа не допускается и процесс обслуживания фоновых заявок начинается только тогда, когда после выхода с прибора основной заявки очередь основных заявок пуста. Времена обслуживания основных и фоновых заявок независимы и имеют произвольные распределения. В системе реализована дисциплина инверсионного порядка обслуживания с вероятностным приоритетом. Предложен метод расчета (в терминах преобразований) основных стационарных показателей функционирования системы, включая стационарные распределения времени ожидания начала обслуживания и времени пребывания заявок основного потока в системе.

**Ключевые слова:** инверсионный порядок обслуживания; вероятностный приоритет; групповой поток; фоновые заявки

**DOI:** 10.14357/19922264200304

## 1 Введение

Статья посвящена развитию результатов работ [1–4] по исследованию стационарных характеристик СМО с так называемыми фоновыми заявками, находящих применение при разработке энерго-сберегающих схем в сенсорных беспроводных сетях (см., например, [5]). Рассматривается СМО с одним прибором, инверсионным порядком обслуживания с вероятностным приоритетом и двумя входящими потоками заявок разных классов: групповым пуассоновским потоком основных заявок и ординарным потоком фоновых заявок. Источником заявок фонового потока служит так называемый бункер, в котором запас заявок не ограничен: если нет очереди основных заявок, то начинается процесс обслуживания фоновых заявок<sup>3</sup>. Заявки основного потока имеют относительный приоритет перед заявками фонового потока. С помощью специального метода, введенного в [6], здесь решена задача отыскания (в терминах преобразований) стационарных вероятностей состояний, а также стационарных распределений времени ожидания начала обслуживания и времени пребывания заявок основ-

ного потока в системе. Заметим, что рассмотренная система может быть трактована как система с прогулками (отключением прибора) при опустошении системы от заявок основного потока (см., например, [7, 8]). Таким образом, полученные здесь результаты обобщают некоторые из известных для этого типа систем результаты и, в частности, позволяют находить стационарное распределение вероятностей состояний при прямом порядке обслуживания (т. е. при дисциплине FIFO — first in, first out).

## 2 Описание системы

Далее основные (фоновые) заявки будем называть заявками первого (второго) типа. Рассмотрим СМО с одним прибором и инверсионным порядком обслуживания с вероятностным приоритетом, в которую поступают заявки двух типов. Заявки первого типа поступают в систему группами в соответствии с пуассоновским потоком с параметром  $\lambda$ . Вероятность поступления  $k$ ,  $k \geq 1$ , заявок в группе обозначим через  $c_k$ , а средний размер группы —

\* Исследование выполнено при финансовой поддержке РФФИ (проект № 20-07-00804).

<sup>1</sup> Российский университет дружбы народов, milovanova-ta@rudn.ru

<sup>2</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, rguzumchik@ipiran.ru

<sup>3</sup> Такой поток заявок иногда называют насыщенным.

через  $\bar{c} = \sum_{k=1}^{\infty} k c_k$ . Заявки первого типа обслуживаются по одной, и их времена обслуживания (длины) являются независимыми случайными величинами с функцией распределения  $B(x)$  и средним значением  $\bar{b} = \int_0^{\infty} (1 - B(x)) dx < \infty$ . Заявки второго типа поступают из накопителя бесконечной емкости, и их длины тоже независимы с функцией распределения  $G(x)$  и средним значением  $\bar{g} = \int_0^{\infty} (1 - G(x)) dx < \infty$ .

Для простоты изложения будем полагать<sup>1</sup>, что существуют плотности  $b(x) = B'(x)$  и  $g(x) = G'(x)$ . Заявки первого типа имеют относительный приоритет перед заявками второго типа, т.е. поступление на прибор заявки второго типа происходит только в том случае, если в системе отсутствуют заявки первого типа (прерывание обслуживания заявок второго типа не допускается). Общее число заявок первого типа в системе ограничено числом  $N$ ,  $N \leq \infty$ . При  $N < \infty$  считается, что поступающая группа теряется целиком, если в момент поступления для хотя бы одной из заявок в группе не хватает места в очереди.

Принятый в системе инверсионный порядок обслуживания с вероятностным приоритетом заключается в следующем:

- (1) если в момент поступления группы заявок первого типа в системе обслуживается заявка первого типа, то длина  $x$  первой из поступающей группы заявки сравнивается с остаточной длиной  $y$  заявки, находящейся на приборе. С вероятностью  $v(y, x)$  поступающая группа занимает первые места в очереди, а заявки, находившиеся в очереди до поступления группы, становятся за ними с учетом порядка. С дополнительной вероятностью  $\bar{v}(y, x) = 1 - v(y, x)$  первая заявка из поступающей группы становится на прибор, остальные заявки из поступающей группы занимают первые места в очереди, заявка с прибора встает за ними. Остальные заявки, находившиеся в очереди до поступления новой группы, становятся после этой заявки с сохранением порядка;
- (2) если группа заявок первого типа в момент поступления застает на приборе заявку второго типа, то длина  $x$  первой из поступающей группы заявки сравнивается с длиной  $y$  заявки, стоящей на первом месте в очереди. С вероятностью  $w(y, x)$  заявка длины  $y$  остается на первом месте в очереди, поступающая группа занимает места в очереди начиная со второ-

го, а остальные заявки, имевшиеся в очереди до поступления новой группы, становятся за ними с сохранением порядка. С дополнительной вероятностью  $\bar{w}(y, x) = 1 - w(y, x)$  поступающая группа заявок занимает первые места в очереди, заявка длины  $y$  становится за поступившей группой заявок, а заявки, находившиеся в очереди до момента поступления группы, становятся за ней с учетом порядка.

### 3 Стационарное распределение числа заявок в системе

Рассмотрим  $\nu(t)$  — общее число заявок в системе в момент  $t$  и  $\vec{\xi}(t) = (\xi_1(t), \dots, \xi_{\nu(t)}(t))$  — вектор, координатой  $\xi_1(t)$  которого является остаточное время обслуживания заявки, находящейся в этот момент на приборе,  $\xi_2(t)$  — первой заявки в очереди,  $\dots$ ,  $\xi_{\nu(t)}(t)$  — последней заявки в очереди. Тогда если обозначить через  $\chi(t)$  — величину, равную в момент  $t$  единице, когда прибор занят обслуживанием заявки первого типа, и нулю в противном случае, то процесс  $(\chi(t), \nu(t), \vec{\xi}(t))$  будет марковским. Введем необходимые для дальнейшего изложения функции:

$q_k(t, x)$ ,  $0 \leq k \leq N-1$ , — стационарная плотность<sup>2</sup> вероятности того, что на приборе обслуживается заявка второго типа длины  $t$  и в очереди находятся  $k$  заявок первого типа, причем заявка, стоящая в очереди первой, имеет длину  $x$ ;

$q_N^*(t, x, y)$  — стационарная плотность вероятности того, что на приборе обслуживается заявка второго типа длины  $t$  и в очереди находятся  $n$  заявок первого типа, причем длины первых двух заявок в очереди  $x$  и  $y$ ;

$p_k(x, y)$ ,  $1 \leq k \leq N-1$ , — стационарная плотность вероятности того, что на приборе обслуживается заявка первого типа длины  $x$  и в очереди находятся  $k-1$  заявок первого типа, причем заявка, стоящая в очереди первой, имеет длину  $y$ ;

$p_N^*(x, y)$  — стационарная плотность вероятности того, что на приборе обслуживается заявка первого типа длины  $x$  и в очереди находятся  $N-1$  заявок первого типа, причем заявка, стоящая в очереди первой, имеет длину  $y$ .

Обозначим через

$$P_k = \int_0^{\infty} \int_0^{\infty} p_k(x, y) dy dx, \quad 1 \leq k \leq N-1;$$

<sup>1</sup>Как и в системе с ординарным потоком заявок первого типа (см. [2]), случай произвольных распределений длин заявок можно трактовать в терминах обобщенных функций.

<sup>2</sup>Определяется как  $q_k(t, x) = \partial^2 Q_k(t, x) / (\partial t \partial x)$ , где  $Q_k(t, x) = \lim_{T \rightarrow \infty} \mathbf{P}\{\chi(T) = 0, \nu(T) = k+1, \xi_1(T) < t, \xi_2(T) < x\}$ . При  $k = 0$  аргумент  $x$  опускается. Можно показать, что эта и другие введенные плотности существуют.

$$P_N^* = \int_0^\infty \int_0^\infty p_N^*(x, y) dy dx$$

стационарные вероятности того, что на приборе обслуживается заявка первого типа и в очереди находятся  $k - 1$  заявок первого типа. Соответственно,

$$Q_k = \int_0^\infty \int_0^\infty q_k(t, x) dx dt, \quad 0 \leq k \leq N - 1;$$

$$Q_N^* = \int_0^\infty \int_0^\infty \int_0^\infty q_N^*(t, x, y) dy dx dt$$

суть стационарные вероятности того, что в системе находятся  $k$  заявок первого типа и обслуживается заявка второго типа.

Для нахождения стационарного распределения числа заявок в системе воспользуемся методом производящих функций (ПФ). Для этого, не связывая систему с конкретным значением  $N$ , рассмотрим набор систем с различными  $N > 1$  и введем ПФ<sup>1</sup>

$$P(s, x) = \sum_{k=1}^\infty p_k(x) s^k; \tag{1}$$

$$Q(s, t, x) = \sum_{k=1}^\infty q_k(t, x) s^k; \tag{2}$$

$$Q^*(s, t, z, x) = \sum_{k=1}^\infty q_{k+1}^*(t, z, x) s^k, \tag{3}$$

где

$$p_k(x) = \int_0^\infty p_k(x, y) dy.$$

Начнем с нахождения ПФ (2) и (3). Система уравнений Колмогорова–Чепмена для стационарных плотностей  $q_k(t, x)$  и  $q_N^*(t, x, y)$  имеет вид<sup>2</sup>:

$$-q_0'(t) = -\lambda q_0(t) + g(t)Q; \tag{4}$$

$$-\frac{dq_1(t, x)}{dt} = -\lambda q_1(t, x) + \lambda c_1 b(x) q_0(t); \tag{5}$$

$$-\frac{dq_k(t, x)}{dt} = -\lambda q_k(t, x) + \lambda c_k b(x) q_0(t) + \sum_{m=1}^{k-1} \lambda c_{k-m} \times \int_0^\infty (b(y)w(x, y)q_m(t, x) + b(x)\bar{w}(y, x)q_m(t, y)) dy, \tag{6}$$

$$2 \leq k \leq N - 1;$$

$$-\frac{dq_N^*(t, z, x)}{dt} = \lambda c_N b(z) b(x) q_0(t) + \lambda c_1 (b(x)w(z, x)q_{N-1}(t, z) + b(z)\bar{w}(x, z)q_{N-1}(t, x)) + \sum_{m=1}^{N-2} \lambda c_{N-m} \int_0^\infty (b(y)w(z, y)q_m(t, z)\delta(y - x) + b(z)b(x)\bar{w}(y, z)q_m(t, y)) dy, \tag{7}$$

где  $Q = q_0(0) + p_1(0)$  — постоянная, которая, как будет видно из дальнейшего, имеет смысл нормирующего множителя, а  $\delta$  здесь и далее — дельта-функция Дирака. Решая уравнения (4)–(6), получаем, что при  $1 \leq k \leq N - 1$  плотности  $q_k(t, x)$  представимы в виде

$$q_k(t, x) = \sum_{i=1}^k a_i^{(k)} A_i(x) q_i(t),$$

причем сомножители определяются рекуррентным образом:

$$q_0(t) = Q e^{\lambda t} \int_t^\infty g(y) e^{-\lambda y} dy, \quad q_k(t) = \lambda e^{\lambda t} \int_t^\infty q_{k-1}(y) e^{-\lambda y} dy; \quad 1 \leq k \leq N - 1;$$

$$a_1^{(k)} = c_k, \quad a_i^{(k)} = \sum_{m=i-1}^{k-1} a_{i-1}^{(m)} c_{k-m}, \quad 2 \leq i \leq k;$$

$$A_1(x) = b(x), \quad A_i(x) = \int_0^\infty (b(y)w(x, y)A_{i-1}(x) + b(x)\bar{w}(y, x)A_{i-1}(y)) dy.$$

Отсюда путем стандартных преобразований находим, что ПФ  $Q(s, t, x)$  имеет вид<sup>3</sup>:

$$Q(s, t, x) = \sum_{i=1}^\infty A_i(x) q_i(t) (C(s))^i,$$

где  $C(s)$  — ПФ размера группы. Интегрируя выражение для  $Q(s, t, x)$  по всем возможным значениям  $t$  и  $x$  и учитывая, что  $\int_0^\infty A_i(x) dx = 1$ , получаем ПФ стационарных вероятностей  $\{Q_k, k \geq 0\}$ :

$$Q(s) + Q_0 = Q \frac{1 - \gamma(\lambda - \lambda C(s))}{\lambda - \lambda C(s)}, \tag{8}$$

где  $\gamma(s)$  — преобразование Лапласа плотности  $g$  в точке  $s$ . Далее, как видно из (7), после умножения

<sup>1</sup> Введенная ПФ  $P(s, x)$  имеет смысл обычной ПФ только при загрузке системы заявками первого типа  $\lambda \bar{c} < 1$ .

<sup>2</sup> Для сокращения записи здесь и далее используются соглашения, что  $\sum_{j=i}^{i-1} \equiv 0$  и  $\prod_{j=i}^{i-1} \equiv 1$ .

<sup>3</sup> Заметим, что в некоторых случаях для определения  $Q(s, t, x)$  может быть удобно воспользоваться соответствующим интегральным уравнением, которое следует из (4)–(6).

левой и правой частей на  $s^N$  и суммирования по  $N > 1$ , ПФ  $Q^*(s, t, z, x)$  задается формулой:

$$\begin{aligned}
 Q^*(s, t, z, x) = & \\
 = Q\lambda \left( \frac{C(s)}{s} - c_1 \right) & b(z)b(x) \int_t^\infty e^{\lambda u} du \int_u^\infty g(y)e^{-\lambda y} dy + \\
 & + \lambda c_1 \int_t^\infty (b(x)w(z, x)Q(s, u, z) + \\
 & + b(z)\bar{w}(x, z)Q(s, u, x)) du + \lambda \left( \frac{C(s)}{s} - c_1 \right) \times \\
 & \times \int_t^\infty du \int_0^\infty (b(y)w(z, y)Q(s, u, z)\delta(y - x) + \\
 & + b(z)b(x)\bar{w}(y, z)Q(s, u, y)) dy. \quad (9)
 \end{aligned}$$

Интегрируя теперь (9) по всем  $t, z$  и  $x$ , с учетом вида ПФ  $Q(s, t, x)$  в (8), получаем ПФ стационарных вероятностей  $\{Q_N^*, N \geq 2\}$ :

$$\begin{aligned}
 Q^*(s) = \frac{Q}{\lambda} \left( \frac{C(s)}{s} \left( \frac{\lambda \bar{g}}{1 - C(s)} - \right. \right. \\
 \left. \left. - \frac{1 - \gamma(\lambda - \lambda C(s))}{(1 - C(s))^2} \right) - c_1 (\lambda \bar{g} - 1 + \gamma(\lambda)) \right). \quad (10)
 \end{aligned}$$

Перейдем к нахождению ПФ (1). Система уравнений Колмогорова–Чепмена для стационарных плотностей  $p_k(x)$  и  $p_N^*(x, y)$  имеет вид:

$$-\frac{dp_1(x)}{dx} = -\lambda p_1(x) + q_1(0, x) + p_2(0, x); \quad (11)$$

$$\begin{aligned}
 -\frac{dp_k(x)}{dx} = & -\lambda p_k(x) + q_k(0, x) + p_{k+1}(0, x) + \\
 & + \sum_{m=1}^{k-1} \lambda c_{k-m} \int_0^\infty (b(u)v(x, u)p_m(x) + \\
 & + b(x)\bar{v}(u, x)p_m(u)) du, \quad 2 \leq k \leq N - 1; \quad (12)
 \end{aligned}$$

$$\begin{aligned}
 -\frac{dp_N^*(x, y)}{dx} = & q_N^*(0, x, y) + \\
 & + \lambda c_1 (b(y)v(x, y)p_{N-1}(x) + b(x)\bar{v}(y, x)p_{N-1}(y)) + \\
 & + \sum_{m=1}^{N-2} \lambda c_{N-m} \int_0^\infty (b(u)v(x, u)p_m(x)\delta(u - y) + \\
 & + b(x)b(y)\bar{v}(u, x)p_m(u)) du. \quad (13)
 \end{aligned}$$

Из метода исключения состояний (см., например, [6, 9–11] и [2, с. 75]) следует, что плотности  $p_k(x)$  и  $p_k(x, y)$  совпадают с точностью до постоянного множителя, в качестве которого, с учетом

найденных выше ПФ (8) и (10), можно выбрать  $Q$ . Поэтому в (12) вместо  $p_{k+1}(0, x)$  можно подставить  $p_{k+1}^*(0, x)$ , предварительно проинтегрировав (13). После этого, умножая левые и правые части (11) и (12) на  $s^k$  и суммируя по  $k \geq 1$ , приходим к уравнению для расчета ПФ  $P(s, x)$ :

$$\begin{aligned}
 -\frac{dP(s, x)}{dx} = & -\lambda(1 - c_1)P(s, x) + \lambda c_1 b(x)P(s) + \\
 & + Q(s, 0, x) + \int_0^\infty Q^*(s, 0, u, x)du + \\
 & + \lambda(C(s) - c_1) \int_0^\infty (b(u)v(x, u)P(s, x) + \\
 & + b(x)\bar{v}(u, x)P(s, u)) du + \lambda \left( \frac{C(s)}{s} - c_1 \right) \times \\
 & \times \int_0^\infty dz \int_0^\infty (b(u)v(z, u)P(s, z)\delta(u - x) + \\
 & + b(z)b(x)\bar{v}(u, z)P(s, u)) du.
 \end{aligned}$$

Наконец, интегрируя (13) и вводя ПФ

$$P^*(s, x, y) = \sum_{N=2}^\infty p_N^*(x, y)s^N,$$

находим:

$$\begin{aligned}
 P^*(s, x, y) = & s \int_x^\infty Q^*(s, 0, u, y) du + \\
 & + \lambda c_1 s \int_x^\infty (b(y)v(u, y)P(s, u) + b(u)\bar{v}(y, u)P(s, y)) du + \\
 & + \lambda(C(s) - c_1 s) \int_x^\infty dz \int_0^\infty (b(u)v(z, u)P(s, z)\delta(u - y) + \\
 & + b(z)b(y)\bar{v}(u, z)P(s, u)) du.
 \end{aligned}$$

Таким образом, в рассматриваемой системе стационарные вероятности  $Q_k, 0 \leq k \leq N - 1$ , и  $P_k, 1 \leq k \leq N - 1$ , определяются как коэффициенты при  $s^k$  разложения в ряд по степеням  $s$  соответственно функций  $Q(s) + Q_0$  и  $P(s) = \int_0^\infty P(s, x) dx$ . Стационарные вероятности  $Q_N^*$  и  $P_N^*$  определяются как коэффициенты при  $s^N$  разложения в ряд по степеням  $s$  соответственно функций  $sQ^*(s)$  и  $P^*(s) = \int_0^\infty \int_0^\infty P^*(s, x, y) dy dx$ . Постоянная  $Q$  определяется из условия нормировки

$$Q_0 + \sum_{k=1}^{N-1} (Q_k + P_k) + Q_N^* + P_N^* = 1$$

или, в случае  $N = \infty$ , из соотношения

$$Q_0 + P(1) + Q(1) = 1.$$

Зная стационарное распределение, в силу пуассоновости входящего потока можно вычислить и стационарную вероятность  $\pi$  потери заявки:

$$\pi = P_N^* + Q_N^* + \sum_{k=2}^N \frac{k c_k}{\bar{c}} \sum_{j=N-k+1}^{N-1} (P_j + Q_j).$$

Моменты стационарного распределения числа заявок в системе вычисляются путем дифференцирования соответствующих ПФ и последующего решения получившихся уравнений. Заметим, что при  $N < \infty$  загрузка системы заявками первого типа, равная  $(\sum_{k=1}^{N-1} P_k + P_N^*) / (1 - \pi)$ , также служит одним из показателей ее производительности.

Ситуация с решением интегральных уравнений аналогична ситуации в случае системы с ординарным потоком: в некоторых частных случаях<sup>1</sup> решения могут быть выписаны в явном виде. Необходимо также отметить, что некоторые эффекты, которые наблюдаются в системе с ординарным потоком, исчезают в системе с групповым потоком. В частности (при фиксированной загрузке заявками первого типа), стационарное распределение числа заявок в рассматриваемой системе при инверсионном порядке обслуживания с прерыванием (т. е.  $w(y, x) = v(y, x) = 0$ ) уже не является инвариантным относительно вида распределения  $B(x)$ .

#### 4 Стационарное распределение времени ожидания и времени пребывания заявки в системе

Для нахождения временных характеристик поступающих в систему заявок первого типа необходимо предварительно определить несколько вспомогательных величин. Обозначим через  $u_m(s; x)$ ,  $1 \leq m \leq N$ , преобразование Лапласа–Стилтьеса (ПЛС) функции распределения времени до того момента, когда в системе останется  $(m - 1)$  заявок при условии, что на приборе начала обслуживаться заявка первого типа длины  $x$  и в системе было  $m$  заявок. Поскольку в соответствии с описанием системы группа заявок, застающая при поступлении в системе  $n$  заявок первого типа, теряется, то  $u_N(s; x) = e^{-sx}$ . Для  $1 \leq m \leq N - 1$ , рассматривая все возможные события и воспользовавшись свойствами ПЛС, получаем:

$$\begin{aligned} u_m(s; x) = & e^{-(\lambda+s)x} + \sum_{k=1}^{N-m} c_k \int_0^x \lambda e^{-(\lambda+s)t} dt \times \\ & \times \int_0^\infty \bar{v}(x-t, y) u_m(s; x-t) u_{m+k}(s; y) \times \\ & \times \prod_{j=2}^k u_{m+k+1-j}(s) b(y) dy + \sum_{k=1}^{N-m} c_k \int_0^x \lambda e^{-(\lambda+s)t} dt \times \\ & \times \int_0^\infty v(x-t, y) u_{m+k}(s; x-t) u_{m+k-1}(s; y) \times \\ & \times \prod_{j=2}^k u_{m+k-j}(s) b(y) dy. \end{aligned}$$

Здесь и далее

$$u_m(s) = \int_0^\infty u_m(s; x) b(x) dx.$$

Решения этих интегральных уравнений в явном виде при  $N < \infty$  выписать не удастся<sup>2</sup>. Но при любом фиксированном  $s$  они могут быть найдены численно, причем расчет необходимо вести в следующем порядке:  $u_{N-1}(s; x)$ ,  $u_{N-1}(s)$ ,  $u_{N-2}(s; x)$ ,  $u_{N-2}(s) \dots$

Пусть теперь в системе находится всего  $m_1$  заявок первого типа, первая заявка в очереди имеет длину  $x_1$ , а на приборе обслуживается заявка второго типа. Выделим заявку в очереди: пусть она находится на  $i_1$ -м месте. Обозначим через  $h_{(m_1, i_1), (m_2, i_2)}(x_1, x_2)$  (условную) плотность вероятности того, что после очередного поступления в систему окажутся  $m_2$ ,  $\min(m_1 + 1, N) \leq m_2 \leq N$ , заявок, выделенная заявка окажется на  $i_2$ -м,  $i_1 \leq i_2 \leq m_2$ , месте, а длина первой заявки в очереди будет  $x_2$ . Доопределим функцию  $h$  в точках  $m_2 = m_1$  следующим образом (здесь и далее  $\mathbf{1}_{(A)}$  — индикатор события  $A$ ):

$$\begin{aligned} h_{(m_1, i_1), (m_1, i_2)}(x_1, x_2) = & \\ = & \begin{cases} 0, & m_1 \neq N; \\ \delta(x_1 - x_2) \mathbf{1}_{(i_1 = i_2)}, & m_1 = N. \end{cases} \end{aligned} \quad (14)$$

Учитывая, что поступающая группа заявок может либо занять первые места в очереди, либо встать в очередь сразу позади первой заявки, по формуле полной вероятности получаем:

<sup>1</sup> Например, когда длины заявок первого типа принимают только конечное число значений, или когда для функций  $v(x, y)$  известна сепарабельная аппроксимация (см. [10, 12–14]). Подробнее см. [2, с. 78].

<sup>2</sup> Однако при  $N = \infty$  функции  $u_m(s; x)$  и  $u_m(s)$  не зависят от  $m$  и, как показано в [15, с. 14], могут быть найдены путем решения одного-единственного уравнения. Аналогично обстоит дело и с ПЛС периода занятости системы заявками первого типа.

$$\begin{aligned}
 & h_{(m_1, i_1), (m_2, i_2)}(x_1, x_2) = \\
 & = c_{m_2 - m_1} \int_0^\infty (\bar{w}(x_1, y) \delta(y - x_2) \mathbf{1}_{(i_2 = i_1 + m_2 - m_1)} + \\
 & + w(x_1, y) \delta(x_1 - x_2) \mathbf{1}_{(i_2 = 1)} \mathbf{1}_{(i_1 = 1)} + w(x_1, y) \times \\
 & \times \delta(x_1 - x_2) \mathbf{1}_{(i_2 = i_1 + m_2 - m_1)} \mathbf{1}_{(i_1 \neq 1)}) b(y) dy. \quad (15)
 \end{aligned}$$

На основе (14) и (15) рассчитываются плотности  $h_{(m_1, i_1), (m_2, i_2)}^{(n)}(x_1, x_2)$ ,  $n \geq 2$ , вероятностей переходов за  $n$  последовательных поступлений:

$$\begin{aligned}
 & h_{(m_1, i_1), (m_2, i_2)}^{(n)}(x_1, x_2) = \\
 & = \begin{cases} \sum_{m=m_1+1}^{m_2} \sum_{i=i_1}^{i_2} \int_0^\infty h_{(m_1, i_1), (m, i)}^{(n-1)}(x_1, y) \times \\ \times h_{(m, i), (m_2, i_2)}(y, x_2) dy, & m_1 \neq N; \\ h_{(N, i_1), (m_2, i_2)}(x_1, x_2), & m_1 = N. \end{cases}
 \end{aligned}$$

Нахождение распределения времени ожидания начала обслуживания удобно разбить на два этапа: один соответствует случаю, когда поступающая группа застала на приборе заявку первого типа, другой — случаю, когда на приборе оказалась заявка второго типа. Определим ПЛС  $\omega_{k1}^{(1)}(s; x)$ ,  $1 \leq k \leq N - 1$ , функции распределения времени ожидания начала обслуживания принятой заявки, заставшей на приборе заявку первого типа, при условии, что она поступила в группе размера  $k$ , была на первом месте в группе и имела длину  $x$ . Ее время ожидания равно нулю, если она заняла место заявки на приборе. Если же она застала в системе  $m$ ,  $1 \leq m \leq N - k$ , заявок первого типа, на приборе — заявку длины  $y$  и не заняла ее место, то время ожидания совпадает с периодом времени, необходимым для уменьшения длины очереди на единицу. В терминах ПЛС имеем:

$$\begin{aligned}
 \omega_{k1}^{(1)}(s; x) = \sum_{m=1}^{N-k} \int_0^\infty p_m(y) (\bar{w}(y, x) + \\ + w(y, x) u_{m+k}(s; y)) dy. \quad (16)
 \end{aligned}$$

Предположим теперь, что выделенная заявка оказалась на  $i$ -м месте в группе. Тогда ее время ожидания зависит от длин заявок, стоящих перед ней, а также от того, произошла ли смена заявки на приборе. Поскольку длины заявок в группе независимы и не зависят от размера группы, то ПЛС  $\omega_{ki}^{(1)}(s; x)$ ,  $1 \leq k \leq N - 1$ ,  $2 \leq i \leq k$ , стационарного (условного) распределения ее времени ожидания начала обслуживания равно:

$$\begin{aligned}
 \omega_{ki}^{(1)}(s; x) = \\
 = \sum_{m=1}^{N-k} \int_0^\infty p_m(y) \int_0^\infty \left( \bar{w}(y, z) u_{m+k}(s; z) \prod_{j=1}^{i-2} u_{m+k-j}(s) + \right. \\
 \left. + w(y, z) u_{m+k}(s; y) u_{m+k-1}(s; z) \prod_{j=2}^{i-1} u_{m+k-j}(s) \right) \times \\
 \times b(z) dz dy. \quad (17)
 \end{aligned}$$

Усредняя  $\omega_{ki}^{(1)}(s; x)$ , полученные в (16) и (17), по распределению длины выделенной заявки и ее месту в поступающей группе, получаем безусловное ПЛС  $\omega^{(1)}(s)$  функции распределения времени ожидания начала обслуживания принятой заявки, заставшей при поступлении на приборе заявку первого типа:

$$\omega^{(1)}(s) = \sum_{k=1}^{N-1} \sum_{i=1}^k \frac{c_k}{\bar{c}} \int_0^\infty \omega_{ki}^{(1)}(s; x) b(x) dx. \quad (18)$$

Перейдем ко второму этапу. Обозначим через  $\theta_{(k,i)(l,j)}(y, z, x)$ ,  $k \leq l \leq N$ ,  $1 \leq i \leq k$ , вероятность того, что выделенная заявка займет  $j$ -е место в очереди, первая заявка в очереди будет иметь длину  $z$ , всего в системе будет  $l$  заявок первого типа и на приборе — заявка второго типа длины  $y$ , при условии, что выделенная заявка поступила в группе размера  $k$ , была в ней на  $i$ -м месте и имела длину  $x$ . Выражения для ненулевых вероятностей  $\theta_{(k,i)(l,j)}(y, z, x)$  находятся по формуле полной вероятности:

$$\begin{aligned}
 \theta_{(k,i)(k,i)}(y, z, x) & = \\
 & = q_0(y) \delta(x - z) \mathbf{1}_{(i=1)} + q_0(y) b(z) \mathbf{1}_{(i \neq 1)}; \\
 \theta_{(k,i)(l,i)}(y, z, x) & = \\
 & = q_{l-k}(y, z) \bar{w}(z, x) \delta(x - z) \mathbf{1}_{(i=1)} + \\
 & + b(z) \int_0^\infty q_{l-k}(y, t) \bar{w}(t, z) dt \mathbf{1}_{(i \neq 1)}; \\
 \theta_{(k,i)(l,i+1)}(y, z, x) & = q_{l-k}(y, z) w(z, x) \mathbf{1}_{(i=1)} + \\
 & + q_{l-k}(y, z) \int_0^\infty b(t) w(z, t) dt \mathbf{1}_{(i \neq 1)}.
 \end{aligned}$$

Оказавшись в очереди, выделенная заявка до поступления на прибор должна ожидать окончания обслуживания заявки на приборе, впереди стоящих заявок (и их потомков), а также окончания обслуживания тех заявок (и их потомков), которые могут поступить в систему (за время обслуживания заявки второго типа на приборе) и встать перед ней. Нетрудно видеть, что вероятность  $\theta_{(k,i)(l,j)}(y, z, x)$ ,

$k \leq l \leq N, i \leq j \leq l$ , того, что (при прочих условиях, остающихся неизменными) через время  $y$  выделенная заявка окажется на  $j$ -м месте в очереди, равна<sup>1</sup>:

$$\begin{aligned} \tilde{\theta}_{(k,i)(l,j)}(y, z, x) &= \\ &= e^{-\lambda y} \theta_{(k,i)(l,j)}(y, z, x) + \sum_{n=1}^{\infty} \frac{(\lambda y)^n}{n!} e^{-\lambda y} \times \\ &\times \sum_{l_1=k}^{l-1} \sum_{j_1=i}^j \int_0^{\infty} \theta_{(k,i)(l_1,j_1)}(y, t, x) h_{(l_1,j_1),(l,j)}^{(n)}(x, t) dt, \end{aligned} \quad (19)$$

причем

$$\tilde{\theta}_{(N,i)(N,j)}(y, z, x) = \theta_{(N,i)(N,j)}(y, z, x) \mathbf{1}_{(j=i)}$$

при  $1 \leq i \leq N$ . Тогда, с учетом независимости длин заявок в группе, ПЛС  $\omega_{ki}^{(2)}(s; x)$ ,  $1 \leq k \leq N, 1 \leq i \leq k$ , стационарного (условного) распределения времени ожидания начала обслуживания заявки, заставшей при поступлении на приборе заявку второго типа, равно:

$$\begin{aligned} \omega_{ki}^{(2)}(s; x) &= \sum_{l=k}^N \sum_{j=i}^l \int_0^{\infty} \int_0^{\infty} e^{-sy} \tilde{\theta}_{(k,i)(l,j)}(y, z, x) \times \\ &\times \left( \mathbf{1}_{(j=1)} + u_l(s; z) \prod_{m=1}^{j-2} u_{l-m}(s) \mathbf{1}_{(j \neq 1)} \right) dz dy, \end{aligned} \quad (20)$$

усредняя которое, как в (18), получаем безусловное ПЛС  $\omega^{(2)}(s)$ . Таким образом, стационарное распределение времени ожидания начала обслуживания принятой в систему заявки имеет ПЛС  $(\omega^{(1)}(s) + \omega^{(2)}(s))(1 - \pi)^{-1}$ .

Аналогичным образом находится и ПЛС стационарного распределения времени пребывания заявки первого типа в системе. Действительно, поскольку это время складывается из времени ожидания начала обслуживания и времени пребывания на приборе с учетом возможных прерываний, то каждое слагаемое в (16), (17) и (20) достаточно домножить на ПЛС времени до уменьшения на единицу числа заявок первого типа в системе, т. е. на  $u_m(s; x)$  при соответствующем  $m$ .

Заметим, что распределение времени пребывания в системе заявок второго типа не представляет интереса: оно равно  $G(x)$ . Что касается периодов времени между моментами начала поступления заявок второго типа на прибор, то они совпадают по распределению с циклом занятости системы заявками первого типа.

<sup>1</sup> Отметим, что при расчетах по формуле (19) приходится суммировать конечное число слагаемых, поскольку из принятого правила постановки новых заявок в очередь (и того, что  $N < \infty$ ) следует, что плотности  $h_{(m_1, i_1), (m_2, i_2)}^{(n)}(x_1, x_2)$  (при фиксированных  $m_1$  и  $m_2$ ) совпадают начиная с некоторого  $n$ .

## 5 Заключение

Дифференцируя полученные в предыдущем разделе формулы необходимое число раз, можно найти моменты стационарных распределений времени ожидания начала обслуживания и времени пребывания заявки в системе. Для нахождения средних значений также можно (и удобнее) пользоваться формулой Литтла.

На основе полученных результатов находятся и различные совместные стационарные распределения (например, периода занятости заявками первого типа и числа обслуженных и потерянных на нем заявок и т. п.). Кроме того, из результатов разд. 4 следует, что стационарные временные характеристики могут быть найдены в случае, когда порядок постановки в очередь заявок первого типа в момент обслуживания заявки второго типа произвольный. В дальнейшем интерес представляет обобщение предложенного метода на случай более общего входного потока (как, например, в [15, 16]), нескольких классов (с различными функциями распределения длин) заявок второго типа и других правил начала обслуживания заявок первого типа (например,  $N$ -политики [5]).

## Литература

1. Lee T. T. M/G/1/N queue with vacation time and exhaustive service discipline // Oper. Res., 1984. Vol. 32. P. 774–785.
2. Печинкин А. В. Двухприоритетная система массового обслуживания с инверсионным порядком обслуживания // Техника средств связи. Сер. СС, 1985. Вып. 1. С. 72–81.
3. Бочаров П. П., Шлумпер Л. О. Однолинейная система массового обслуживания с фоновыми заявками // Автомат. телемех., 2005. № 6. С. 74–88.
4. Razumchik R. Two-priority queueing system with LCFS service, probabilistic priority and batch arrivals // AIP Conf. Proc., 2019. Vol. 2116. No. 1. P. 090011-1–090011-3.
5. Kempa W. M. Analytical model of a wireless sensor network (WSN) node operation with a modified threshold-type energy saving mechanism // Sensors Basel, 2019. Vol. 19. Iss. 14. Art. ID: 3114.
6. Печинкин А. В. Об одной инвариантной системе массового обслуживания // Math. Operationsforsch. Statist. Ser. Optimization, 1983. Vol. 14. No. 3. P. 433–444.

7. Kim C., Dudin A., Dudina O., Klimenok V. Analysis of queueing system with non-preemptive time limited service and impatient customers // *Methodol. Comput. Appl.*, 2020. Vol. 22. P. 401–432.
8. Kempa W. M., Marjasz R. Distribution of the time to buffer overflow in the  $M/G/1/N$ -type queueing model with batch arrivals and multiple vacation policy // *J. Oper. Res. Soc.*, 2020. Vol. 71. Iss. 3. P. 447–455.
9. Нагоненко В. А. О характеристиках одной нестандартной системы массового обслуживания. I, II // *Изв. АН СССР. Технич. кибернет.*, 1981. № 1. С. 187–195; № 3. С. 91–99.
10. Милованова Т. А., Печинкин А. В. Стационарные характеристики системы обслуживания с инверсионным порядком обслуживания, вероятностным приоритетом и гистерезисной политикой // *Информатика и её применения*, 2013. Т. 7. Вып. 1. С. 22–36.
11. Мейханаджян Л. А., Милованова Т. А., Печинкин А. В., Разумчик Р. В. Стационарные вероятности состояний в системе обслуживания с инверсионным порядком обслуживания и обобщенным вероятностным приоритетом // *Информатика и её применения*, 2014. Т. 8. Вып. 3. С. 16–26.
12. Поснелов В. В. О погрешности приближения функции двух переменных суммами произведений функций одного переменного // *Ж. вычисл. мат. мат. физ.*, 1978. Т. 18. Вып. 5. С. 1307–1308.
13. Uschmajew A. Regularity of tensor product approximations to square integrable functions // *Constr. Approx.*, 2011. Vol. 34. Iss. 3. P. 371–391.
14. Townsend A., Trefethen L. N. An extension of chebfun to two dimensions // *SIAM J. Sci. Comput.*, 2013. Vol. 35. Iss. 6. P. 495–518.
15. Разумчик Р. В. Стационарные характеристики системы обслуживания с инверсионным порядком обслуживания, вероятностным приоритетом и групповым поступлением разнородных заявок // *Информатика и её применения*, 2017. Т. 11. Вып. 4. С. 10–18.
16. Razumchik R. On  $M/G/1$  queue with state-dependent heterogeneous batch arrivals, inverse service order and probabilistic priority // *AIP Conf. Proc.*, 2017. Vol. 1863. No. 1. P. 090006-1–090006-3.

Поступила в редакцию 15.07.20

## A SINGLE-SERVER QUEUEING SYSTEM WITH LIFO SERVICE, PROBABILISTIC PRIORITY, BATCH POISSON ARRIVALS, AND BACKGROUND CUSTOMERS

T. A. Milovanova<sup>1</sup> and R. V. Razumchik<sup>2</sup>

<sup>1</sup>Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

<sup>2</sup>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

**Abstract:** Consideration is given to the single-server queueing system with two independent flows of customers: a batch Poisson flow of (primary) customers and a saturated flow of background customers. Primary customers have relative priority over background customers, i. e., the service of a background customer cannot be interrupted. A background customer is instantly taken for service every time the buffer for primary customers is empty upon the service completion. The service times of primary and background customers are independent and are allowed to be generally distributed. The implemented service policy is LIFO (last in, first out) with the probabilistic priority. The method and analytic expressions for the computation (in terms of transforms) of the system's main stationary performance characteristics, including the stationary distribution of the waiting and sojourn times of the primary customers, are presented.

**Keywords:** queueing system; LIFO service; probabilistic priority; batch arrivals; background customers

**DOI:** 10.14357/19922264200304

### Acknowledgments

The reported study was funded by RFBR, project No. 20-07-00804.

## References

1. Lee, T. T. 1984.  $M/G/1/N$  queue with vacation time and exhaustive service discipline. *Oper. Res.* 32: 774–785.
2. Pechinkin, A. V. 1985. Dvukhpriornitnaya sistema massovogo obsluzhivaniya s inversionnym poryadkom obsluzhivaniya [Two-priority queuing system with inversion order of service]. *Tekhnika sredstv svyazi Ser. SS* [Communication technology. Communication systems series] 1:72–81.
3. Bocharov, P. P., and L. O. Shlumper. 2005. A single-server queueing system with background customers. *Automat. Rem. Contr.* 66(6):917–930.
4. Razumchik, R. 2019. Two-priority queueing system with LCFS service, probabilistic priority and batch arrivals. *AIP Conf. Proc.* 2116(1):090011-1–090011-3.
5. Kempa, W. M. 2019. Analytical model of a wireless sensor network (WSN) node operation with a modified threshold-type energy saving mechanism. *Sensors Basel* 19(14):3114.
6. Pechinkin, A. V. 1983. Ob odnoy invariantnoy sisteme massovogo obsluzhivaniya [On an invariant queuing system]. *Math. Operationsforsch. Statist. Ser. Optimization* 14(3):433–444.
7. Kim, C., A. Dudin, O. Dudina, and V. Klimenok. 2020. Analysis of queueing system with non-preemptive time limited service and impatient customers. *Methodol. Comput. Appl.* 22:401–432.
8. Kempa, W. M., and R. Marjasz. 2020. Distribution of the time to buffer overflow in the  $M/G/1/N$ -type queueing model with batch arrivals and multiple vacation policy. *J. Oper. Res. Soc.* 71(3):447–455.
9. Nagonenko, V. A. 1981. O kharakteristikakh odnoy nestandartnoy sistemy massovogo obsluzhivaniya [On the characteristics of one non-standard queuing system]. I, II *J. Comput. Sys. Sc. Int.* 1:187–195; 3:91–99.
10. Milovanova, T. A., and A. V. Pechinkin. 2013. Statsionarnye kharakteristiki sistemy obsluzhivaniya s inversionnym poryadkom obsluzhivaniya, veroyatnostnym prioritetom i gisterezisnoy politikoy [Stationary characteristics of queuing system with an inversion procedure service probabilistic priority and hysteresis policy] *Informatika i ee Primeneniya — Inform. Appl.* 7(1):22–35.
11. Meykhanadzhyan, L. A., T. A. Milovanova, A. V. Pechinkin, and R. V. Razumchik. 2014. Statsionarnye veroyatnosti sostoyaniy v sisteme obsluzhivaniya s inversionnym poryadkom obsluzhivaniya i obobshchennym veroyatnostnym prioritetom [Stationary distribution in a queueing system with inverse service order and generalized probabilistic priority]. *Informatika i ee Primeneniya — Inform. Appl.* 8(3):16–26.
12. Pospelov, V. V. 1978. O pogreshnosti priblizheniya funktsii dvuh peremennykh summami proizvedeniy funktsiy odnogo peremennogo [The error of approximation of a function of two variables by sums of the products of functions of one variable]. *USSR Comp. Math. Math.* 18(5):1307–1308.
13. Uschmajew, A. 2011. Regularity of tensor product approximations to square integrable functions. *Constr. Approx.* 34(3):371–391.
14. Townsend, A., and L. N. Trefethen. 2013. An extension of chebfun to two dimensions. *SIAM J. Sci. Comput.* 35(6):495–518.
15. Razumchik, R. V. 2017. Statsionarnye kharakteristiki sistemy obsluzhivaniya s inversionnym poryadkom obsluzhivaniya, veroyatnostnym prioritetom i gruppovym postupleniyem raznorodnykh zayavok [ $M/G/1$  queue with state-dependent heterogeneous batch arrivals, inverse service order, and probabilistic priority]. *Informatika i ee Primeneniya — Inform. Appl.* 8(3):16–26.
16. Razumchik, R. 2017. On  $M/G/1$  queue with state-dependent heterogeneous batch arrivals, inverse service order and probabilistic priority. *AIP Conf. Proc.* 1863(1):090006-1–090006-3.

Received July 15, 2020

## Contributors

**Milovanova Tatiana A.** (b. 1977) — Candidate of Science (PhD) in physics and mathematics, senior lecturer, Department of Applied Informatics and Probability Theory, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; [tmilovanova77@mail.ru](mailto:tmilovanova77@mail.ru)

**Razumchik Rostislav V.** (b. 1984) — Candidate of Science (PhD) in physics and mathematics, 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; [rrazumchik@ipiran.ru](mailto:rrazumchik@ipiran.ru)

# О РАСПРЕДЕЛЕНИИ ОТНОШЕНИЯ СУММЫ ЭЛЕМЕНТОВ ВЫБОРКИ, ПРЕВОСХОДЯЩИХ НЕКОТОРЫЙ ПОРОГ, К СУММЕ ВСЕХ ЭЛЕМЕНТОВ ВЫБОРКИ. I\*

В. Ю. Королев<sup>1</sup>

**Аннотация:** Рассматривается задача описания распределения доли суммы независимых случайных величин, которая составлена из слагаемых, превосходящих некоторый заданный порог. В отличие от известных вариантов такой задачи, в которых фиксируется число суммируемых крайних порядковых статистик, особенность рассматриваемой здесь задачи состоит в том, что заданный порог может быть превзойден непредсказуемым числом элементов выборки. В статье в терминах функции распределения отдельного слагаемого формально представлен явный вид распределения отношения суммы элементов выборки, превосходящих заданный порог, к сумме всех наблюдений. На эвристическом уровне выведены асимптотические и предельные распределения этого отношения при фиксированном пороге, удобные для использования в качестве асимптотических аппроксимаций в практических вычислениях. Рассмотрены ситуации, в которых распределение слагаемых имеет легкий хвост (конечны вторые моменты), и ситуации, в которых распределение слагаемых имеет тяжелый хвост (принадлежит к области притяжения устойчивого закона). Во всех случаях описана нормировка отношения, гарантирующая невырожденность предельного (при неограниченном увеличении числа слагаемых) распределения, и само предельное распределение (нормальное в случае легких хвостов и устойчивое в случае тяжелых хвостов).

**Ключевые слова:** сумма независимых случайных величин; случайная сумма; биномиальное распределение; смесь распределений вероятностей; нормальное распределение; устойчивое распределение; экстремальная порядковая статистика

DOI: 10.14357/19922264200305

## 1 Постановка задачи

В 1897 г. итальянский экономист и социолог Вильфридо Парето выявил эмпирическую закономерность, заключающуюся в том, что 80% дохода страны аккумулируются в 20% семей [1]. Эту закономерность многие пытались обосновать или опровергнуть. В данной заметке предпринята попытка рассмотреть связанную с этим принципом вероятностно-статистическую задачу о том, какую долю суммы наблюдений составляют наблюдения, превосходящие заданный порог.

Эта задача имеет большое значение не только для экономики, но и для других областей знания. Например, проследивая изменение во времени параметров распределения отношения суммы элементов выборки, превосходящих некоторый порог, к сумме всех элементов выборки при исследовании метеорологических данных (температура, осадки, теплообмен между атмосферой и океаном) во времени (например, в скользящем режиме, когда выборка — это «окно», сдвигающееся в направлении

астрономического времени при исследовании соответствующего временного ряда), можно получить информацию об особенностях проявления процесса изменения климата, в просторечии называемого «глобальным потеплением».

С математической точки зрения эта задача тесно связана со статистикой цензурированных выборок. Известны разные варианты этой задачи. Например, в работах [2–10] рассмотрена задача о распределении суммы заданного числа крайних порядковых статистик. Однако в опубликованных на эту тему работах рассматривались либо конкретные распределения, либо теоретические результаты типа законов больших чисел.

Пусть  $n \in \mathbb{N}$ ,  $X_1, \dots, X_n$  — выборка, т.е. набор независимых одинаково распределенных случайных величин с общей функцией распределения  $F(x) = P(X_1 < x)$ ,  $x \in \mathbb{R}$ . Без существенного ограничения общности для удобства будем считать, что  $F(x)$  непрерывна и  $F(0) = 0$ , т.е.  $P(X_1 \geq 0) = 1$ . Пусть  $u \in (0, \infty)$  — рассматриваемый порог.

\* Работа выполнена при частичной финансовой поддержке РФФИ (проект 19-07-00914) и в соответствии с программой Московского центра фундаментальной и прикладной математики.

<sup>1</sup> Факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vkorolev@cs.msu.ru

Пусть  $X_{(1)}, \dots, X_{(n)}$  — вариационный ряд, построенный по выборке  $X_1, \dots, X_n$ . В качестве приближенного решения указанной задачи можно рассмотреть распределение случайной величины  $\sum_{j=[nF(u)]}^n X_{(j)} / \sum_{j=1}^n X_{(j)}$ . Здесь в числителе стоит сумма крайних порядковых статистик, начиная с выборочной квантили порядка  $F(u)$ , которая при  $n \rightarrow \infty$  сходится к теоретической квантили порядка  $F(u)$ , т. е. к порогу  $u$ .

Еще один подход к решению общей задачи заключается в учете того, что число  $N_n(u)$  элементов выборки, превосходящих порог  $u$ , является случайной величиной, имеющей биномиальное распределение с параметрами  $n$  и  $1 - F(u)$ . Тогда отношение суммы элементов выборки, превосходящих порог  $u$ , к сумме всех наблюдений принимает вид:  $\sum_{j=n-N_n(u)+1}^n X_{(j)} / \sum_{j=1}^n X_{(j)}$ . Для изучения распределения этой случайной величины в некоторых частных случаях можно было бы применить результаты работ [1, 2, 4, 6, 9], но, к сожалению, в приведенных выше отношениях участвуют случайные величины, не являющиеся независимыми.

Чтобы обойти указанное препятствие, в данной статье использован альтернативный подход. Отметим, что, в отличие от упоминавшихся выше работ, в рассматриваемой здесь задаче фиксируется не число суммируемых крайних порядковых статистик, а порог, который может быть превзойден непредсказуемым числом элементов выборки. Ниже будет формально представлен явный вид распределения отношения суммы элементов выборки, превосходящих заданный порог, к сумме всех наблюдений и на эвристическом уровне будут выведены асимптотические и предельные распределения этого отношения при фиксированном пороге, удобные для использования в качестве асимптотических аппроксимаций в практических вычислениях.

Обозначим  $S_n = X_1 + \dots + X_n$ . Индикатор множества (события)  $A$  обозначим  $\mathbb{I}(A)$ .

Пусть  $u > 0$  таково, что  $0 < F(u) < 1$ . Очевидно,  $X_j = X_j \mathbb{I}(X_j < u) + X_j \mathbb{I}(X_j \geq u)$ . Тогда

$$\begin{aligned} S_n &= \sum_{j=1}^n X_j \mathbb{I}(X_j < u) + \sum_{j=1}^n X_j \mathbb{I}(X_j \geq u) \equiv \\ &\equiv S_n^{(<u)} + S_n^{(\geq u)}. \end{aligned}$$

Основным объектом изучения будут распределения случайных величин  $S_n^{(\geq u)}$  и  $S_n^{(<u)}$ , а в первую очередь — распределение отношения  $R(u) = S_n^{(\geq u)} / S_n$ . Очевидно, что  $0 \leq R(u) \leq 1$  и

$$R(u) = \left( 1 + \frac{S_n^{(<u)}}{S_n^{(\geq u)}} \right)^{-1}.$$

Поэтому если обозначить

$$\underline{W}_u(r) = P(S_n^{(<u)} < r S_n^{(\geq u)}), \quad r \geq 0,$$

то для  $r \in [0, 1]$

$$P(R(u) < r) = 1 - \underline{W}_u\left(\frac{1}{r} - 1\right).$$

Так что сосредоточимся на описании функций распределения  $\underline{W}_u(r)$  и  $\overline{W}_u(r)$ :

$$\overline{W}_u(r) \equiv P(S_n^{(\geq u)} < r S_n^{(<u)}) = 1 - \underline{W}_u\left(\frac{1}{r}\right).$$

## 2 Точное решение

Очевидно,

$$X_j \mathbb{I}(X_j \geq u) = \begin{cases} 0, & X_j < u; \\ X_j, & X_j \geq u. \end{cases}$$

Если условиться считать равенство нулю индикатора  $\mathbb{I}(X_j < u)$  «неудачей», а противоположное событие — «успехом», то число  $N_n(u)$  ненулевых слагаемых в сумме  $S_n^{(\geq u)}$  будет случайной величиной, имеющей биномиальное распределение с параметрами  $n$  и  $1 - F(u)$ . Таким образом, для  $x \geq 0$

$$P\left(S_n^{(\geq u)} < x\right) = P\left(\sum_{j=0}^{N_n(u)} X_j^{(\geq u)} < x\right), \quad (1)$$

где случайные величины  $X_1^{(\geq u)}, X_2^{(\geq u)}, \dots$  независимы и имеют одну и ту же функцию распределения:

$$\begin{aligned} F^{(\geq u)}(x) &\equiv P\left(X_1^{(\geq u)} < x\right) = 1 - P\left(X_1^{(\geq u)} \geq x\right) = \\ &= 1 - P\left(X_1 \mathbb{I}(X_1 \geq u) \geq x \mid X_1 \geq u\right) = \\ &= 1 - \frac{P\left(X_1 \geq x, X_1 \geq u\right)}{P\left(X_1 \geq u\right)} = \\ &= \begin{cases} 0, & x < u; \\ \frac{F(x) - F(u)}{1 - F(u)}, & x \geq u \end{cases} = \\ &= \frac{F(\max\{x, u\}) - F(u)}{1 - F(u)}. \end{aligned}$$

При этом случайная величина  $N_n(u)$  независима от последовательности  $X_1^{(\geq u)}, X_2^{(\geq u)}, \dots$

Аналогично для  $x \geq 0$

$$P\left(S_n^{(<u)} < x\right) = P\left(\sum_{j=0}^{n-N_n(u)} X_j^{(<u)} < x\right), \quad (2)$$

где случайные величины  $X_1^{(<u)}, X_2^{(<u)}, \dots$  независимы и имеют одну и ту же функцию распределения:

$$\begin{aligned}
 F^{(<u)}(x) &\equiv \mathbb{P}\left(X_1^{(<u)} < x\right) = \\
 &= \mathbb{P}\left(X_1 \mathbb{I}(X_1 < u) < x \mid X_1 < u\right) = \\
 &= \frac{\mathbb{P}\left(X_1 < x, X_1 < u\right)}{\mathbb{P}\left(X_1 < u\right)} = \begin{cases} 1, & x \geq u; \\ \frac{F(x)}{F(u)}, & 0 \leq x < u \end{cases} = \\
 &= \frac{F(\min\{x, u\})}{F(u)}.
 \end{aligned}$$

При этом случайная величина  $N_n(u)$  также независима от последовательности  $X_1^{(<u)}, X_2^{(<u)}, \dots$ . Более того, случайные величины  $X_1^{(\geq u)}, X_2^{(\geq u)}, \dots, X_1^{(<u)}, X_2^{(<u)}, \dots$  независимы в совокупности.

Таким образом, для  $r \geq 0$  по формуле полной вероятности имеем:

$$\begin{aligned}
 \underline{W}_u(r) &= \mathbb{P}\left(\sum_{j=0}^{n-N_n(u)} X_j^{(<u)} < r \sum_{j=0}^{N_n(u)} X_j^{(\geq u)}\right) = \\
 &= \sum_{k=0}^n C_n^k (1 - F(u))^k (F(u))^{n-k} \times \\
 &\quad \times \mathbb{P}\left(\sum_{j=0}^{n-k} X_j^{(<u)} < r \sum_{j=0}^k X_j^{(\geq u)}\right). \quad (3)
 \end{aligned}$$

Следует особо отметить, что в последнем соотношении под знаком вероятности стоят суммы случайных величин, *независимых в совокупности*. Имеем:

$$\begin{aligned}
 \underline{W}_u(r) &= \sum_{k=0}^n C_n^k (1 - F(u))^k (F(u))^{n-k} \times \\
 &\quad \times \int_0^\infty B_{n-k}^{(<u)}(rz) dB_k^{(\geq u)}(z),
 \end{aligned}$$

где для  $m \geq 0$

$$B_m^{(<u)}(x) = (F(u))^{-m} F^{\nabla m}(x),$$

$$F^{\nabla 0}(x) = \begin{cases} 0, & x \leq 0; \\ 1, & x > 0, \end{cases}$$

$$F^{\nabla(m+1)}(x) = \int_0^{\min\{x, u\}} F^{\nabla m}(x - y) dF(y);$$

$$B_m^{(\geq u)}(x) = (1 - F(u))^{-m} F^{\blacktriangle m}(x),$$

$$F^{\blacktriangle 0}(x) = \begin{cases} 0, & x \leq 0; \\ 1, & x > 0, \end{cases}$$

$$F^{\blacktriangle(m+1)}(x) = \int_u^\infty [F(\max\{x - y, u\}) - F(u)] dF^{\blacktriangle m}(y).$$

Тогда

$$\underline{W}_u(r) = \sum_{k=0}^n C_n^k \int_0^\infty F^{\nabla(n-k)}(rz) dF^{\blacktriangle k}(z). \quad (4)$$

Таким образом, в принципе, зная функцию распределения  $F(x)$ , по формуле (4) можно вычислить функцию распределения  $\underline{W}_u(r)$ .

Однако очевидно, что для непосредственного использования формулы (4) на практике придется применять громоздкие численные процедуры. Поэтому наиболее целесообразный подход заключается в применении асимптотических аппроксимаций, основанных на представлениях (1)–(3) и получаемых с помощью предельных теорем для сумм случайного числа случайных величин.

### 3 Приближенные решения

#### 3.1 Асимптотические аппроксимации для распределения отношения $S_n^{(\geq u)} / S_n^{(<u)}$ при фиксированном $u$

Поскольку случайные величины  $S_n^{(\geq u)}$  и  $S_n^{(<u)}$  не являются независимыми, изучение асимптотического распределения отношения  $S_n^{(\geq u)} / S_n^{(<u)}$  довольно трудоемко. Однако если хвосты распределения случайных величин достаточно легки (существуют математические ожидания), то при большом  $n$  некоторое представление об этом отношении даст оно само, так как по закону больших чисел это отношение будет стабилизироваться около некоторого числа.

Для более аккуратных выводов, учитывающих вероятностные свойства рассматриваемого отношения, приведем более подробные рассуждения.

Пусть  $u$  фиксировано. Обозначим:

$$p = p(u) = 1 - F(u).$$

Пусть  $\epsilon \in (0, 1)$  — малое число. Для  $r \geq 0$  по формуле полной вероятности имеем:

$$\begin{aligned}
 \underline{W}_u(r) &= \mathbb{P}\left(\sum_{j=0}^{n-N_n(u)} X_j^{(<u)} < r \sum_{j=0}^{N_n(u)} X_j^{(\geq u)}\right) = \\
 &= \sum_{k: |k/(np) - 1| \leq \epsilon} C_n^k (1 - F(u))^k (F(u))^{n-k} \times \\
 &\quad \times \mathbb{P}\left(\sum_{j=0}^{n-k} X_j^{(<u)} < r \sum_{j=0}^k X_j^{(\geq u)}\right) +
 \end{aligned}$$

$$\begin{aligned}
 & + \sum_{k: |k/(np)-1| > \epsilon} C_n^k (1-F(u))^k (F(u))^{n-k} \times \\
 & \times P \left( \sum_{j=0}^{n-k} X_j^{(<u)} < r \sum_{j=0}^k X_j^{(\geq u)} \right) \equiv I_1 + I_2. \quad (5)
 \end{aligned}$$

Рассмотрим  $I_2$ . Очевидно,  $EN_n(u) = np(u)$ ,  $DN_n(u) = np(u)(1-p(u))$ . Тогда по неравенству Чебышёва имеем:

$$\begin{aligned}
 I_2 & \leq \sum_{k: |k/(np)-1| > \epsilon} C_n^k (1-F(u))^k (F(u))^{n-k} = \\
 & = P \left( \left| \frac{N_n(u)}{np(u)} - 1 \right| > \epsilon \right) \leq \frac{1-p(u)}{np(u)\epsilon^2} = \\
 & = \frac{F(u)}{n(1-F(u))\epsilon^2}.
 \end{aligned}$$

Таким образом, если  $u$  фиксировано, то  $I_2 = O(n^{-1}) \rightarrow 0$  при  $n \rightarrow \infty$ .

Рассмотрим  $I_1$ . Если  $k \in \{k : |k/(np) - 1| \leq \epsilon\}$ , то  $np(1-\epsilon) \leq k$  и, следовательно,

$$\min \left\{ k : \left| \frac{k}{np} - 1 \right| \leq \epsilon \right\} \rightarrow \infty$$

при  $n \rightarrow \infty$ . Одновременно если  $k \in \{k : |k/(np) - 1| \leq \epsilon\}$ , то  $k \leq np(1+\epsilon)$  и, следовательно,  $n-k \geq n[1-p(1+\epsilon)]$ . Поэтому если

$$\epsilon < \min \left\{ 1, \frac{F(u)}{1-F(u)} \right\},$$

то  $p(1+\epsilon) < 1$  и

$$\min \left\{ n-k : \left| \frac{k}{np} - 1 \right| \leq \epsilon \right\} \rightarrow \infty$$

при  $n \rightarrow \infty$ . Таким образом, при надлежащем выборе  $\epsilon$  в  $I_1$  в обеих суммах под знаком вероятности в (5) число слагаемых неограниченно возрастает при  $n \rightarrow \infty$ .

Случайные величины  $X_j^{(<u)}$  ограничены. Поэтому существуют  $EX_j^{(<u)} = \underline{a}_u$  и  $DX_j^{(<u)} = \underline{\sigma}_u^2$ . При этом при больших  $n$  для  $k \in \{k : |k/(np) - 1| \leq \epsilon\}$

$$P \left( \sum_{j=1}^{n-k} X_j^{(<u)} < x \right) \approx \Phi \left( \frac{x - (n-k)\underline{a}_u}{\underline{\sigma}_u \sqrt{n-k}} \right).$$

Предположим, что функция распределения  $F^{(\geq u)}(x)$  принадлежит к области притяжения некоторого устойчивого распределения  $G_\alpha(x)$  с характеристическим показателем  $\alpha \in (0, 2]$ . Это означает, что существуют числа  $c_k$  и  $d_k > 0$  такие, что при больших  $n$  для  $k \in \{k : |k/(np) - 1| \leq \epsilon\}$

$$P \left( \sum_{j=1}^k X_j^{(\geq u)} < x \right) \approx G_\alpha \left( \frac{x - c_k}{d_k} \right).$$

Тогда при  $n \rightarrow \infty$

$$\begin{aligned}
 \underline{W}_u(r) & = P \left( \sum_{j=0}^{n-N_n(u)} X_j^{(<u)} < r \sum_{j=0}^{N_n(u)} X_j^{(\geq u)} \right) = \\
 & = \sum_{k: |k/(np)-1| \leq \epsilon} C_n^k (1-F(u))^k (F(u))^{n-k} \times \\
 & \times P \left( \sum_{j=0}^{n-k} X_j^{(<u)} < r \sum_{j=0}^k X_j^{(\geq u)} \right) + O \left( \frac{1}{n} \right) \approx \\
 & \approx \sum_{k: |k/(np)-1| \leq \epsilon} C_n^k (1-F(u))^k (F(u))^{n-k} \times \\
 & \times \int_0^\infty \Phi \left( \frac{rz - (n-k)\underline{a}_u}{\underline{\sigma}_u \sqrt{n-k}} \right) dG_\alpha \left( \frac{z - c_k}{d_k} \right) \approx \\
 & \approx \sum_{k=0}^n C_n^k (1-F(u))^k (F(u))^{n-k} \times \\
 & \times \int_0^\infty \Phi \left( \frac{r - ((n-k)/z)\underline{a}_u}{\underline{\sigma}_u (\sqrt{n-k}/z)} \right) dG_\alpha \left( \frac{z - c_k}{d_k} \right), \quad (6)
 \end{aligned}$$

а если существует  $DX_j^{(\geq u)} = \bar{\sigma}_u^2$ , то  $\alpha = 2$ ,  $G_\alpha = \Phi$ ,  $c_k = kEX_j^{(\geq u)} \equiv k\bar{a}_u$ ,  $d_k = \bar{\sigma}_u \sqrt{k}$  и при  $n \rightarrow \infty$

$$\begin{aligned}
 \underline{W}_u(r) & \approx \underline{W}_u^{(n)}(r) \equiv \sum_{k=0}^n C_n^k (1-F(u))^k (F(u))^{n-k} \times \\
 & \times \int_0^\infty \Phi \left( \frac{r - ((n-k)/z)\underline{a}_u}{\underline{\sigma}_u (\sqrt{n-k}/z)} \right) d\Phi \left( \frac{z - k\bar{a}_u}{\bar{\sigma}_u \sqrt{k}} \right), \quad (7)
 \end{aligned}$$

где для определенности сумма с нулевым числом слагаемых полагается равной нулю. Таким образом, в данном случае при больших  $n$  функция распределения  $\underline{W}_n(r)$  может быть приближена специальной сдвиг-масштабной смесью нормальных законов, в которой смешивающее распределение (вообще говоря, двумерное) может иметь тяжелый хвост по своей непрерывной переменной.

На практике можно использовать приближенные функции распределения  $\underline{W}_n(r)$  конечной смесью нормальных законов, параметры которой статистически оцениваются, скажем, с помощью EM (expectation-maximization) алгоритма с учетом возможности оценить значение  $F(u)$  соответствующим значением эмпирической функции распределения.

3.1.1 Случай легких хвостов

Если объем выборки  $n$  велик, то аккуратные вычисления по формулам (6) и (7) затруднены из-за возможных искажений окончательного результата вследствие накапливающихся погрешностей. Поэтому дальнейшей целью будет получение аппроксимирующих распределений, не зависящих от  $n$  и адекватных при больших значениях объема выборки.

Пусть  $\xi_0$  и  $\xi_1$  — случайные величины с одним и тем же стандартным нормальным распределением, независимые от случайной величины  $N_n(u)$ . Тогда функция распределения  $W_u^{(n)}(r)$  (см. (7)) соответствует случайной величине

$$R_u^{(n)} = \frac{\sigma_u \sqrt{n - N_n(u)} \xi_0 + (n - N_n(u)) \underline{a}_u}{\bar{\sigma}_u \sqrt{N_n(u)} \xi_1 + N_n(u) \bar{a}_u} =$$

$$= \left( \sqrt{1 - \frac{N_n(u)}{n}} \frac{1}{\sqrt{n}} \sigma_u \xi_0 + \left(1 - \frac{N_n(u)}{n}\right) \frac{\underline{a}_u}{\bar{a}_u} \right) \times$$

$$\times \left( \sqrt{\frac{N_n(u)}{n}} \frac{1}{\sqrt{n}} \bar{\sigma}_u \xi_1 + \frac{N_n(u)}{n} \frac{\bar{a}_u}{\underline{a}_u} \right)^{-1} \rightarrow$$

$$\rightarrow \frac{F(u)}{1 - F(u)} \frac{\underline{a}_u}{\bar{a}_u} = \frac{\int_0^u x dF(x)}{\int_u^\infty x dF(x)}$$

по вероятности при  $n \rightarrow \infty$ . Поэтому в качестве статистической оценки математического ожидания отношения суммы наблюдений, меньших порога  $u$ , к сумме наблюдений, превосходящих этот порог, можно взять само это отношение: если  $F_n(x)$  — эмпирическая функция распределения, построенная по выборке  $X_1, \dots, X_n$ ,  $F_n(x) = (1/n) \sum_{j=1}^n \mathbb{I}(X_j < x)$ ,  $x \in \mathbb{R}$ , то

$$\frac{\int_0^u x dF(x)}{\int_u^\infty x dF(x)} \approx \frac{\int_0^u x dF_n(x)}{\int_u^\infty x dF_n(x)} =$$

$$= \frac{\sum_{j=1}^n X_j \mathbb{I}(X_j < u)}{\sum_{j=1}^n X_j \mathbb{I}(X_j \geq u)}.$$

Для более точных выводов, выполняя очевидные преобразования и применяя теорему Муавра–Лапласа, получаем:

$$\sqrt{n} \left( R_u^{(n)} - \frac{F(u)}{1 - F(u)} \frac{\underline{a}_u}{\bar{a}_u} \right) =$$

$$= \frac{\sqrt{1 - N_n(u)/n} \sigma_u \xi_0 + \sqrt{n}(1 - N_n(u)/n) \underline{a}_u}{\sqrt{N_n(u)/n} (1/\sqrt{n}) \bar{\sigma}_u \xi_1 + (N_n(u)/n) \bar{a}_u} -$$

$$- \frac{\sqrt{n} F(u)}{1 - F(u)} \frac{\underline{a}_u}{\bar{a}_u} =$$

$$= \frac{\sqrt{1 - N_n(u)/n} \sigma_u \xi_0}{\sqrt{N_n(u)/n} (1/\sqrt{n}) \bar{\sigma}_u \xi_1 + (N_n(u)/n) \bar{a}_u} +$$

$$+ \frac{\sqrt{n}(1 - N_n(u)/n) \underline{a}_u}{\sqrt{N_n(u)/n} (1/\sqrt{n}) \bar{\sigma}_u \xi_1 + (N_n(u)/n) \bar{a}_u} -$$

$$- \frac{\sqrt{n} F(u)}{1 - F(u)} \frac{\underline{a}_u}{\bar{a}_u} = \frac{\sqrt{F(u)} \sigma_u}{(1 - F(u)) \bar{a}_u + o_P(1)} \xi_0 +$$

$$+ \frac{(n - N_n(u)/\sqrt{n}) \underline{a}_u}{(1 - F(u)) \bar{a}_u + o_P(1)} - \frac{\sqrt{n} F(u)}{1 - F(u)} \frac{\underline{a}_u}{\bar{a}_u} + o_P(1) =$$

$$= \frac{\sqrt{F(u)} \sigma_u}{(1 - F(u)) \bar{a}_u + o_P(1)} \xi_0 + \frac{\sqrt{n} F(u) \underline{a}_u}{(1 - F(u)) \bar{a}_u + o_P(1)} -$$

$$- \frac{\sqrt{n} F(u)}{1 - F(u)} \frac{\underline{a}_u}{\bar{a}_u} -$$

$$- \frac{N_n - n(1 - F(u))}{\sqrt{n} F(u) (1 - F(u))} \frac{\sqrt{F(u)} \underline{a}_u}{\sqrt{1 - F(u)} \bar{a}_u + o_P(1)} +$$

$$+ o_P(1) \xrightarrow{d} \frac{\sqrt{F(u)} \sigma_u}{(1 - F(u)) \bar{a}_u} \xi_0 + \frac{\sqrt{F(u)} \underline{a}_u}{\sqrt{1 - F(u)} \bar{a}_u} \xi_1 \stackrel{d}{=} \xi_0 \frac{1}{\bar{a}_u} \sqrt{\frac{F(u)}{1 - F(u)} \left( \frac{\sigma_u^2}{1 - F(u)} + \underline{a}_u^2 \right)},$$

где символом  $o_P(1)$  обозначены разные случайные величины, сходящиеся к нулю по вероятности при  $n \rightarrow \infty$ , а символ  $\xrightarrow{d}$  обозначает сходимость по распределению. Таким образом, в случае легких хвостов распределения исходных случайных величин  $X_j$  (т.е. существования дисперсий случайных величин  $X_j^{(\geq u)}$ ) для любого  $x$

$$\lim_{n \rightarrow \infty} P \left( \sqrt{n} \left( R_u^{(n)} - \frac{F(u)}{1 - F(u)} \frac{\underline{a}_u}{\bar{a}_u} \right) < x \right) =$$

$$= \Phi \left( x \bar{a}_u \left[ \frac{F(u)}{1 - F(u)} \left( \frac{\sigma_u^2}{1 - F(u)} + \underline{a}_u^2 \right) \right]^{-1/2} \right).$$

Другими словами, в указанных выше предположениях случайная величина  $R_u^{(n)}$  асимптотически нормальна с математическим ожиданием

$$\frac{F(u)}{1 - F(u)} \frac{\underline{a}_u}{\bar{a}_u} = \frac{\int_0^u x dF(x)}{\int_u^\infty x dF(x)}$$

и дисперсией

$$\frac{1}{\bar{a}_u^2 n} \frac{F(u)}{1-F(u)} \left( \frac{\sigma_u^2}{1-F(u)} + \bar{a}_u^2 \right) = \frac{1}{n} \frac{\int_0^u x^2 dF(x)}{\left( \int_u^\infty x dF(x) \right)^2}.$$

### 3.1.2 Случай тяжелых хвостов

Предположим, что распределение случайных величин  $X_j^{(\geq u)}$  имеет тяжелый хвост, а именно: будем считать, что это распределение принадлежит к области нормального притяжения строго устойчивого распределения с характеристическим показателем  $\alpha \in (0, 2]$ .

Сначала рассмотрим случай  $\alpha \in (1, 2]$ . В такой ситуации у случайной величины  $X_j^{(\geq u)}$  существует математическое ожидание  $\bar{a}_u$  и условие принадлежности ее распределения к области нормального притяжения строго устойчивого распределения с характеристическим показателем  $\alpha \in (1, 2]$  заключается в существовании числа  $b_u \in (0, \infty)$ , тако-го что

$$\frac{1}{b_u n^{1/\alpha}} \sum_{j=1}^n \left( X_j^{(\geq u)} - \bar{a}_u \right) \xrightarrow{d} \zeta_\alpha$$

при  $n \rightarrow \infty$ , где  $P(\zeta_\alpha < x) \equiv G_\alpha(x)$ . Другими словами, в (6)  $c_k = k\bar{a}_u$  и  $d_k = b_u k^{1/\alpha}$ , так что функция распределения в правой части (6) соответствует случайной величине

$$\bar{R}_u^{(n)} = \frac{\sigma_u \sqrt{n - N_n(u)} \xi_0 + (n - N_n(u)) \bar{a}_u}{b_u (N_n(u))^{1/\alpha} \zeta_\alpha + N_n(u) \bar{a}_u}.$$

Вместо случайной величины  $\bar{R}_u^{(n)}$  удобнее рассматривать обратную случайную величину  $\tilde{R}_u^{(n)} = \left( \bar{R}_u^{(n)} \right)^{-1}$ . Имеем:

$$\tilde{R}_u^{(n)} = \frac{b_u N_n(u)^{1/\alpha} \zeta_\alpha}{\sigma_u \sqrt{n - N_n(u)} \xi_0 + (n - N_n(u)) \bar{a}_u} + \frac{(N_n(u)/n) \bar{a}_u}{\sigma_u \sqrt{(n - N_n(u))/n} (1/\sqrt{n}) \xi_0 + (1 - N_n(u)/n) \bar{a}_u}.$$

При этом

$$\frac{N_n(u)}{n} \bar{a}_u \left( \frac{\sigma_u \sqrt{\frac{n - N_n(u)}{n}}}{1} \frac{1}{\sqrt{n}} \xi_0 + \left( 1 - \frac{N_n(u)}{n} \right) \bar{a}_u \right)^{-1} \rightarrow \frac{(1 - F(u)) \bar{a}_u}{F(u) \bar{a}_u}$$

по вероятности при  $n \rightarrow \infty$ . В то же время

$$\begin{aligned} n^{1-1/\alpha} \left( \tilde{R}_u^{(n)} - \frac{(1 - F(u)) \bar{a}_u}{F(u) \bar{a}_u} \right) &= \\ &= \frac{b_u (N_n(u)/n)^{1/\alpha} \zeta_\alpha}{\sigma_u (\sqrt{n - N_n(u)/n} \xi_0 + (1 - N_n(u)/n) \bar{a}_u)} + \\ + o_P(1) &= \frac{b_u (1 - F(u))^{1/\alpha} \zeta_\alpha + o_P(1)}{F(u) \bar{a}_u + o_P(1)} + o_P(1) \xrightarrow{d} \\ &\xrightarrow{d} \frac{b_u (1 - F(u))^{1/\alpha}}{F(u) \bar{a}_u} \zeta_\alpha \end{aligned}$$

при  $n \rightarrow \infty$ .

Пусть теперь  $\alpha \in (0, 1]$ . В такой ситуации у случайной величины  $X_j^{(\geq u)}$  отсутствует математическое ожидание. Будем считать, что распределение случайной величины  $X_j^{(\geq u)}$  принадлежит к области нормального притяжения *одностороннего* строго устойчивого распределения с характеристическим показателем  $\alpha \in (0, 1]$ . Это означает, что существует такое число  $b_u > 0$ , что

$$\frac{1}{b_u n^{1/\alpha}} \sum_{j=1}^n X_j^{(\geq u)} \xrightarrow{d} \zeta_\alpha \quad (8)$$

при  $n \rightarrow \infty$ , где  $P(\zeta_\alpha < x) \equiv G_\alpha(x)$ . Таким образом, можно считать, что в (6)  $c_k = 0$  и  $d_k = b_u k^{1/\alpha}$ . При этом функция распределения, стоящая в правой части (6), соответствует случайной величине

$$\hat{R}_u^{(n)} = \frac{\sigma_u (n - N_n(u))^{1/2} \xi_0 + (n - N_n(u)) \bar{a}_u}{b_u N_n^{1/\alpha}(u) \zeta_\alpha}.$$

В то же время имеем:

$$\begin{aligned} n^{1/\alpha-1} \hat{R}_u^{(n)} &= \frac{\sigma_u ((n - N_n(u))^{1/2} / n^{1/\alpha}) \xi_0}{b_u (N_n/n)^{1/\alpha} \zeta_\alpha} + \\ &+ \frac{((n - N_n(u)) / n^{1/\alpha}) \bar{a}_u}{b_u (N_n/n)^{1/\alpha} \zeta_\alpha} = \\ &= n^{1/\alpha-1} \frac{\sigma_u \sqrt{(n - N_n(u)) / n} (1/n^{1/\alpha-1/2}) \xi_0}{b_u (N_n/n)^{1/\alpha} \zeta_\alpha} + \\ &+ \frac{n^{1-1/\alpha} F(u) \bar{a}_u}{b_u (N_n/n)^{1/\alpha} \zeta_\alpha} - \\ &- \frac{\left( ((N_n(u) - n(1 - F(u))) / \sqrt{n}) (1/n^{1/\alpha-1/2}) \right) \bar{a}_u}{b_u (N_n/n)^{1/\alpha} \zeta_\alpha} = \\ &= n^{1/\alpha-1} \frac{n^{1-1/\alpha} F(u) \bar{a}_u + o_P(1)}{b_u (1 - F(u))^{1/\alpha} \zeta_\alpha + o_P(1)} \xrightarrow{d} \\ &\xrightarrow{d} \frac{\bar{a}_u F(u)}{b_u (1 - F(u))^{1/\alpha} \zeta_\alpha} \end{aligned}$$

при  $n \rightarrow \infty$ . Следовательно, в случае тяжелых хвостов (отсутствия дисперсии у случайных величин  $X_j^{(\geq u)}$ ) имеет место сходимость:

$$n^{1-1/\alpha} \frac{S_n^{(\geq u)}}{S_n^{(<u)}} \xrightarrow{d} \frac{b_u (1 - F(u))^{1/\alpha}}{\underline{a}_u F(u)} \zeta_\alpha \quad (n \rightarrow \infty).$$

К такому выводу можно прийти и другим, более строгим путем. Для этого понадобится следующее вспомогательное утверждение.

Рассмотрим последовательность случайных величин  $\zeta_1, \zeta_2, \dots$ . Пусть  $N_1, N_2, \dots$  — натурально-значные случайные величины такие, что при каждом  $n$  случайная величина  $N_n$  независима от последовательности  $\zeta_1, \zeta_2, \dots$ . В следующей лемме сходимость подразумевается при  $n \rightarrow \infty$ .

**Лемма.** *Предположим, что существуют неограниченно возрастающая (убывающая к нулю) последовательность положительных чисел  $\{A_n\}_{n \geq 1}$  и случайная величина  $\zeta$  такие, что*

$$A_n^{-1} \zeta_n \implies \zeta. \quad (9)$$

*Если существуют неограниченно возрастающая (убывающая к нулю) последовательность положительных чисел  $\{B_n\}_{n \geq 1}$  и случайная величина  $N$  такие, что*

$$B_n^{-1} A_{N_n} \implies N, \quad (10)$$

то

$$B_n^{-1} \zeta_{N_n} \implies \zeta N, \quad (11)$$

*причем случайные сомножители в правой части (11) независимы. Если дополнительно  $N_n \rightarrow \infty$  по вероятности и семейство масштабных смесей функции распределения случайной величины  $\zeta$  идентифицируемо, то условие (10) не только достаточно для (11), но и необходимо.*

Доказательство см. в [11] (случай  $A_n, B_n \rightarrow \infty$ ) и [12] (случай  $A_n, B_n \rightarrow 0$ ).

Очевидно представление:

$$\begin{aligned} n^{1-1/\alpha} \frac{S_n^{(\geq u)}}{S_n^{(<u)}} &= \frac{S_n^{(\geq u)}}{n^{1/\alpha}} \frac{n}{S_n^{(<u)}} = \\ &= \frac{1}{n^{1/\alpha}} \sum_{j=0}^{N_n(u)} X_j^{(\geq u)} \left( \frac{1}{n} \sum_{j=0}^{n-N_n(u)} X_j^{(<u)} \right)^{-1}. \end{aligned} \quad (12)$$

В этом представлении перемножаемые суммы не являются независимыми. Рассмотрим второй множитель. В нем случайная величина  $\bar{N}_n(u) = n - N_n(u)$  имеет биномиальное распределение с параметрами  $n$  и  $F(u)$  и независима от случайных

величин  $X_1^{(<u)}, X_2^{(<u)}, \dots$ . При этом по усиленному закону больших чисел

$$\frac{1}{n} \sum_{j=0}^n X_j^{(<u)} \longrightarrow \underline{a}_u \quad (13)$$

с вероятностью единица, а значит, и по распределению. В то же время по усиленному закону больших чисел для бернуллиевых величин

$$n^{-1} \bar{N}_n(u) \longrightarrow F(u) \quad (14)$$

с вероятностью единица, а значит, и по распределению. Применяя лемму с  $A_n = B_n = n$  и соотношениями (13) и (14) в роли (9) и (10) соответственно, получаем:

$$n^{-1} S_n^{(<u)} \xrightarrow{d} \underline{a}_u F(u). \quad (15)$$

Поскольку константа независима от любой случайной величины, в представлении (12) множители асимптотически независимы.

Теперь рассмотрим первый множитель в правой части (12). В нем случайная величина  $N_n(u)$  имеет биномиальное распределение с параметрами  $n$  и  $1 - F(u)$  и независима от случайных величин  $X_1^{(\geq u)}, X_2^{(\geq u)}, \dots$ . По усиленному закону больших чисел для бернуллиевых величин

$$n^{-1} N_n(u) \longrightarrow 1 - F(u)$$

с вероятностью единица, а значит, и по распределению. При этом, очевидно,

$$n^{-1/\alpha} N_n^{1/\alpha}(u) \longrightarrow (1 - F(u))^{1/\alpha}. \quad (16)$$

Применяя лемму с  $A_n = B_n = n^{1/\alpha}$ , соотношениями (8) и (16) в роли (9) и (10) соответственно, получаем:

$$n^{-1/\alpha} S_n^{(\geq u)} \xrightarrow{d} b_u (1 - F(u))^{1/\alpha} \zeta_\alpha. \quad (17)$$

Используя соотношения (15) и (17), убеждаемся, что

$$\begin{aligned} n^{1-1/\alpha} \frac{S_n^{(\geq u)}}{S_n^{(<u)}} &= \frac{S_n^{(\geq u)}}{n^{1/\alpha}} \frac{n}{S_n^{(<u)}} \xrightarrow{d} \\ &\xrightarrow{d} \frac{b_u (1 - F(u))^{1/\alpha}}{\underline{a}_u F(u)} \zeta_\alpha \end{aligned}$$

при  $n \rightarrow \infty$ .

Таким образом, при указанной нормировке отношение суммы наблюдений, превосходящих порог  $u$ , к сумме наблюдений, меньших этого порога, асимптотически устойчиво с тем же самым характеристическим показателем  $\alpha$ .

## Литература

1. Koch R. The 80/20 principle: The secret of achieving more with less. — London: Nicholas Brealey Publishing, 1998. 304 p.
2. Prakasa Rao B. L. S. Limit theorems for sums of order statistics // Z. Wahrscheinlichkeit., 1976. Vol. 33. No. 4. P. 285–307.
3. Csörgő S., Horváth L., Mason D. M. What portion of the sample makes a partial sum asymptotically stable or normal? // Probab. Theory Rel., 1986. Vol. 72. P. 1–16.
4. Csörgő S., Haeusler E., Mason D. M. The asymptotic distribution of trimmed sums // Ann. Probab., 1988. Vol. 16. P. 672–699.
5. Csörgő S., Haeusler E., Mason D. M. A probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables // Adv. Appl. Math., 1988. Vol. 9. P. 259–333.
6. Csörgő S. Limit theorems for sums of order statistics // 6th Summer School (International) in Probability theory and Mathematical Statistics. — Sofia: Publishing House Bulgarian Acad. Sci., 1989. P. 5–37.
7. Sums, trimmed sums and extremes / Eds. M. G. Hahn, D. M. Mason, D. C. Weiner. — Boston, MA, USA: Birkhäuser, 1991. 428 p.
8. Kesten H., Maller R. A. Ratios of trimmed sums and order statistics // Ann. Probab., 1992. Vol. 20. No. 4. P. 1805–1842.
9. Janssen A. Invariance principles for sums of extreme sequential order statistics attracted to Levy processes // Stoch. Proc. Appl., 2000. Vol. 85. P. 255–277.
10. Ben Rached N., Botev Z., Kammoun A., Alouini M.-S., Tempone R. On the sum of order statistics and applications to wireless communication systems performances // IEEE T. Wirel. Commun., 2018. Vol. 17. No. 11. P. 7801–7813. doi: 10.1109/twc.2018.2871201.
11. Королев В. Ю. Сходимость случайных последовательностей с независимыми случайными индексами. I // Теория вероятн. и ее примен., 1994. Т. 39. Вып. 2. С. 313–333.
12. Королев В. Ю. Сходимость случайных последовательностей с независимыми случайными индексами. II // Теория вероятн. и ее примен., 1995. Т. 40. Вып. 4. С. 907–910.

Поступила в редакцию 28.11.19

## ON THE DISTRIBUTION OF THE RATIO OF THE SUM OF SAMPLE ELEMENTS EXCEEDING A THRESHOLD TO THE TOTAL SUM OF SAMPLE ELEMENTS. I

V. Yu. Korolev<sup>1,2</sup>

<sup>1</sup>Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, 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:** The problem of description of the distribution of the ratio of the sum of sample elements exceeding a threshold to the total sum of sample elements is considered. Unlike known versions of this problem in which the number of summed extreme order statistics is fixed, here, the specified threshold can be exceeded by an unpredictable number of sample elements. In the paper, in terms of the distribution function of a separate summand, the explicit form of the distribution of the ratio of the sum of sample elements exceeding a threshold to the total sum of sample elements is formally presented. The asymptotic and limit distributions are heuristically deduced for this ratio. These distributions are convenient for practical computations. The cases are considered in which the distributions of the summands have light tails (the second moments are finite) as well as the cases in which these distributions have heavy tails (belong to the domain of attraction of a stable law). In all cases, the normalization of the ratio is described that provides the existence of a nondegenerate limit (as the number of summands infinitely increases) distribution as well as the limit distribution itself (normal for the case of light tails and stable for the case of heavy tails).

**Keywords:** sum of independent random variables; random sum; binomial distribution; mixture of probability distributions; normal distribution; stable distribution; extreme order statistics

**DOI:** 10.14357/19922264200305

## Acknowledgments

The research was conducted in accordance with the scientific program of the Moscow Center for Fundamental and Applied Mathematics and supported by the Russian Foundation for Basic Research (project 19-07-00914).

## References

1. Koch, R. 1998. *The 80/20 principle: The secret of achieving more with less*. London: Nicholas Brealey Publishing. 304 p.
2. Prakasa Rao, B. L. S. 1976. Limit theorems for sums of order statistics. *Z. Wahrscheinlichkeit*. 33(4):285–307.
3. Csörgő, S., L. Horváth, and D. M. Mason. 1986. What portion of the sample makes a partial sum asymptotically stable or normal? *Probab. Theory Rel.* 72:1–16.
4. Csörgő, S., E. Haeusler, and D. M. Mason. 1988. The asymptotic distribution of trimmed sums. *Ann. Probab.* 16:672–699.
5. Csörgő, S., E. Haeusler, and D. M. Mason. 1988. A probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables. *Adv. Appl. Math.* 9:259–333.
6. Csörgő, S. 1989. Limit theorems for sums of order statistics. *6th Summer School (International) in Probability Theory and Mathematical Statistics*. Sofia: Publishing House Bulgarian Acad. Sci. 5–37.
7. Hahn, M. G., D. M. Mason, and D. C. Weiner, eds. 1991. *Sums, trimmed sums and extremes*. Boston, MA: Birkhäuser. 428 p.
8. Kesten, H., and R. A. Maller. 1992. Ratios of trimmed sums and order statistics. *Ann. Probab.* 20(4):1805–1842.
9. Janssen, A. 2000. Invariance principles for sums of extreme sequential order statistics attracted to Levy processes. *Stoch. Proc. Appl.* 85:255–277.
10. Ben Rached, N., Z. Botev, A. Kammoun, M.-S. Alouini, and R. Tempone. 2018. On the sum of order statistics and applications to wireless communication systems performances. *IEEE T. Wirel. Commun.* 17(11):7801–7813. doi: 10.1109/twc.2018.2871201.
11. Korolev, V. Yu. 1994. Convergence of random sequences with the independent random indices. I. *Theor. Probab. Appl.* 39(2):282–297.
12. Korolev, V. Yu. 1995. Convergence of random sequences with independent random indices. II. *Theor. Probab. Appl.* 40(4):770–772.

Received November 28, 2019

## Contributor

**Korolev Victor Yu.** (b. 1954) — Doctor of Science in physics and mathematics, professor, head of department, Faculty of Computational Mathematics and Cybernetics, and principal scientist, Moscow Center for Fundamental and Applied Mathematics, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilova Str., Moscow 119333, Russian Federation; vkorolev@cs.msu.ru

# О СТАТИСТИЧЕСКИХ СВОЙСТВАХ ОЦЕНКИ РИСКА В ЗАДАЧЕ ОБРАЩЕНИЯ ПРЕОБРАЗОВАНИЯ РАДОНА ПРИ СЛУЧАЙНОМ ОБЪЕМЕ ПРОЕКЦИОННЫХ ДАННЫХ\*

О. В. Шестаков<sup>1</sup>

**Аннотация:** При реконструкции томографических изображений необходимо решать задачу подавления шума, возникающего при регистрации проекционных данных. Методы решения этой задачи, основанные на вейвлет-алгоритмах и процедурах пороговой обработки, обладают рядом преимуществ, включая вычислительную эффективность и возможность адаптации к локальным особенностям изображений. Анализ погрешностей этих методов представляет собой важную практическую задачу, поскольку дает возможность оценить качество как самих методов, так и используемого оборудования. При использовании методов пороговой обработки обычно предполагается, что число коэффициентов разложения фиксировано, а распределение шума гауссово. Эта модель хорошо изучена в литературе, и для разных классов функций вычислены оптимальные значения порогов. Однако в некоторых ситуациях объем выборки заранее не фиксирован и его приходится моделировать некоторой случайной величиной. В данной работе рассматривается модель со случайным числом наблюдений и исследуются асимптотические свойства оценки среднеквадратичного риска. Доказывается, что предельное распределение этой оценки принадлежит классу сдвиг-масштабных смесей нормальных законов.

**Ключевые слова:** пороговая обработка; случайный объем выборки; преобразование Радона; оценка среднеквадратичного риска

**DOI:** 10.14357/19922264200306

## 1 Введение

Задачи реконструкции изображений по проекциям возникают во многих прикладных областях, включая медицину, биологию, астрономию, физику плазмы и многие другие. Большинство математических методов реконструкции основано на обращении так называемого преобразования Радона, которое описывает проекционные данные.

Преобразование Радона является линейным и однородным, и поэтому для его обращения оказываются применимыми методы вейвлет-анализа и пороговой обработки, позволяющие эффективно подавлять шум, возникающий при регистрации проекционных данных. Эти методы получили развитие в работах [1–3], в которых оценивается порядок среднеквадратичного риска и описываются алгоритмы вычисления оптимальных параметров пороговой обработки. Также изучены статистические свойства оценки среднеквадратичного риска, позволяющей судить о качестве реконструкции на основе только наблюдаемых данных. Показано, что при определенных условиях она является сильно состоятельной и асимптотически нормальной [4].

В некоторых случаях объем данных, доступных для анализа, заранее не известен. Такие ситуации могут возникать, например, в случае пропуска данных или недостатке информации о характеристиках используемого оборудования. В таком случае предполагается, что объем выборки данных представляет собой случайную величину с некоторым заданным распределением.

В данной работе рассматривается модель со случайным числом «загрязненных» белым гауссовым шумом коэффициентов разложения функции, описывающей проекционные данные томографического изображения. Показано, что предельное распределение оценки среднеквадратичного риска уже не обязано быть нормальным.

## 2 Обращение преобразования Радона с помощью метода вейвлет–вейглет-разложения

Математическая модель задачи реконструкции томографических изображений основана на так называемом преобразовании Радона. Обозначим

\* Работа выполнена при финансовой поддержке РФФИ (проект 19-07-00352) и в соответствии с программой Московского центра фундаментальной и прикладной математики.

<sup>1</sup> Московский государственный университет им. М. В. Ломоносова, кафедра математической статистики факультета вычислительной математики и кибернетики; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, oshestakov@cs.msu.ru

через  $\text{Lip}(\gamma, L)$  класс равномерно регулярных по Липшицу функций, где  $\gamma$  — показатель, а  $L$  — константа Липшица [5]. Пусть изображение описывается функцией  $f(x, y) \in \text{Lip}(\gamma, L)$  с компактным носителем (без ограничения общности будем считать, что это круг единичного радиуса с центром в начале координат) и некоторым показателем  $\gamma > 0$ . Преобразованием Радона функции  $f$  называется преобразование вида:

$$\text{Rf}(s, \theta) = \int_{L_{s, \theta}} f(x, y) dl,$$

где

$$L_{s, \theta} = \{(x, y) : x \cos \theta + y \sin \theta - s = 0\}.$$

Одним из возможных методов обращения преобразования  $\text{Rf}$  служит метод вейвлет–вейглет–разложения [2].

Пусть заданы  $\phi(x)$  и  $\psi(x)$  — отцовская и материнская вейвлет-функции. Определим

$$\left. \begin{aligned} \psi_{j, k_1, k_2}^{[1]}(x, y) &= 2^j \phi(2^j x - k_1) \psi(2^j y - k_2); \\ \psi_{j, k_1, k_2}^{[2]}(x, y) &= 2^j \psi(2^j x - k_1) \phi(2^j y - k_2); \\ \psi_{j, k_1, k_2}^{[3]}(x, y) &= 2^j \psi(2^j x - k_1) \psi(2^j y - k_2). \end{aligned} \right\} (1)$$

Семейство  $\{\psi_{j, k_1, k_2}^{[\lambda]}\}$  образует ортонормированный базис в  $L^2(\mathbb{R}^2)$ . Индекс  $j$  в (1) называется масштабом, а индексы  $k_1$  и  $k_2$  — сдвигами. В данной работе предполагается, что используются вейвлеты Мейера [5], имеющие  $M$  непрерывных производных ( $M \geq \gamma$ ).

Вейвлет-разложение функции  $f$  имеет вид:

$$f = \sum_{\lambda, j, k_1, k_2} \langle f, \psi_{j, k_1, k_2}^{[\lambda]} \rangle \psi_{j, k_1, k_2}^{[\lambda]}.$$

Определим функции  $\xi_{j, k_1, k_2}^{[\lambda]}(s, \theta)$ :

$$\xi_{j, k_1, k_2}^{[\lambda]}(s, \theta) = \frac{2^{-j/2}}{4\pi} I^{-1} \left[ R \psi_{j, k_1, k_2}^{[\lambda]} \right](s, \theta),$$

где  $I^\alpha$  — потенциал Рисса, определяемый в пространстве Фурье по формуле  $\widehat{I^\alpha g}(\omega) = |\omega|^{-\alpha} \widehat{g}(\omega)$ . Эти функции называются «вейглетами» [2]. Для них справедливо соотношение

$$\langle f, \psi_{j, k_1, k_2}^{[\lambda]} \rangle = 2^{j/2} \langle \text{Rf}, \xi_{j, k_1, k_2}^{[\lambda]} \rangle.$$

Последовательность  $\{\xi_{j, k_1, k_2}^{[\lambda]}\}$  образует устойчивый базис [3], и вейвлет-вейглет–разложение  $f$  имеет вид:

$$f = \sum_{\lambda, j, k_1, k_2} 2^{j/2} \langle \text{Rf}, \xi_{j, k_1, k_2}^{[\lambda]} \rangle \psi_{j, k_1, k_2}^{[\lambda]}. \quad (2)$$

В разложении (2) используются только проекционные данные, и оно служит основой метода реконструкции.

### 3 Метод подавления шума, основанный на пороговой обработке

В практических задачах изображение описывается дискретными отсчетами некоторой функции. При этом, поскольку носителем функции выступает единичный круг,  $(s, \theta) \in [-1, 1] \times [0, \pi]$  и модель зашумленных проекционных данных, рассматриваемая в данной работе, выглядит следующим образом:

$$Y_{i, j} = \text{Rf} \left( -1 + \frac{2i}{N}, \frac{j\pi}{N} \right) + e_{i, j},$$

$$i = 1, \dots, N, \quad j = 1, \dots, N,$$

где  $N = 2^J$  для некоторого  $J > 0$ . Предполагается, что  $e_{i, j}$  независимы и имеют нормальное распределение с нулевым средним и дисперсией  $\sigma^2$ . Тогда в дискретном аналоге вейвлет–вейглет–разложения коэффициенты описываются моделью [6]:

$$X_{j, k_1, k_2}^{[\lambda]} = \mu_{j, k_1, k_2}^{[\lambda]} + e_{j, k_1, k_2}^{[\lambda]}, \quad (3)$$

где

$$\mu_{j, k_1, k_2}^{[\lambda]} = 2^J \langle \text{Rf}, \xi_{j, k_1, k_2}^{[\lambda]} \rangle;$$

$e_{j, k_1, k_2}^{[\lambda]}$  имеют нормальное распределение с нулевым средним и дисперсией  $\sigma_\lambda^2$  ( $\lambda = 1, 2, 3$ ). Эти коэффициенты уже не являются независимыми. Значение  $\sigma_\lambda^2$  зависит от выбранного вейвлет-базиса и  $\lambda$ , но не зависит от  $k_1, k_2$  и  $j$ .

Для подавления шума и построения оценок коэффициентов  $\mu_{j, k_1, k_2}^{[\lambda]}$ , как правило, используется пороговая обработка зашумленных коэффициентов модели (3). Смысл этой обработки заключается в удалении достаточно маленьких коэффициентов, которые считаются шумом. Оценки  $\mu_{j, k_1, k_2}^{[\lambda]}$  вычисляются с помощью функции пороговой обработки  $\rho(x, T_\lambda)$  с некоторым порогом  $T_\lambda$ :

$$\widehat{\mu}_{j, k_1, k_2}^{[\lambda]} = \rho \left( X_{j, k_1, k_2}^{[\lambda]}, T_\lambda \right).$$

В данной работе предполагается, что используется функция жесткой пороговой обработки  $\rho_H(x, T_\lambda) = x \mathbf{1}(|x| > T_\lambda)$  или мягкой пороговой обработки  $\rho_S(x, T_\lambda) = \text{sgn}(x) (|x| - T_\lambda)_+$ .

Среднеквадратичный риск метода пороговой обработки определяется по формуле:

$$r_J = \sum_{j=0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j E \left( \widehat{\mu}_{j,k_1,k_2}^{[\lambda]} - \mu_{j,k_1,k_2}^{[\lambda]} \right)^2. \quad (4)$$

Выражение (4) можно использовать только для вычисления величины риска в «модельных» ситуациях (при известной функции изображения) для тестирования оборудования, поскольку оно содержит значения «чистых» коэффициентов  $\mu_{j,k_1,k_2}^{[\lambda]}$ . В случае когда функция сигнала неизвестна, можно построить статистическую оценку  $r_J$ . Если в слагаемом  $|X_{j,k_1,k_2}^{[\lambda]}| > T$ , то вклад этого слагаемого в риск составляет  $2^j \sigma^2$  в случае жесткой и  $2^j(\sigma^2 + T^2)$  в случае мягкой пороговой обработки, а если  $|X_{j,k_1,k_2}^{[\lambda]}| \leq T$ , то вклад составляет  $2^j(\mu_{j,k_1,k_2}^{[\lambda]})^2$  в обоих случаях. Неизвестную величину  $(\mu_{j,k_1,k_2}^{[\lambda]})^2$  можно оценить разностью  $(X_{j,k_1,k_2}^{[\lambda]})^2 - \sigma^2$ , поскольку

$$E \left( X_{j,k_1,k_2}^{[\lambda]} \right)^2 = \sigma^2 + \left( \mu_{j,k_1,k_2}^{[\lambda]} \right)^2.$$

Таким образом, в качестве оценки среднеквадратичного риска можно использовать величину:

$$\widehat{r}_J = \sum_{j=0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j F \left[ X_{j,k_1,k_2}^{[\lambda]} \right],$$

где

$$F \left[ X_{j,k_1,k_2}^{[\lambda]} \right] = \begin{cases} \left( \left( X_{j,k_1,k_2}^{[\lambda]} \right)^2 - \sigma^2 \right) \mathbf{1} \left( \left| X_{j,k_1,k_2}^{[\lambda]} \right| \leq T \right) + \\ + \sigma^2 \mathbf{1} \left( \left| X_{j,k_1,k_2}^{[\lambda]} \right| > T \right) & \text{в случае} \\ \text{жесткой пороговой обработки;} \\ \left( \left( X_{j,k_1,k_2}^{[\lambda]} \right)^2 - \sigma^2 \right) \mathbf{1} \left( \left| X_{j,k_1,k_2}^{[\lambda]} \right| \leq T \right) + \\ + \left( \sigma^2 + T^2 \right) \mathbf{1} \left( \left| X_{j,k_1,k_2}^{[\lambda]} \right| > T \right) & \text{в случае} \\ \text{мягкой пороговой обработки.} \end{cases}$$

При мягкой пороговой обработке эта оценка оказывается несмещенной [5]. Кроме того, справедливо утверждение об асимптотической нормальности  $\widehat{r}_J$  [4].

**Теорема 1.** Если  $f \in \text{Lip}(\gamma, L)$  с  $\gamma > 1/2$ , то при выборе асимптотически оптимального порога

$$P \left( \frac{\widehat{r}_J - r_J}{D_J} < x \right) \rightarrow \Phi(x) \text{ при } J \rightarrow \infty,$$

где  $\Phi(x)$  — функция распределения стандартного нормального закона, а  $D_J = C_R 2^{2J}$ , где константа  $C_R$  зависит от используемого базиса. Метод вычисления этой константы описан в [4].

Асимптотически оптимальный порог в теореме 1 при  $J \rightarrow \infty$  удовлетворяет соотношению [7]:

$$T \simeq \sigma \sqrt{\frac{6\gamma + 3}{2\gamma + 3} \ln 2^{2J}}.$$

В следующем разделе показано, что в модели со случайным числом вейвлет-коэффициентов предельное распределение оценки риска уже не обязательно быть нормальным.

## 4 Случайное число коэффициентов разложения

Пусть  $M$  — положительная целочисленная случайная величина и  $N = 2^M$ . Тогда оценка среднеквадратичного риска принимает вид:

$$\widehat{R}_M = \sum_{j=0}^{M-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^3 2^j F \left[ X_{j,k_1,k_2}^{[\lambda]} \right]. \quad (5)$$

Чтобы получились осмысленные асимптотические результаты для оценки (5), величина  $M$  должна быть «большой». Рассмотрим последовательность  $M_J, J = 1, \dots$ , и исследуем поведение (5) при  $J \rightarrow \infty$ .

**Теорема 2.** Пусть выполнены условия теоремы 1. Пусть также  $M_J \xrightarrow{P} \infty$  и для некоторой последовательности  $R_J$  совместное распределение пар

$$(U_J, V_J) \equiv \left( \frac{2^{4M_J}}{2^{4J}}, \frac{r_{M_J} - R_J}{D_J} \right)$$

сходится к распределению некоторого случайного вектора  $(U, V)$  при  $J \rightarrow \infty$ . Тогда

$$P \left( \frac{\widehat{R}_{M_J} - R_J}{D_J} < x \right) \rightarrow H(x) \text{ при } J \rightarrow \infty, \quad (6)$$

где  $H(x)$  — функция распределения, соответствующая характеристической функции  $h(t) = E \varphi(t\sqrt{U}) e^{itV}$  ( $\varphi(t)$  — характеристическая функция стандартного нормального закона).

Таким образом, предельные распределения в (6) представляют собой сдвиг-масштабные смеси нормальных законов.

**Доказательство.** Запишем

$$\frac{\widehat{R}_{M_J} - R_J}{D_J} = \frac{D_{M_J}}{D_J} \frac{\widehat{R}_{M_J} - r_{M_J}}{D_{M_J}} + \frac{r_{M_J} - R_J}{D_J}.$$

Тогда

$$\frac{D_{M_J}}{D_J} = \frac{2^{2M_J}}{2^{2J}}$$

и характеристическая функция  $h_{J,J}(t)$  случайной величины  $(\widehat{R}_{M_J} - R_J)/D_J$  запишется в виде:

$$h_J(t) = \sum_{j=1}^{\infty} P(M_J = j) \exp\left(it \frac{r_j - R_J}{D_J}\right) \varphi_j\left(t \frac{2^{2j}}{2^{2J}}\right),$$

где  $\varphi_j(t) = \varphi((\hat{r}_j - r_j)/D_j)$  — характеристическая функция  $(\hat{r}_j - r_j)/D_j$ . Обозначим

$$g_J(t) = E\varphi\left(t\sqrt{U_J}\right) e^{itV_J} = \sum_{j=1}^{\infty} P(M_J = j) \exp\left(it \frac{r_j - R_J}{D_J}\right) \varphi\left(t \frac{2^{2j}}{2^{2J}}\right).$$

Имеем

$$|h_J(t) - g_J(t)| = \left| \sum_{j=1}^{\infty} P(M_J = j) \times \exp\left(it \frac{r_j - R_J}{D_J}\right) \left(\varphi_j\left(t \frac{2^{2j}}{2^{2J}}\right) - \varphi\left(t \frac{2^{2j}}{2^{2J}}\right)\right) \right|.$$

Фиксируем  $t \neq 0$ . Рассуждая, как в работе [8], для произвольного  $\varepsilon > 0$  можно выбрать такое число  $\kappa > 0$ , что

$$\sum_{j > \kappa + J} P(M_J = j) = P(U_J > 2^{4\kappa}) < \varepsilon$$

и начиная с некоторого  $J$ , зависящего от  $\varepsilon$  и  $\kappa$ ,

$$\left| \sum_{j > \kappa + J} P(M_J = j) \exp\left(it \frac{r_j - R_J}{D_J}\right) \times \left(\varphi_j\left(t \frac{2^{2j}}{2^{2J}}\right) - \varphi\left(t \frac{2^{2j}}{2^{2J}}\right)\right) \right| < 2\varepsilon.$$

Далее в силу теоремы 1

$$\begin{aligned} & \left| \sum_{j \leq \kappa + J} P(M_J = j) \exp\left(it \frac{r_j - R_J}{D_J}\right) \times \left(\varphi_j\left(t \frac{2^{2j}}{2^{2J}}\right) - \varphi\left(t \frac{2^{2j}}{2^{2J}}\right)\right) \right| \leq \\ & \leq \sum_{j \leq \kappa + J} P(M_J = j) \sup_{\tau \leq 2^{2\kappa} t} |\varphi_j(\tau) - \varphi(\tau)| < \\ & < E \sup_{\tau \leq 2^{2\kappa} t} |\varphi_{M_J}(\tau) - \varphi(\tau)| < \varepsilon, \end{aligned}$$

так как  $M_J \xrightarrow{P} \infty$  [8].

Таким образом, для произвольного  $t$

$$|h_J(t) - g_J(t)| \rightarrow 0 \text{ при } J \rightarrow \infty.$$

Далее при любом фиксированном  $t$

$$|g_J(t) - h(t)| = \left| E\varphi\left(t\sqrt{U_J}\right) e^{itV_J} - E\varphi\left(t\sqrt{U}\right) e^{itV} \right| \rightarrow 0 \text{ при } J \rightarrow \infty,$$

поскольку функция  $q_t(u, v) = \varphi(t\sqrt{u})e^{itv}$  непрерывна и ограничена, и  $(U_J, V_J) \Rightarrow (U, V)$  при  $J \rightarrow \infty$ . Следовательно,

$$|h_J(t) - h(t)| \rightarrow 0 \text{ при } J \rightarrow \infty$$

и имеет место (6). Теорема доказана.

**Замечание.** Распределения, получаемые в качестве предельных в теореме 2, могут существенно отличаться по своим свойствам от нормального закона. В частности, они могут иметь гораздо более тяжелые хвосты. Кроме того, порядок теоретического среднеквадратичного риска в модели со случайным числом вейвлет-коэффициентов может оказаться существенно больше порядка риска в модели с фиксированным числом коэффициентов [9].

## Литература

1. *Kolaczyk E. D.* Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noisy data. — Stanford, CA, USA: Stanford University, 1994. PhD Diss. 152 p.
2. *Donoho D.* Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition // Appl. Comput. Harmon. A., 1995. Vol. 2. P. 101–126.
3. *Lee N.* Wavelet-vaguelette decompositions and homogeneous equations. — West Lafayette, IN, USA: Purdue University, 1997. PhD Diss. 103 p.
4. *Шестаков О. В.* Вероятностно-статистические методы анализа и обработки сигналов на основе вейвлет-алгоритмов. — М.: Аргмак-Медиа, 2016. 200 с.
5. *Mallat S.* A wavelet tour of signal processing. — New York, NY, USA: Academic Press, 1999. 857 p.
6. *Ерошенко А. А., Шестаков О. В.* Асимптотические свойства оценки риска в задаче восстановления изображения с коррелированным шумом при обращении преобразования Радона // Информатика и её применения, 2014. Т. 8. Вып. 4. С. 32–40.
7. *Jansen M.* Noise reduction by wavelet thresholding. — Lecture notes in statistics ser. — New York, NY, USA: Springer-Verlag, 2001. Vol. 161. 217 p.
8. *Korolev V. Yu., Zeifman A. I.* On convergence of the distributions of random sequences with independent random indexes to variance-mean mixtures // Stoch. Models, 2016. Vol. 32. Iss. 3. P. 414–432.
9. *Shestakov O. V.* Mean-square risk of the threshold processing in the problem of inverting the Radon transform with a random sample size // J. Math. Sci., 2020. Vol. 248. No. 1. P. 46–50.

Поступила в редакцию 01.07.20

# ON THE STATISTICAL PROPERTIES OF RISK ESTIMATE IN THE PROBLEM OF INVERTING THE RADON TRANSFORM WITH A RANDOM VOLUME OF PROJECTION DATA

O. V. Shestakov<sup>1,2</sup>

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

<sup>2</sup>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

**Abstract:** When reconstructing tomographic images, it is necessary to solve the problem of suppressing the noise arising from registration of projection data. Methods for solving this problem based on wavelet algorithms and threshold processing procedures have several advantages, including computational efficiency and ability to adapt to local features of images. An analysis of the errors of these methods is an important practical task, since it makes it possible to evaluate the quality of both the methods themselves and the equipment used. When using threshold processing procedures, it is usually assumed that the number of decomposition coefficients is fixed and the noise distribution is Gaussian. This model has been well studied in the literature, and the optimal threshold values have been calculated for different classes of functions. However, in some situations, the sample size is not fixed in advance and must be modeled with some random variable. This paper considers a model with a random number of observations and investigates the asymptotic properties of the mean-square risk estimate. It is proved that the limiting distribution of this estimate belongs to the class of shift-scale mixtures of normal laws.

**Keywords:** threshold processing; random sample size; Radon transform; grid; mean-square risk estimate

**DOI:** 10.14357/19922264200306

## Acknowledgments

The work is supported by the Russian Foundation for Basic Research (project No. 19-07-00352). The research was conducted in accordance with the Program of Moscow Center for Fundamental and Applied Mathematics.

## References

1. Kolaczyk, E. D. 1994. Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noisy data. Stanford, CA: Stanford University. PhD Diss. 152 p.
2. Donoho, D. 1995. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. A* 2:101–126.
3. Lee, N. 1997. Wavelet-vaguelette decompositions and homogeneous equations. West Lafayette, IN: Purdue University. PhD Diss. 103 p.
4. Shestakov, O. V. 2016. *Veroyatnostno-statisticheskie metody analiza i obrabotki signalov na osnove veyvlet-algoritmov* [Probabilistic-statistical methods of signal analysis and processing based on wavelet algorithms]. Moscow: ARGAMAK-MEDIA. 200 p.
5. Mallat, S. 1999. *A wavelet tour of signal processing*. New York, NY: Academic Press. 857 p.
6. Eroshenko, A. A., and O. V. Shestakov. 2014. Asimptoticheskie svoystva otsenki riska v zadache vosstanovleniya izobrazheniya s korrelirovannym shumom pri obrashchenii preobrazovaniya Radona [Asymptotic properties of risk estimate in the problem of reconstructing images with correlated noise by inverting the Radon transform]. *Informatika i ee Primeneniya — Inform. Appl.* 8(4):32–40.
7. Jansen, M. 2001. *Noise reduction by wavelet thresholding*. Lecture notes in statistics ser. New York, NY: Springer-Verlag. 217 p.
8. Korolev, V. Yu., and A. I. Zeifman. 2016. On convergence of the distributions of random sequences with independent random indexes to variance-mean mixtures. *Stoch. Models* 32(3):414–432.
9. Shestakov, O. V. 2020. Mean-square risk of the threshold processing in the problem of inverting the Radon transform with a random sample size. *J. Math. Sci.* 248(1):46–50.

Received July 1, 2020

## Contributor

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

# МЕТОД ЛОГАРИФМИЧЕСКИХ МОМЕНТОВ ДЛЯ ОЦЕНИВАНИЯ ПАРАМЕТРОВ ГАММА-ЭКСПОНЕНЦИАЛЬНОГО РАСПРЕДЕЛЕНИЯ\*

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

**Аннотация:** Рассматривается модифицированный метод моментов для оценивания параметров гамма-экспоненциального распределения. Обосновывается сильная состоятельность полученных оценок. Гамма-экспоненциальное распределение представляет собой удобный механизм при моделировании процессов и явлений с помощью масштабных смесей обобщенных гамма-распределений. Подобные задачи возникают во многих естественно-научных областях в предположении о рандомизированности рассматриваемых параметров и могут быть единообразно описаны в терминах байесовских моделей баланса. Полученные результаты могут применяться в широком классе задач, использующих для моделирования распределения с положительным неограниченным носителем, без дополнительных предположений о представлении исследуемого объекта в терминах масштабной смеси, ввиду большого разнообразия видов плотности пятипараметрического гамма-экспоненциального распределения.

**Ключевые слова:** оценивание параметров; гамма-экспоненциальное распределение; смешанные распределения; обобщенное гамма-распределение; метод моментов; состоятельная оценка

**DOI:** 10.14357/19922264200307

## 1 Введение

В различных областях жизнедеятельности присутствуют характеристики систем, выражаемые при помощи отношения двух величин.

Так, в демографии уровень младенческой смертности определяется как отношение числа умерших в возрасте до года к числу родившихся за некоторый период времени [1], а индекс разводимости — как отношение коэффициента суммарной разводимости к коэффициенту суммарной брачности [2]; в физике коэффициент трансформации — это отношение выходного напряжения к входному, а универсальная функция Кирхгофа — отношение излучательной способности тела к поглощательной [3]; в теории массового обслуживания отношение интенсивности входящего потока к интенсивности обслуживания определяет коэффициент загрузки системы [4]; при моделировании чрезвычайных ситуаций пожароопасность объекта определяется отношением угрозы возникновения пожара к фактору пожарозащиты [5]; в теории надежности ожидаемое время безотказной работы представимо в виде отношения среднего времени безотказной работы к среднему времени восстановления [6].

Ряд примеров можно продолжить. Каждую из таких характеристик можно рассматривать как индекс баланса системы [7]. Применение рандомизационного байесовского подхода к описанным моделям дает возможность изучать характеристики индекса баланса как масштабной смеси вероятностных законов.

В случае когда для описания процессов и явлений используются распределения с неограниченным неотрицательным носителем, широкую популярность приобрел гамма-класс распределений. В работе [8] было описано так называемое гамма-экспоненциальное распределение, в терминах которого, в частности, выражаются характеристики масштабных смесей обобщенных гамма-распределений, применяемых в байесовских моделях баланса [7].

На практике исследователь имеет дело с наблюдаемыми величинами, отражающими проявление анализируемого реального процесса, в отношении которых делаются некоторые модельные предположения о виде их распределения. При этом предполагается, что истинное значение параметров распределения неизвестно и требует оценки. Заметим, что нередко при моделировании используют

\* Работа выполнена при частичной финансовой поддержке РФФИ (проект 20-07-00655) и в соответствии с программой Московского центра фундаментальной и прикладной математики.

<sup>1</sup> Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, pubigena@mail.ru

<sup>2</sup> Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, oshestakov@cs.msu.su

ся частные случаи обобщенного гамма-распределения, такие как экспоненциальное распределение;  $\chi^2$ -распределение; распределение Эрланга; гамма-распределение; полунормальное распределение, или распределение максимума процесса броуновского движения; распределение Рэлея; распределение Максвелла–Больцмана;  $\chi$ -распределение;  $m$ -распределение Накагами; распределение Вильсона–Хильферти; распределение Вейбулла–Гнеденко и многие другие, включая масштабированные и обратные аналоги перечисленных.

Дадим необходимые определения. Обозначим обобщенное гамма-распределение, или распределение Крицкого–Менкеля [9, 10], с плотностью

$$f(x) = \frac{|v|x^{vq-1}e^{-(x/\theta)^v}}{\theta^{vq}\Gamma(q)}, \quad v \neq 0, \quad q > 0, \quad \theta > 0, \quad x > 0,$$

через  $GG(v, q, \theta)$ .

**Определение 1.** Назовем функцию вида

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

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

гамма-экспоненциальной функцией [11].

Функция (1) обобщает на случай  $\beta \neq 1$  преобразование, введенное Леруа [12] для исследования производящих функций специального вида. Кроме того, функцию (1) можно рассматривать (при некоторых допущениях) как частный случай функции Сривастава–Томовски [13], обобщающей функцию Миттаг–Леффлера [14].

**Определение 2.** Будем говорить, что случайная величина  $\zeta$  имеет гамма-экспоненциальное распределение  $GE(r, \nu, s, t, \delta)$  с параметрами изгиба  $0 \leq r < 1$ , формы  $\nu \neq 0$ , концентрации  $s, t > 0$  и масштаба  $\delta > 0$ , если ее плотность при  $x > 0$  задается соотношением:

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

где  $E = (r, \nu, s, t, \delta)$  [8].

В работе [8] было показано, что гамма-экспоненциальное распределение обладает следующими свойствами.

**Лемма 1.**

1. Пусть независимые случайные величины  $\lambda$  и  $\mu$  имеют соответственно распределения  $GG(v, q, \theta)$  и  $GG(u, p, \alpha)$ ,  $uv > 0$ . Тогда распределение  $\lambda$  совпадает с  $GE(0, v, \cdot, q, \theta)$ ; распределение  $\lambda/\mu$  при  $|u| > |v|$  совпадает с  $GE(v/u, v, p, q, \theta/\alpha)$ ; распределение  $\lambda/\mu$  при  $|v| > |u|$  совпадает с  $GE(u/v, -u, q, p, \theta/\alpha)$ .

2. При  $0 < r < 1$  плотность  $g_E(x)$ ,  $E = (r, \nu, s, t, \delta)$ , совпадает с плотностью отношения независимых случайных величин, имеющих обобщенные гамма-распределения  $GG(\nu, t, \delta)$  и  $GG(\nu/r, s, 1)$ .

В случае моделирования реального процесса при помощи гамма-экспоненциального распределения неизбежно возникает вопрос оценивания неизвестных параметров по реальным данным. Ввиду представления плотности  $g_E(x)$  в терминах специальной гамма-экспоненциальной функции (1) метод максимального правдоподобия представляется затруднительным. То же можно сказать и о прямом методе моментов. По этой причине предлагается оценивать параметры гамма-экспоненциального распределения при помощи модифицированного метода, основанного на логарифмических моментах.

## 2 Логарифмические моменты гамма-экспоненциального распределения

Введем обозначение для дигамма-функции:  $\psi(z) = \Gamma'(z)/\Gamma(z)$ .

Найдем первые два логарифмических момента случайной величины  $\zeta \sim GE(r, \nu, s, t, \delta)$ . Рассмотрим преобразование Меллина

$$\mathcal{M}_\zeta(z) = \int_0^\infty x^z dF_\zeta(x), \quad z \in \mathbb{C},$$

распределения  $\zeta$ . Воспользуемся леммой 1 и представлением  $\zeta \stackrel{d}{=} \lambda/\mu$ , где независимые случайные величины  $\lambda$  и  $\mu$  имеют соответственно распределения  $GG(\nu, t, \delta)$  и  $GG(\nu/r, s, 1)$ . Поскольку для  $\lambda \sim GG(\nu, t, \delta)$  преобразование Меллина имеет вид:

$$\mathcal{M}_\lambda(z) = \frac{\delta^z}{\Gamma(t)} \Gamma\left(t + \frac{z}{\nu}\right), \quad t + \frac{\operatorname{Re}(z)}{\nu} > 0,$$

для отношения  $\lambda \sim GG(\nu, t, \delta)$  к  $\mu \sim GG(\nu/r, s, 1)$  выполнено

$$\mathcal{M}_{\lambda/\mu}(z) = \frac{\delta^z}{\Gamma(t)\Gamma(s)} \Gamma\left(t + \frac{z}{\nu}\right) \Gamma\left(s - \frac{rz}{\nu}\right),$$

$$t + \frac{\operatorname{Re}(z)}{\nu} > 0, \quad s - \frac{r\operatorname{Re}(z)}{\nu} > 0,$$

откуда получаем вид моментов  $\zeta \stackrel{d}{=} \lambda/\mu$ :

$$E\zeta^k = \frac{\delta^k}{\Gamma(t)\Gamma(s)} \Gamma\left(t + \frac{k}{\nu}\right) \Gamma\left(s - \frac{rk}{\nu}\right),$$

$$t + \frac{k}{\nu} > 0, \quad s - \frac{rk}{\nu} > 0, \quad (2)$$

и характеристической функции логарифма  $\zeta$ :

$$E e^{iy \ln \zeta} = \frac{\delta^{iy}}{\Gamma(t)\Gamma(s)} \Gamma\left(t + \frac{iy}{\nu}\right) \Gamma\left(s - \frac{iry}{\nu}\right),$$

$$y \in \mathbb{R}. \quad (3)$$

Дважды продифференцировав соотношение (3), получаем:

$$E \ln \zeta = \frac{\nu \ln \delta + \psi(t) - r\psi(s)}{\nu};$$

$$E \ln^2 \zeta = \frac{[\nu \ln \delta + \psi(t) - r\psi(s)]^2}{\nu^2} + \frac{\psi'(t)}{\nu^2} + \frac{r^2 \psi'(s)}{\nu^2}.$$

### 3 Оценивание параметров изгиба и масштаба

Зафиксируем параметры формы  $\nu$  и концентрации  $t$  и  $s$ . Заметим, что, поскольку функция, стоящая в правой части (2), определена не для всех значений параметров и необратима, использовать классический метод моментов напрямую не представляется возможным. Оценим при помощи модифицированного метода моментов параметры изгиба  $r$  и масштаба  $\delta$  распределения  $GE(r, \nu, s, t, \delta)$  с учетом соотношения  $\zeta \stackrel{d}{=} \lambda/\mu$ , где независимые случайные величины  $\lambda$  и  $\mu$  имеют соответственно распределения  $GG(\nu, t, \delta)$  и  $GG(\nu/r, s, 1)$ .

Введем обозначение для выборочных логарифмических моментов  $\zeta$ :

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

где  $X = (X_1, \dots, X_n)$  — выборка из распределения  $\zeta$ .

Составим систему из двух уравнений с двумя неизвестными  $r$  и  $\delta$ :

$$E \ln \zeta = L_1(X); \quad E \ln^2 \zeta = L_2(X).$$

Очевидно, что исходная система уравнений эквивалентна следующей:

$$\frac{\nu \ln \delta + \psi(t) - r\psi(s)}{\nu} = L_1(X); \quad (4)$$

$$\frac{\psi'(t) + r^2 \psi'(s)}{\nu^2} = L_2(X) - L_1^2(X). \quad (5)$$

Из последнего уравнения с учетом  $0 \leq r < 1$  получаем, что оценка параметра изгиба имеет вид:

$$\hat{r}(X) = \sqrt{\frac{\nu^2(L_2(X) - L_1^2(X)) - \psi'(t)}{\psi'(s)}}, \quad (6)$$

откуда следует вид оценки параметра масштаба:

$$\hat{\delta}_r(X) = \exp \left\{ L_1(X) - \frac{\psi(t)}{\nu} + \right.$$

$$\left. + \psi(s) \sqrt{\frac{\nu^2(L_2(X) - L_1^2(X)) - \psi'(t)}{\nu^2 \psi'(s)}} \right\}. \quad (7)$$

Из усиленного закона больших чисел и непрерывности функций, стоящих в правых частях (6) и (7), следует свойство полученных оценок.

**Теорема 1.** При фиксированных параметрах  $\nu$ ,  $t$  и  $s$  распределения  $GE(r, \nu, s, t, \delta)$  оценки  $\hat{r}(X)$  и  $\hat{\delta}_r(X)$  параметров  $r$  и  $\delta$  обладают свойством сильной состоятельности.

### 4 Оценивание параметров формы и масштаба

Зафиксируем параметры изгиба  $r$  и концентрации  $t$  и  $s$ . Оценим при помощи модифицированного метода моментов параметры формы  $\nu$  и масштаба  $\delta$  распределения  $GE(r, \nu, s, t, \delta)$ .

Проводя рассуждения, аналогичные приведенным в предыдущем разделе, можно заметить, что двух уравнений (4) и (5) недостаточно для решения поставленной задачи ввиду невозможности определения знака параметра  $\nu$  исходя из равенства

$$\nu^2 = \frac{\psi'(t) + r^2 \psi'(s)}{L_2(X) - L_1^2(X)}.$$

Рассмотрим дополнительное уравнение:

$$E \ln^3 \zeta = L_3(X).$$

Поскольку выполнено (3), имеем:

$$E \ln^3 \zeta = \frac{[\nu \ln \delta + \psi(t) - r\psi(s)]^3}{\nu^3} +$$

$$+ \frac{3[\nu \ln \delta + \psi(t) - r\psi(s)] \psi'(t)}{\nu^3} +$$

$$+ \frac{3r^2 [\nu \ln \delta + \psi(t) - r\psi(s)] \psi'(s)}{\nu^3} +$$

$$+ \frac{\psi''(t)}{\nu^3} - \frac{r^3 \psi''(s)}{\nu^3},$$

откуда с учетом (4) и (5) получаем

$$\frac{\psi''(t)}{\nu^3} - \frac{r^3 \psi''(s)}{\nu^3} = L_3(X) - 3L_1(X)L_2(X) + 2L_1^3(X).$$

Таким образом, оценка параметра формы  $\nu$  имеет вид:

$$\hat{\nu}(X) = \sqrt[3]{\frac{\psi''(t) - r^3\psi''(s)}{L_3(X) - 3L_1(X)L_2(X) + 2L_1^3(X)}}.$$

Из (4) следует вид оценки параметра масштаба  $\delta$ :

$$\hat{\delta}_\nu(X) = \exp \left\{ L_1(X) - (\psi(t) - r\psi(s)) \times \right. \\ \left. \times \sqrt[3]{\frac{L_3(X) - 3L_1(X)L_2(X) + 2L_1^3(X)}{\psi''(t) - r^3\psi''(s)}} \right\}.$$

Аналогично заключению разд. 3 справедливо следующее утверждение.

**Теорема 2.** При фиксированных параметрах  $r, t$  и  $s$  распределения  $GE(r, \nu, s, t, \delta)$  оценки  $\hat{\nu}(X)$  и  $\hat{\delta}_\nu(X)$  параметров  $\nu$  и  $\delta$  обладают свойством сильной состоятельности.

## 5 Заключение

При получении сильно состоятельных оценок двух пар параметров гамма-экспоненциального распределения, описанных в статье, активно использовалось представление гамма-экспоненциального распределения как масштабной смеси обобщенных гамма-распределений. При этом изначально предполагалось, что исследуемое распределение представляет собой удобный механизм анализа байесовских моделей баланса, в основе которых лежит представление целевого объекта как частного двух независимых случайных величин. Однако пятипараметрическое гамма-экспоненциальное распределение обладает широким набором возможных видов плотности и может оказаться полезным при изучении и других моделей, использующих для описания процессов и явлений распределения с неограниченным положительным носителем.

## Литература

1. Борисов В. А. Демография. — М.: NOTABENE, 2001. 272 с.
2. Волгин Н. А., Рыбаковский Л. Л., Калмыкова Н. М. и др. Демография. — М.: Логос, 2005. 280 с.
3. Кузнецов С. И., Рогозин К. И. Справочник по физике. — Томск: ТПУ, 2012. 224 с.
4. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. 529 с.
5. Шантала В. Г., Радоуцкий В. Ю., Шантала В. В. Основы моделирования чрезвычайных ситуаций. — Белгород: БГТУ, 2010. 166 с.
6. Здоровцов И. А., Королев В. Ю. Основы теории надежности волоконно-оптических линий передачи железнодорожного транспорта. — М.: МАКС Пресс, 2004. 308 с.
7. Кудрявцев А. А. Байесовские модели баланса // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 18–27.
8. Кудрявцев А. А. О представлении гамма-экспоненциального и обобщенного отрицательного биномиального распределений // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 78–82.
9. Крицкий С. Н., Менкель М. Ф. О приемах исследования случайных колебаний речного стока // Труды НИУ ГУГМС. Сер. IV, 1946. Вып. 29. С. 3–32.
10. Крицкий С. Н., Менкель М. Ф. Выбор кривых распределения вероятностей для расчетов речного стока // Известия АН СССР. Отд. техн. наук, 1948. № 6. С. 15–21.
11. Кудрявцев А. А., Тутова А. И. Гамма-экспоненциальная функция в байесовских моделях массового обслуживания // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 104–108.
12. Le Roy É. Sur les séries divergentes et les fonctions définies par un développement de Taylor // Ann. Faculté Sci. Toulouse 2 Ser., 1900. Vol. 2. No. 3. P. 317–384.
13. Srivastava H. M., Tomovski Ž. Fractional calculus with an integral operator containing a generalized Mittag-Leffler function in the kernel // Appl. Math. Comput., 2009. Vol. 211. P. 198–210.
14. Gorenflo R., Kilbas A. A., Mainardi F., Rogosin S. V. Mittag-Leffler functions, related topics and applications. — Berlin–Heidelberg: Springer-Verlag, 2014. 443 p.

Поступила в редакцию 04.07.20

# METHOD OF LOGARITHMIC MOMENTS FOR ESTIMATING THE GAMMA-EXPONENTIAL DISTRIBUTION PARAMETERS

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

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

<sup>2</sup>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

**Abstract:** The article discusses a modified method of moments for estimating the parameters of gamma-exponential distribution. The strong consistency of the estimates obtained is proved. Gamma-exponential distribution is a convenient mechanism for modeling the processes and phenomena using scale mixtures of generalized gamma distributions. Such problems arise in many fields of science under the assumption that the considered parameters are randomized and can be described in terms of Bayesian balance models. The obtained results can be applied in a wide class of problems that use for modeling the distribution with positive unlimited support, without additional assumptions about the representation of the studied object in terms of a scale mixture, due to the wide variety of density types of the five-parameter gamma-exponential distribution.

**Keywords:** parameter estimation; gamma-exponential distribution; mixed distributions; generalized gamma distribution; method of moments; consistent estimate

**DOI:** 10.14357/19922264200307

## Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 20-07-00655). The research was conducted in accordance with the Program of the Moscow Center for Fundamental and Applied Mathematics.

## References

1. Borisov, V. A. 2001. *Demografiya* [Demography]. Moscow: NOTABENE. 272 p.
2. Volgin, N. A., L. L. Rybakovskiy, N. M. Kalmykova, et al. 2005. *Demografiya* [Demography]. Moscow: Logos. 280 p.
3. Kuznetsov, S. I., and K. I. Rogozin. 2012. *Spravochnik po fizike* [Handbook of physics]. Tomsk: TPU. 224 p.
4. Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.
5. Shaptala, V. G., V. Yu. Radoutskiy, and V. V. Shaptala. 2010. *Osnovy modelirovaniya chrezvychaynykh situatsiy* [Basics of modeling of emergency situations]. Belgorod: BGU. 166 p.
6. Zdorovtsov, I. A., and V. Yu. Korolev. 2004. *Osnovy teorii nadezhnosti volokonno-opticheskikh liniy peredachi zheleznodorozhnogo transporta* [Fundamentals of reliability theory of fiber optic transmission lines for railway transport]. Moscow: MAKS Press. 308 p.
7. Kudryavtsev, A. A. 2018. Bayesovskie modeli balansa [Bayesian balance models]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):18–27.
8. Kudryavtsev, A. A. 2019. O predstavlenii gamma-eksponentsial'nogo i obobshchennogo otritsatel'nogo binomial'nogo raspredeleniy [On the representation of gamma-exponential and generalized negative binomial distributions]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):78–82.
9. Kritsky, S. N., and M. F. Menkel. 1946. O priemakh issledovaniya sluchaynykh kolebaniy rechnogo stoka [Methods of investigation of random fluctuations of river flow]. *Trudy NIU GUGMS Ser. IV* [Proceedings of GUGMS research institutions. Ser. IV] 29:3–32.
10. Kritsky, S. N., and M. F. Menkel. 1948. Vychor krivyykh raspredeleniya veroyatnostey dlya raschetov rechnogo stoka [Selection of probability distribution curves for river flow calculations]. *Izvestiya AN SSSR. Otd. tekhn. nauk* [Herald of the USSR Academy of Sciences. Technical Sciences] 6:15–21.
11. Kudryavtsev, A. A., and A. I. Titova. 2017. Gamma-eksponentsial'naya funktsiya v bayesovskikh modelyakh massovogo obsluzhivaniya [Gamma-exponential function in Bayesian queueing models]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):104–108.
12. Le Roy, É. 1900. Sur les series divergentes et les fonctions definies par un developpement de Taylor. *Ann. Faculté Sci. Toulouse 2 Ser.* 2(3):317–384.
13. Srivastava, H. M., and Z. Tomovski. 2009. Fractional calculus with an integral operator containing a generalized Mittag-Leffler function in the kernel. *Appl. Math. Comput.* 211:198–210.
14. Gorenlo, R., A. A. Kilbas, F. Mainardi, and S. V. Rogosin. 2014. *Mittag-Leffler functions, related topics and applications*. Berlin–Heidelberg: Springer-Verlag. 443 p.

Received July 4, 2020

## Contributors

**Kudryavtsev Alexey A.** (b. 1978) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; nubigena@mail.ru

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

# ОСНОВНЫЕ ПОНЯТИЯ ПРОГРАММИРОВАНИЯ В ИЗЛОЖЕНИИ ДЛЯ ДОШКОЛЬНИКОВ\*

В. Б. Бетелин<sup>1</sup>, А. Г. Кушниренко<sup>2</sup>, А. Г. Леонов<sup>3</sup>

**Аннотация:** Развитие информационных технологий сформировало социально-экономический запрос на снижение возраста знакомства детей с программированием. В результате шестилетних усилий авторам удалось разработать и массово внедрить годовой курс программирования для дошкольников, построенный на метафоре программного управления. В процессе развития курса удалось отобрать и сформулировать набор основных понятий программирования, который может быть освоен дошкольниками возраста 6+ в деятельностно-игровой форме. Этот набор понятий вводится на примерах программ управления движущимися и неподвижными объектами с интуитивно понятными, обозримыми системами команд. Курс строится на базе бестекстовой пиктографической системы программирования «ПиктоМир» разработки ФНЦ НИИСИ РАН. Разработанное для курса программно-методическое наполнение позволяет каждому дошкольнику к концу курса получить опыт составления и отладки 120–150 простейших программ.

**Ключевые слова:** информатика; робот; программа; компьютер; язык программирования; дошкольник; «ПиктоМир»; пиктограмма

**DOI:** 10.14357/19922264200308

## 1 Введение

Законодательными структурами власти России федерального уровня выдвигаются предложения по понижению возраста знакомства детей с информатикой и программированием на уровень системы дошкольного образования [1]. Настоящая статья посвящена описанию конкретной методики такого понижения и суммирует шестилетний опыт разработки и массового внедрения годового курса «Алгоритмика для дошколят», проводимого совместными усилиями ФНЦ НИИСИ РАН и Департамента образования администрации г. Сургута. Начиная с осени 2018 г. годовой курс проходят все выпускники всех подготовительных групп всех дошкольно-образовательных организаций г. Сургута — около 6 тыс. детей. В курсе используется бестекстовая учебная система программирования «ПиктоМир», разработка которой была начата в ФНЦ НИИСИ РАН около 10 лет назад [2] и продолжается сегодня в работах по теме госзадания Минобрнауки РФ «Разработка, реализация и внедрение семейства интегрированных многоязыковых сред программирования. . .» [3]. Система

«ПиктоМир» и методическое обеспечение курса — свободно распространяемые. Их можно загрузить с сайта НИИСИ РАН или работать с ними с помощью браузера через веб-интерфейс [4].

Информатизация дошкольного образования — очень широкая и важная для современной цивилизации тема [5]. В данной статье, в рамках более общей работы над 4-летним курсом программирования для дошкольников и младших школьников, упор делается на конкретной задаче — организации первых занятий курса программирования для дошкольников. Этот вопрос особенно важен, поскольку именно на первых занятиях закладывается фундамент курса, осваивается набор основных понятий.

Накопленный опыт позволил авторам предложить набор основных понятий программирования, который может быть органично освоен дошкольниками в деятельностно-игровой форме. Этот набор понятий призван раскрыть одну из «больших идей» нашей цивилизации — *принцип программного управления*.

Согласно рубрикации А. Л. Семенова [6, п. 12], одной из целей курса информатики должно

\* Работа выполнена по теме 0065-2019-0010 госзадания 2020 года в отделе учебной информатики Федерального научного центра «Научно-исследовательский институт системных исследований» Российской академии наук.

<sup>1</sup>Федеральный научный центр «Научно-исследовательский институт системных исследований» Российской академии наук, betelin@niisi.msk.ru

<sup>2</sup>Федеральный научный центр «Научно-исследовательский институт системных исследований» Российской академии наук, agk\_@mail.ru

<sup>3</sup>Московский государственный университет им. М. В. Ломоносова; Федеральный научный центр «Научно-исследовательский институт системных исследований» Российской академии наук; Московский педагогический государственный университет, dr.l@vip.niisi.ru

быть «освоение информационной картины мира». Принцип программного управления и строит такую картину в терминах и образах, понятных шестилетнему ребенку. Схематично принцип может быть объяснен так: **любую работу, которую человек может выполнить, команду механическим помощником-роботом, можно перепоручить компьютеру, если человеку удастся составить программу выполнения той деятельности, которую роботу надлежит выполнить.**

Методически правильным представляется двухэтапное изложение принципа программного управления, при котором на первом этапе излагается простейший вариант принципа — без обратной связи. Практика работы с дошкольниками показала, что именно этот этап оказывается основополагающим и более трудным, чем второй этап, на котором вводится обратная связь.

## 2 Принцип программного управления без обратной связи

**Программа** — это план будущей деятельности, в процессе которой один объект — **компьютер** — управляет другим объектом — **роботом** — по программе, заранее составленной человеком — **программистом** — по заранее известным **правилам составления программ (язык программирования)**. Процесс **выполнения** программы компьютером состоит в том, что компьютер, следуя программе, некоторым заранее установленным способом дает роботу **команды**, которые тот **исполняет**, докладывая компьютеру об окончании исполнения каждой команды и готовности к приему следующей команды. Чтобы компьютер мог выполнить программу, она должна быть ему предварительно сообщена (**загружена в память компьютера**).

Это описание принципа явно вводит 12 понятий: 6 **объектов**, 1 **субъект** и 5 **взаимодействий** между объектами и субъектом.

*Объекты:*

**программа; компьютер; память компьютера; робот; правила составления программ (язык программирования); команда.**

*Субъект:*

**программист.**

*Взаимодействия:*

- программист **составляет** программу;
- компьютер **выполняет** программу, **давая** роботу команды;
- получив команду, робот ее **исполняет** и ждет поступления следующей команды;
- компьютер **загружает в свою память** сообщенную ему программу.

Разумеется, выбранный выше набор понятий и акценты, сделанные в объяснении принципа, могли бы быть другими. Авторы выбрали указанный набор из 12 понятий, исходя из сугубо практических соображений. Дело в том, что на первых занятиях с дошкольниками педагог должен одновременно решить две задачи:

- (1) добиться интуитивного понимания детьми «правил игры», интуитивного осознания детьми предложенной системы понятий;
- (2) пополнить словарный запас детей, научить их использовать в речевой практике термины, выражающие освоенные понятия.

Предлагаемый выбор понятий позволяет педагогу вчерне решить обе задачи примерно за 10–15 получасовых групповых занятий и добиться твердого усвоения системы понятий к концу годового курса. Важно, что данные понятия образуют некоторую замкнутую систему. Скорее всего, для ребенка 6 лет, пришедшего на занятия алгоритмикой, данная система понятий окажется первой в его жизни изученной взаимосвязанной **системой научных понятий**. И это изучение должно быть организовано так, чтобы система понятий была понята до конца каждым ребенком.

Согласно Л. С. Выготскому [7], осознание любого общего принципа требует комплексного освоения ребенком некоторой **системы научных понятий**: «Научные понятия являются воротами, через которые осознанность входит в царство детских понятий».

**Принцип программного управления может быть понят, осознан ребенком только после усвоения изложенной выше достаточно сложной системы из 12 научных понятий.**

Разумеется, доведение перечисленных понятий до ребенка возраста 6–7 лет в вербальной форме, путем устных объяснений взрослого, невозможно. Осознанное усвоение этих понятий станет возможным, только если ребенку будут предложены виды деятельности, позволяющие ему в игровой форме «вжиться» **во все** эти 12 понятий, «пропустить их через себя».

Подчеркнем, что принцип программного управления можно вводить по-разному, варьируя набор понятий, которые преподносятся как основные, и понятий, которые вводятся неявно. Предложенный выше набор из 12 основных понятий был подобран так, чтобы максимально облегчить ребенку освоение **каждого** из этих понятий и **всей системы** понятий в деятельностно-игровой форме.

Авторам удалось построить первые занятия курса так, чтобы каждый ребенок в группе смог про-

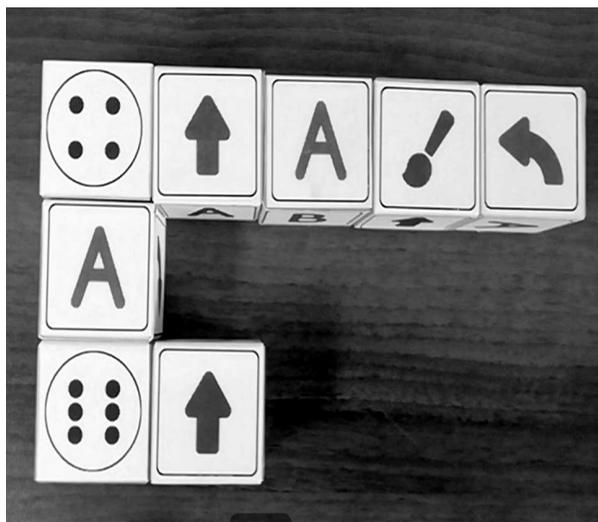


(а)

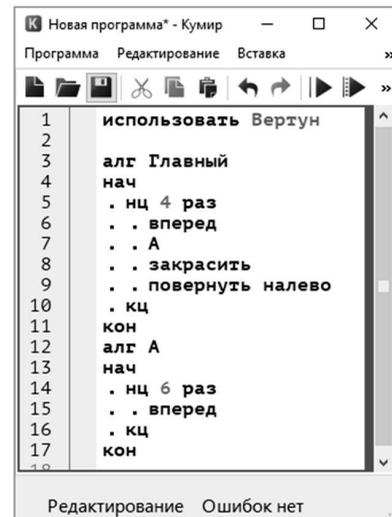


(б)

Рис. 1 Виртуальные роботы (а) и реальный робот-игрушка (б)



(а)



(б)

Рис. 2 Программа управления роботом, выложенная дошкольником из кубиков (а), и та же программа, составленная семиклассником (б)

имитировать выполнение **всех** пяти перечисленных выше действий-взаимодействий, выполняя поочередно роль робота, компьютера и программиста, и смог поработать с материальным воплощением программы, составляя ее и загружая в память компьютера и сталкиваясь с трудностями в случае нарушения правил составления программы. Это удалось сделать за счет использования на первых занятиях следующих методических и технических решений.

Как было предложено Пейпертом еще полвека назад [8], дети работают не только с виртуальными (экранными) роботами, но и с реальными роботами-игрушками, которые перемещаются по полу игровой комнаты, имитируя перемещения виртуальных роботов на экране планшета (рис. 1).

Реальные роботы управляются звуковыми командами. Эти команды «наблюдаемы» (слышимы) детьми.

Программы составляются из материальных объектов, кубиков, с нанесенными на их грани пиктограммами команд, повторителей и других конструкций языка программирования (рис. 2). Функции компьютера выполняет планшет.

Загрузка программы в память компьютера (планшета) состоит в явно проводимом ребенком фотографировании камерой планшета выложенной на столе программы, т. е. некоторой конфигурации кубиков. За этим явным действием невидимо для ребенка, скрытно, следует «понимание» программы компьютером — распознавание фотографии с ку-

биками процессором планшета с помощью нейронных сетей. Результат этого понимания ребенок видит на экране планшета, а как это понимание происходит — с ребенком не обсуждается.

Первое синтаксическое правило составления программы из кубиков требует, чтобы кубики были выложены в достаточно ровные ряды, ряды располагались друг под другом. Первое семантическое правило выполнения программы гласит: при выполнении программы пиктограммы в рядах читаются слева направо, а ряды читаются сверху вниз. При этом ребенок сталкивается с тем, что компьютеру не удастся «понять» программы, составленные с нарушением правил, т. е. компьютеру «непонятны» расположения кубиков на столе, в которых трудно или невозможно мысленно разбить выложенные кубики на ряды (рис. 3)<sup>1</sup>.

Еще две группы правил описывают две конструкции языка программирования: числовой повтори-тель и подпрограмма с однобуквенным именем. Эти правила описывают и синтаксис, и семантику и применяются в ситуациях, когда ребенок, имитируя компьютер, пытается «понять» выложенную другим ребенком программу и далее пытается «понять» программу выполнить. Эти правила требуют, чтобы пиктограммы располагались в рядах в определенном порядке. Дети осваивают эти правила без затруднений.

Рассматриваемый курс построен так, что на первых занятиях ребенок играет со сверстниками и учебными пособиями (роботом и набором кубиков с пиктограммами команд) в сюжетно-ролевые игры. **Компьютер на первых порах используется только по его главному назначению, для исполнения загруженных в его память программ**, и не используется для других целей: ни для составления программ, ни для генерации изображений виртуальных роботов и виртуальных обстановок на экране. На первых занятиях курса **программа, робот и обстановка, в которой робот действует, являются реальными, а не виртуальными объектами**, и все взаимодействия представляют собой реальные процессы с участием материальных объектов. И ребенок может осваивать роли объектов в игре. Ребенок может выступать в роли робота, исполняя звуковые команды, поступающие от компьютера или от другого ребенка, выступающего в роли компьютера; ребенок может выступать в роли компьютера, выполняя составленную другим ребенком программу и командуя при этом третьим ребенком, играющим роль робота; ребенок может поработать программистом, составляя самостоятельно программу путем перемещения материальных кубиков на столе и переходя позднее



Рис. 3 Иллюстрация художника Михаила Гладковского к докладу А. П. Ершова «Программирование — вторая грамотность»

к составлению программ путем псевдоматериального перемещения рукой пиктограмм на сенсорном экране планшета.

Важно, чтобы по мере того, как основные понятия программирования осваиваются детьми на интуитивном уровне при работе с реальными роботами на ковриках, кубиками на столе и виртуальными роботами, ковриками и программами на экранах планшетов, на бескомпьютерных половинах занятий происходил перевод этих интуитивных представлений в вербальную форму. Под руководством воспитателя дети должны обсуждать значения слов *программист, робот, программа*, значения фраз типа «я выполнил программу, которую составил Коля», «программа составлена из 6 пиктограмм», «робот выполнил 10 команд». Это пополнение словарного запаса детей и развитие навыков монолога и диалога с использованием накопленного «профессионального» словарного запаса является столь же важной целью курса, как и обретение навыков самостоятельного составления простейших программ в учебно-игровой системе программирования.

<sup>1</sup>Невыровненность кубиков может рассматриваться как непрерывный аналог «синтаксической неправильности» программы.

### 3 Принцип программного управления с обратной связью

Большую часть курса «Алгоритмика для дошколят» занимает составление программ управления без обратной связи. Каждая такая программа решает одну задачу: обеспечивает нужное поведение робота в одной-единственной внешней обстановке, предъявляемой ребенку на полу в игровой комнате или в графической форме на экране планшета. Программы без обратной связи состояются с использованием всего **двух явных** управляющих конструкций: числовой повторитель и подпрограмма с однобуквенным именем и **одной неявной** конструкцией — последовательного выполнения линейного участка программы.

В «поминутной» методичке годового курса «Алгоритмика для дошколят» описаны 30 занятий, еще 4 занятия предусмотрены как резервные. Последовательное выполнение линейного участка программы появляется на первом же занятии. Конструкция *повторитель* впервые появляется на занятии № 10 и вводится как способ сокращения размера программы. Конструкция *подпрограмма* впервые появляется на занятии № 15 и вводится как способ «шифровки» фрагментов программы. Позднее эта конструкция рассматривается еще и как способ сокращения размера программы. Разумеется, выразительная сила двух выбранных управляющих конструкций невелика. Все программы, которые можно составить с использованием этих конструкций, являются линейными. Однако эти линейные программы могут иметь достаточно сложную структуру управления.

Практика показала, что набора содержательных задач, решаемых с использованием этих двух управляющих конструкций, достаточно для удержания внимания детей в течение года (первые 25 занятий из 30) при условии создания достаточного разнообразия роботов и их графических представлений. В настоящее время в курсе «Алгоритмика для дошколят» используются 5 виртуальных роботов и 1 реальный робот (Ползун), имитирующий одного из виртуальных. Несмотря на весьма малую продолжительность компьютерной части каждого занятия курса — от 15 мин в первом полугодии до 20 мин во втором — удается добиться того, что каждый ребенок на каждом занятии выполняет 4–5 заданий, т. е. **в годовом курсе каждый дошкольник самостоятельно составляет 120–150 простейших программ**. Самостоятельное выполнение более сотни упражнений представляется необходимым условием устойчивого освоения теоретического и практического материала курса.

В конце первого года обучения, на последних занятиях, начинается (но не завершается) переход от управления без обратной связи к управлению с ее использованием:

- в системы команд роботов вводятся команды-вопросы, и в предоставляемые системой «ПиктоМир» конструкции языка программирования добавляются ветвление и повторение;
- в систему основных понятий вводятся новый вид команды — **команда-вопрос** и новый вид взаимодействия: робот **отвечает** на команду-вопрос компьютера **да** или **нет**.

Параллельно с введением понятия «обратная связь» в систему основных понятий вводится и понятие **число** (неотрицательное целое число). Для «материализации» понятия «число» вводится виртуальный исполнитель «Волшебный кувшин с камнями», играющий роль счетчика. Наличие счетчика позволяет с помощью подсчета числа шагов решать задачи управления роботом типа «дойти до ближайшей стены и вернуться в исходную точку».

Разумеется, введение новых понятий сопровождается играми. Исполняя роль робота, дети отвечают на команды-вопросы «да» или «нет»; исполняя роль кувшина-счетчика, ребенок по команде добавляет или удаляет камешек из реального кувшина и отвечает на команды-вопросы «кувшин пуст?», «кувшин не пуст?» и «сколько камней в кувшине?». На введение обратной связи в курсе «Алгоритмика для дошколят» требуется не менее 5 занятий. На устойчивое освоение этих понятий на следующем году обучения необходимо еще 15 занятий.

После введения обратной связи становится возможным давать задачи на составление «универсальных» программ, работающих не в одной, а в нескольких разных обстановках. Эти задачи также даются в графической форме, без словесного описания класса обстановок, в которых должна работать программа. Просто на очередном уровне игры требуется составить программу, которая работает не в одной, как раньше, а в двух или трех заданных обстановках.

### 4 Выводы

Опыт показал, что дети возраста 6–7 лет без труда и с энтузиазмом осваивают азы программирования с использованием описанного выше подхода и готовы продолжать занятия программированием в школе. Авторы считают раннее освоение основ программирования в описанном выше объеме необходимым.

## Литература

1. Глава профильного комитета Думы считает нужным ввести информатику в дошкольную программу: Матлы XVIII съезда «Единой России» // ТАСС, 8 декабря 2018. <https://tass.ru/obschestvo/5888487>.
2. Rogozhkina I., Kushnirenko A. PictoMir: Teaching programming concepts to preschoolers with a new tutorial environment // *Procd. Soc. Behv.*, 2011. Vol. 28. P. 601–605. doi: 10.1016/j.sbspro.2011.11.114.
3. Бесиановников Н. О., Кушниренко А. Г., Леонов А. Г., Малы́й А. А. Проект двуязыковой пиктограммно-текстовой учебной среды программирования ПиктоМир-К // Свободное программное обеспечение в высшей школе: Сб. тезисов XIV конф. — М.: МАКС Пресс, 2019. С. 64–66.
4. ПиктоМир: Стартовая страница проекта на сайте Федерального научного центра «Научно-исследовательский институт системных исследований» Российской академии наук. <https://www.niisi.ru/piktomir>.
5. Калаш И. Возможности информационных и коммуникационных технологий в дошкольном образовании: Аналитический обзор / Пер. с англ. под ред. А. Л. Семенова. — Институт ЮНЕСКО по информационным технологиям в дошкольном образовании, 2010. 176 с. <https://iite.unesco.org/pics/publications/ru/files/3214673.pdf>. (*Kabas, I.* Recognizing the potential of ICT in early childhood education: Analytical survey. — UNESCO Institute for Information Technologies in Education, 2010. 148 p. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000190433.pdf> (accessed September 2, 2020).)
6. Семёнов А. Л. Концептуальные проблемы информатики, алгоритмики и программирования в школе // Вестник кибернетики, 2016. № 2(22). С. 12–16.
7. Выготский Л. С. Мышление и речь. — Изд. 5-е, испр. — М.: Лабиринт, 1999. Гл. 6.
8. Пейперт С. Переворот в сознании: Дети, компьютеры и плодотворные идеи / Пер. с англ. — М.: Педагогика, 1989. 224 с. (*Papert S.* Mindstorms: Children, computers and powerful ideas. — New York, NY, USA: Basic Books, 1980. 252 p.)

Поступила в редакцию 20.08.19 (последняя правка 21.07.20)

---

---

## BASIC CONCEPTS OF PROGRAMMING EXPOUNDED FOR PRESCHOOLERS

V. B. Betelin<sup>1</sup>, A. G. Kushnirenko<sup>1</sup>, and A. G. Leonov<sup>1,2,3</sup>

<sup>1</sup>Federal Research Center “Scientific Research Institute for System Analysis of the Russian Academy of Sciences,” 36-1 Nakhimovsky Prosp., Moscow 117218, Russian Federation

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

<sup>3</sup>Moscow Pedagogical State University, 1-1 Malaya Pirogovskaya Str., Moscow 119991, Russian Federation

**Abstract:** The development of information technology has formed a socioeconomic demand for reducing the age of acquaintance of children with programming. As a result of 6 years of efforts, the authors managed to develop and massively introduce an annual programming course for preschoolers built on the metaphor of program control. In the process of developing the course, the authors were able to select and formulate a set of basic programming concepts which fully reveals this metaphor and, at the same time, can be mastered by preschool children age 6+ in an active-play form. This set of concepts is introduced using examples of control programs for moving and stationary objects with an intuitive, visible command system. At the beginning of the course, control without feedback is introduced, the concept of feedback is introduced and used only at the end of the course. As a basic pedagogical software product, the PictoMir text-free pictographic system developed by the Federal Research Center “Scientific Research Institute for System Analysis of the Russian Academy of Sciences” and its programmatic and methodological content is used, allowing each preschooler to gain experience in writing and debugging at least 120–150 simplest programs by the end of the course.

**Keywords:** informatics; robot; program; computer; programming language; preschooler; PictoMir; pictogram

**DOI:** 10.14357/19922264200308

## Acknowledgments

The work was completed on the subject of the Government order 0065-2019-0010 in 2020 in the Department of Educational Informatics, SRISA RAS.

## References

1. Materialy TASS XVIII s"ezda "Edinoy Rossii" [TASS Materials of the 18th All-Russian political party "UNITED RUSSIA" Congress]. Available at: <https://tass.ru/obschestvo/5888487> (accessed July 24, 2020).
2. Rogozhkina, I., and A. Kushnirenko. 2011. PiktoMir: Teaching programming concepts to preschoolers with a new tutorial environment. *Procd. Soc. Behv.* 28:601–605. doi: 10.1016/j.sbspro.2011.11.114.
3. Besshaposnikov, N. O., A. G. Kushnirenko, A. G. Leonov, and A. A. Malyy. 2019. Proekt dvuyazykovoy piktogrammno-tekstovoy uchebnoy sredy programirovaniya PiktoMir-K [The project of the bilingual pictogram-text educational environment for programming PiktoMir-K]. *Sbornik tezisov XIV konf. "Svobodnoe programnoe obespechenie v vysshey shkole"* [14th Conference "Free Software in Higher Education" Proceedings]. Moscow: MAKS Press. 64–66.
4. PiktoMir: Startovaya stranitsa proekta na sayte Federal'nogo nauchnogo tsentra "Nauchno-issledovatel'skiy institut sistemnykh issledovaniy" Rossiyskoy akademii nauk [The start page of the PictoMir project on the website of the SRISA/NIISI RAS]. Available at: <https://www.niisi.ru/piktomir/> (accessed July 24, 2020).
5. Kabaš, I. 2010. Recognizing the potential of ICT in early childhood education: Analytical survey. UNESCO Institute for Information Technologies in Education. 148 p. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000190433.pdf> (accessed September 2, 2020).
6. Semyonov, A. L. 2016. Kontseptual'nye problemy informatiki, algoritmiki i programmirovaniya v shkole [Conceptual problems of teaching computer science, algorithm studies, and programming at school]. *Vestnik kibernetiki* [Proceedings in Cybernetics] 2(22):11–15.
7. Vygotskiy, L. S. 1999. *Myshlenie i rech'* [Thinking and saying]. Moscow: Labirint. Ch. 6.
8. Papert, S. 1980. *Mindstorms: Children, computers and powerful ideas*. New York, NY: Basic Books. 252 p.

*Received August 20, 2019 (last revision July 21, 2020)*

## Contributors

**Betelin Vladimir B.** (b. 1946) — Doctor of Science in physics and mathematics, professor, Academician of RAS, research advisor, Federal Research Center "Scientific Research Institute for System Analysis of the Russian Academy of Sciences," 36-1 Nakhimovsky Prosp., Moscow 117218, Russian Federation; [betelin@niisi.msk.ru](mailto:betelin@niisi.msk.ru)

**Kushnirenko Anatoliy G.** (b. 1944) — Candidate of Science (PhD) in physics and mathematics, Head of Department, Federal Research Center "Scientific Research Institute for System Analysis of the Russian Academy of Sciences," 36 b1 Nakhimovsky Prosp., Moscow 117218, Russian Federation; [agk.\\_@mail.ru](mailto:agk._@mail.ru).

**Leonov Aleksandr G.** (b. 1961) — Candidate of Science (PhD) in physics and mathematics, associate professor, leading scientist, Department of Mechanics and Mathematics, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; head of laboratory, Federal Research Center "Scientific Research Institute for System Analysis of the Russian Academy of Sciences," 36-1 Nakhimovsky Prosp., Moscow 117218, Russian Federation; professor, senior scientist, Moscow Pedagogical State University, 1-1 Malaya Pirogovskaya Str., Moscow 119991, Russian Federation; [dr.l@vip.niisi.ru](mailto:dr.l@vip.niisi.ru)

# ВЫЧИСЛИТЕЛЬНЫЕ АСПЕКТЫ ПРИМЕНЕНИЯ CC-VaR НА СОВОКУПНОСТИ РЫНКОВ\*

Г. А. Агасандян<sup>1</sup>

**Аннотация:** Работа служит непосредственным продолжением предыдущей работы автора, посвященной применению континуального критерия VaR на совокупности нескольких рынков разных размерностей, связанных между собой базовыми активами. Исследование нацелено на приложение идей и методов, развитых для теоретической континуальной модели, к дискретным сценарным рынкам. В модели совокупности одного двумерного и двух одномерных рынков, а также ее усеченных вариантов приводятся конструкции оптимальных сценарных портфелей из базисных инструментов всех рынков совокупности с применением рандомизации. Предлагаемые конструкции проверяются на числовых примерах с использованием потенциально типичных двумерных расширений бета-распределений для описания прогноза будущих цен базовых активов и картины текущих цен базисных инструментов. Изложение сопровождается расчетами весовых коэффициентов базисных инструментов оптимальных портфелей и иллюстрируется графиками портфельных доходов.

**Ключевые слова:** базовые активы; функция рискованных предпочтений; континуальный критерий VaR (CC-VaR); стоимостная и прогнозная плотности; функция относительных доходов; процедура Неймана–Пирсона; комбинированный портфель; суррогатный портфель; идеалистичный портфель

DOI: 10.14357/19922264200309

## 1 Введение

Работа затрагивает круг проблем, связанных с применением на финансовых рынках континуального критерия VaR (CC-VaR), который был введен и изучался автором ранее [1–5]. Исследование служит непосредственным продолжением работы [6] и использует ее идеи и конструкции. В работе [6] последовательно излагалась методология применения CC-VaR сразу на совокупности трех теоретических континуальных рынков разных размерностей, связанных между собой базовыми активами, вводились правила переключения между рынками, формировались базисы из инструментов этих рынков с привлечением механизма рандомизации, а также строились в таких базисах оптимальные по CC-VaR комбинированные портфели.

В данной работе методология, развитая в [6] для модели теоретических континуальных рынков, переносится на дискретные сценарные рынки с целью проверки действенности модели на примерах и иллюстрации результатов.

Вновь рассматривается совокупность трех рынков. Один из них является двумерным рынком, а два других — одномерными. При этом базовые активы одномерных рынков образуют пару базовых активов двумерного. Решение ищется в форме

совмещения трех сценарных портфелей на своих рынках. Оно основывается на анализе расхождений в относительных доходах по сценариям, которые обусловлены расхождениями в ценах на разных рынках. При этом наряду с совокупностью из трех рынков для полноты исследования рассматриваются и две усеченные, состоящие из пары рынков. Оптимальный комбинированный портфель строится с применением идей и техники рандомизации.

Предлагаемые конструкции иллюстрируются числовыми примерами и графиками портфельных доходов с использованием двумерных расширений стандартных бета-распределений для описания прогноза будущих цен базовых активов и картины текущих цен базисных инструментов.

## 2 Сценарные рынки

Рассматривается однопериодный тройственный рынок, состоящий из двух одномерных рынков  $\#X$  и  $\#Y$  с базовыми активами  $X$  и  $Y$  соответственно и одного двумерного  $\#0$  с парой активов  $(X, Y)$ . Возможные цены базовых активов образуют множества  $X$  и  $Y$ . Сценарная дискретизация вводится разбиением множества  $X = [x_0, x_n)$  на  $n$  сценариев  $S_i = [x_{i-1}, x_i) \subset X$ ,  $x_{i-1} < x_i$ ,  $i \in I = \{1, \dots, n\}$ , и множества  $Y = [y_0, y_m)$  на  $m$  сце-

\* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 17-01-00816.

<sup>1</sup> Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, agasand17@yandex.ru

нариев  $T_j = [y_{j-1}, y_j) \subset Y$ ,  $y_{j-1} < y_j$ ,  $j \in J = \{1, \dots, m\}$ . Прямым произведением одномерных сценариев разных рынков получаются  $n \times m$  двумерных сценариев. Равномерное разбиение задается правилом  $x_i = x_0 + ih_1$ ,  $h_1 = (x_n - x_0)/n$ ,  $i \in I$ ;  $y_j = y_0 + jh_2$ ,  $h_2 = (y_m - y_0)/m$ ,  $j \in J$ . Стоимость инструмента  $G$  записывается как  $|G|$ , а средний доход —  $\|G\|$ .

Связующим звеном теоретических и сценарных рынков становятся инструментальные индикаторы сценариев. Их платежными функциями служат характеристические функции сценариев, и они играют на сценарных рынках роль  $\delta$ -инструментов теоретического рынка.

Как и для теоретического тройственного рынка, задается двумерная прогнозная плотность  $p(x, y)$ ,  $x \in X$ ,  $y \in Y$ . Ее маргинальные плотности  $p_1(x)$  и  $p_2(y)$  служат одновременно и прогнозными плотностями для рынков #X и #Y соответственно. Однако подобное свойство в отношении стоимостных плотностей может и не выполняться, и для рынка #0 задается стоимостная плотность  $c(x, y)$ , для рынка #X —  $c_X(x)$ , для рынка #Y —  $c_Y(y)$ ,  $x \in X$ ,  $y \in Y$ . Отметим еще, что эти плотности вводятся для удобства формирования картины цен для сценарных рынков, хотя достаточно ограничиться дискретным набором стоимостей.

Индикаторы сценариев  $D_{ij} = H_{ij} (= H\{S_i \times T_j\})$  служат базисными инструментами на рынке #0,  $D_{X;i} = H_{X;i}$  — на рынке #X,  $D_{Y;j} = H_{Y;j}$  — на рынке #Y. Их цены и вероятности (средние доходы) образуют векторы  $c$  и  $p$ ,  $c_X$  и  $p_X$ ,  $c_Y$  и  $p_Y$  с компонентами соответственно ( $i \in I, j \in J$ ):

$$c_{ij} = |D_{ij}| = \int_{y_{j-1}}^{y_j} \int_{x_{i-1}}^{x_i} c_{ij}(x, y) dx dy,$$

$$p_{ij} = \|D_{ij}\| = \int_{y_{j-1}}^{y_j} \int_{x_{i-1}}^{x_i} p_{ij}(x, y) dx dy;$$

$$c_{X;i} = |D_{X;i}| = \int_{x_{i-1}}^{x_i} c_X(x) dx,$$

$$p_{X;i} = \|D_{X;i}\| = \int_{x_{i-1}}^{x_i} p_1(x) dx;$$

$$c_{Y;j} = |D_{Y;j}| = \int_{y_{j-1}}^{y_j} c_Y(y) dy,$$

$$p_{Y;j} = \|D_{Y;j}\| = \int_{y_{j-1}}^{y_j} p_2(y) dy.$$

### 3 Оптимизация на сценарном тройственном рынке

Под оптимизацией на сценарном рынке понимается построение сценарного портфеля применением дискретного алгоритма, получаемого проецированием на сценарный рынок алгоритма оптимизации для теоретического рынка. Как обычно в задачах с CC-VaR, в качестве дискретного здесь используется *стандартный дискретный алгоритм* [2–5], основанный на процедуре Неймана–Пирсона [7]. В алгоритме основные агрегаты, такие как  $p$  и  $c$ , участвуют в векторной форме (так же и для двумерного рынка).

В краткой форме он записывается последовательностью операций:

$$\rho = \frac{p}{c}, \quad \xi = O(\rho), \quad \eta = O(\xi), \quad d = p(\xi),$$

$$T = \left\{ t_{ij} = \begin{cases} 1, & i \leq j; \\ 0, & i > j \end{cases} \right\}, \quad \varepsilon = T \cdot d;$$

$$b = \phi(\varepsilon), \quad g = b(\eta),$$

где  $O$  — оператор упорядочения векторов в порядке возрастания их компонент.

Перед запуском его работы с целью построения оптимального комбинированного портфеля в базе, составленном из инструментов трех рынков, потребуется специальная подготовка исходных данных.

#### 3.1 Правила замещения

Представления портфелей #0, #X и #Y с весовыми векторами  $g$ ,  $g_X$  и  $g_Y$  соответственно имеют вид:

$$G = \sum_{\substack{i \in I \\ j \in J}} g_{ij} D_{ij};$$

$$G_X = \sum_{i \in I} g_{X;i} D_{X;i};$$

$$G_Y = \sum_{j \in J} g_{Y;j} D_{Y;j}.$$

Применение CC-VaR предполагает анализ относительных доходов  $\rho_{ij} \equiv p_{ij}/c_{ij}$ ,  $\rho_{X;i} \equiv p_{1;i}/c_{X;i}$  и  $\rho_{Y;j} \equiv p_{2;j}/c_{Y;j}$ ,  $i \in I, j \in J$ , для рынков #0, #X и #Y соответственно.

С помощью этих характеристик основные *правила замещения* (4)–(6) в [6], одновременно вводящие двумерные множества индексов (сценариев)  $M_0, M_1, M_2 \subset I \times J$ , теперь записываются в виде:

$$(i, j) \in M_0 \Leftrightarrow \{\rho_{ij} \geq \rho_{X;i} \& \rho_{ij} \geq \rho_{Y;j}\}; \quad (1)$$

$$(i, j) \in M_1 \Leftrightarrow \{\rho_{X;i} > \rho_{ij} \& \rho_{X;i} \geq \rho_{Y;j}\}; \quad (2)$$

$$(i, j) \in M_2 \Leftrightarrow \{\rho_{Y;j} > \rho_{ij} \& \rho_{Y;j} > \rho_{X;i}\}. \quad (3)$$

Множества  $M_0$ ,  $M_1$  и  $M_2$  взаимно не пересекаются и в объединении дают полное множество  $X \times Y$ . Функция замещения в дискретном случае трансформируется в *матрицу замещений* (помечающую сценарии индексами 1, 2, 3):

$$\mathbf{A} = \|a_{ij}\|, \quad a_{ij} = k \Leftrightarrow (i, j) \in M_k, \quad k = 0, 1, 2, \quad i \in I, \quad j \in J. \quad (4)$$

Для рынков #X и #Y используются векторы  $\mathbf{a}_X$  и  $\mathbf{a}_Y$  с компонентами:

$$a_{X;i} = \begin{cases} 1, & \exists j \in J : a_{ij} = 1; \\ 0 & \text{иначе} \end{cases}, \quad i \in I; \quad (5)$$

$$a_{Y;j} = \begin{cases} 1, & \exists i \in I : a_{ij} = 2; \\ 0 & \text{иначе} \end{cases}, \quad j \in J. \quad (6)$$

Первый выделяет сценарии замещения на рынке #X, второй — на рынке #Y.

Рассматриваемую задачу, совмещающую все три рынка, назовем задачей *A*. Аналогично могут быть рассмотрены и две ее усеченные постановки — *B* и *C*. В задаче *B* исключается рынок #Y и метка  $k = 2$  не используется, а ее правила замещения:

$$(i, j) \in M_0 \Leftrightarrow \{\rho_{ij} \geq \rho_{X;i}\};$$

$$(i, j) \in M_1 \Leftrightarrow \{\rho_{X;i} > \rho_{ij}\}.$$

В задаче *C* исключается рынок #0 и не используется метка  $k = 0$ , а ее правила замещения:

$$(i, j) \in M_1 \Leftrightarrow \{\rho_{X;i} \geq \rho_{Y;j}\};$$

$$(i, j) \in M_2 \Leftrightarrow \{\rho_{Y;j} > \rho_{X;i}\}.$$

### 3.2 Комбинированный портфель

По аналогии с теоретической схемой и в соответствии с правилами (3)–(5) строится *комбинированный* портфель замещением тех базисных инструментов рынка #0, для которых имеется базисный инструмент рынка #X или #Y с большим относительным доходом. При этом также применяется рандомизация.

Формулы (8)–(20) из [6] для теоретического рынка переписываются в несколько сокращенном виде применительно к сценарному рынку с естественной заменой переменных  $s$  и  $t$  индексами  $i$  и  $j$ , а множества  $X$  и  $Y$  — множествами индексов  $I$  и  $J$ .

Вводятся множества  $M_{1;i} \subset J$  и  $M_{2;j} \subset I$  как максимальные подмножества множеств  $M_1$  и  $M_2$  для фиксированных значений  $i \in I$  и  $j \in J$  соответственно. Также определяются индикаторы  $M_{1;i}$  и  $M_{2;j}$  рынка #0 как объединение базисных инструментов  $D_{ij}$  по  $j \in M_{1;i}$  и  $i \in M_{2;j}$  соответственно:

$$M_{1;i} = \sum_{j \in M_{1;i}} D_{ij} = D_{1;i} \times H_2 \{M_{1;i}\},$$

$$|M_{1;i}| = \sum_{j \in M_{1;i}} c_{ij}, \quad \|M_{1;i}\| = \sum_{j \in M_{1;i}} p_{ij}; \quad (7)$$

$$M_{2;j} = \sum_{i \in M_{2;j}} D_{ij} = D_{2;j} \times H_1 \{M_{2;j}\},$$

$$|M_{2;j}| = \sum_{i \in M_{2;j}} c_{ij}, \quad \|M_{2;j}\| = \sum_{i \in M_{2;j}} p_{ij}.$$

Эти инструменты желательно было бы заместить «гибридными», совмещающими рынки #0 и #X или #0 и #Y, а именно:

$$\left. \begin{aligned} M_{X;i} &= D_{X;i} \times H_2 \{M_{1;i}\}; \\ M_{Y;j} &= D_{Y;j} \times H_1 \{M_{2;j}\}. \end{aligned} \right\} \quad (8)$$

Поскольку таких инструментов на рассматриваемых рынках нет, для образования их реализуемого аналога на рынках #X и #Y применяется рандомизация.

На рынке #X для всех сценариев  $i \in I$  вводятся взаимонезависимые *биномиальные* случайные величины  $\vartheta_{X;i}$ , вероятности *успеха* (замещения) которых соответственно равны  $\theta_{X;i}$  и в совокупности образуют вектор  $\theta_X$ :

$$\theta_{X;i} = P \{M_{1;i}|i\} = \sum_{j \in M_{1;i}} \frac{p_{ij}}{p_{1;i}}, \quad \theta_X = \{\theta_{X;i}, i \in I\}. \quad (9)$$

Эта рандомизация встраивается непосредственно в базисные инструменты одномерной части #X комбинированного портфеля. Новые базисные инструменты становятся случайными, совпадающими с инструментами  $D_{X;i}$  с вероятностью  $\theta_{X;i}$  и с *нулевыми* инструментами (с нулевым доходом и нулевой стоимостью)  $N_{X;i}$  с вероятностью  $1 - \theta_{X;i}$ . Формально эти инструменты, их стоимость и средний доход (также вероятность) соответственно равны:

$$\begin{aligned} D_{X;i}^{\text{cmb}} &= \vartheta_{X;i} D_{X;i}, \quad |D_{X;i}^{\text{cmb}}| = c_{X;i}^{\text{cmb}} = \theta_{X;i} c_{X;i}, \\ \|D_{X;i}^{\text{cmb}}\| &= p_{X;i}^{\text{cmb}} = \theta_{X;i} p_{1;i}, \quad i \in I. \end{aligned} \quad (10)$$

Назначение (9) параметров  $\theta_{X;i}$  уравнивает вероятности, связанные с инструментами  $M_{X;i}$  (8)

и  $M_{1;i}$  (7). И подтверждением этому служит именно третье соотношение в (10).

Аналогично для рынка #Y вводятся случайные величины  $\vartheta_{Y;j}$ ,  $j \in J$ , с вероятностью замещения  $\theta_{Y;j}$ , и тогда

$$\left. \begin{aligned} D_{Y;j}^{\text{cmb}} &= \vartheta_{Y;j} D_{Y;j}, \quad |D_{Y;j}^{\text{cmb}}| = c_{Y;j}^{\text{cmb}} = \theta_{Y;j} c_{Y;j}, \\ \|D_{Y;j}^{\text{cmb}}\| &= p_{Y;j}^{\text{cmb}} = \theta_{Y;j} p_{2;j}, \\ \theta_{Y;j} &= P\{M_{2;j}|j\} = \sum_{i \in M_{2;j}} \frac{p_{ij}}{p_{2;j}}, \\ \theta_Y &= \{\theta_{Y;j}, j \in J\}. \end{aligned} \right\} (11)$$

Затем для нового базиса формируется новая *единая функция* относительных доходов, и к ней применяется комбинированный алгоритм оптимизации, находящий весовые коэффициенты оптимального (случайного) портфеля. Имеем:

$$G^{\text{cmb}} = \sum_{i,j=0} g_{ij}^{\text{cmb}} D_{ij} + \sum_{i \in I} g_{X;i}^{\text{cmb}} \vartheta_{1;i} D_{X;i} + \sum_{j \in J} g_{Y;j}^{\text{cmb}} \vartheta_{2;j} D_{Y;j}. \quad (12)$$

Это представление порождает и дискретную *идеалистичную* версию комбинированного портфеля с эквивалентной платежной функцией. Она получается заменой случайных базисных инструментов нереализуемыми на рынке детерминированными (8) (с сохранением всех весовых коэффициентов):

$$G^{\text{idl}} = \sum_{(i,j) \in M_0} g_{ij}^{\text{cmb}} D_{ij} + \sum_{i \in I} g_{X;i}^{\text{cmb}} M_{X;i} + \sum_{j \in J} g_{Y;j}^{\text{cmb}} M_{Y;j}.$$

Для графической иллюстрации платежной функции идеалистичного портфеля ее удобно рассчитывать по формуле:

$$\pi_{ij}^{\text{idl}}(x, y) = \max(\pi_{0;ij}^{\text{idl}}(x, y), \pi_{X;i}^{\text{idl}}(x), \pi_{Y;j}^{\text{idl}}(y)). \quad (13)$$

В задачах B и C соответственно:

$$G^{\text{cmb}} = \sum_{i,j=0} g_{ij}^{\text{cmb}} D_{ij} + \sum_{i \in I} g_{X;i}^{\text{cmb}} \vartheta_{1;i} D_{X;i};$$

$$G^{\text{cmb}} = \sum_{i \in I} g_{X;i}^{\text{cmb}} \vartheta_{1;i} D_{X;i} + \sum_{j \in J} g_{Y;j}^{\text{cmb}} \vartheta_{2;j} D_{Y;j}.$$

В этих двух представлениях сохраняются те же обозначения для весовых коэффициентов и величин  $\vartheta_{X;i}$  и  $\vartheta_{Y;j}$ , что и в задаче A, но все они, вообще говоря, другие, так как получаются из иных *правил замещений*. Также соответственно

$$\left. \begin{aligned} G^{\text{idl}} &= \sum_{(i,j) \in M_0} g_{ij}^{\text{cmb}} D_{ij} + \sum_{i \in I} g_{X;i}^{\text{cmb}} M_{X;i}; \\ G^{\text{idl}} &= \sum_{i \in I} g_{X;i}^{\text{cmb}} M_{X;i} + \sum_{j \in J} g_{Y;j}^{\text{cmb}} M_{Y;j}; \\ \pi_{ij}^{\text{idl}}(x, y) &= \max(\pi_{0;ij}^{\text{idl}}(x, y), \pi_{X;i}^{\text{idl}}(x)); \\ \pi_{ij}^{\text{idl}}(x, y) &= \max(\pi_{X;i}^{\text{idl}}(x), \pi_{Y;j}^{\text{idl}}(y)). \end{aligned} \right\} (14)$$

### 3.3 Суррогатный портфель

Наряду с комбинированным строится и портфель, называемый *суррогатным*. Он получается в результате формальной замены базисных инструментов для всех сценариев (по необходимости) новыми, для которых платежные функции и вероятности те же, что и для инструментов  $D_{ij}$ ,  $i \in I$ ,  $j \in J$ , рынка #0, но цены корректируются соответственно правилам (1)–(3) и ценам рынков #X и #Y:

$$c_{ij}^{\text{srg}} = \begin{cases} c_{ij}, & (i, j) \in M_0; \\ \frac{p_{ij}}{\rho_{X;i}}, & (i, j) \in M_1; \\ \frac{p_{ij}}{\rho_{Y;j}}, & (i, j) \in M_2. \end{cases} \quad (15)$$

С этими ценами вновь образуется матрица относительных доходов, алгоритм находит матрицу новых весов портфеля, а сам портфель имеет вид:

$$G^{\text{srg}} = \sum_{i \in I, j \in J} g_{ij}^{\text{srg}} D_{ij}^{\text{srg}}. \quad (16)$$

На сценарном рынке числовые показатели для суррогатного портфеля из-за замены части базисных инструментов будут, как правило, слегка отличаться от комбинированного (и идеалистичного) портфеля. Тем не менее относительно просто конструируемый *суррогатный* портфель можно использовать как дополнительное средство проверки всех расчетов (и графиков доходов).

В задачах B и C иными становятся цены:

$$c_{ij}^{\text{srg}} = \begin{cases} c_{ij}, & (i, j) \in M_0; \\ \frac{p_{ij}}{\rho_{X;i}}, & (i, j) \in M_1; \\ \frac{p_{ij}}{\rho_{Y;j}}, & (i, j) \in M_2. \end{cases}$$

По этим ценам пересчитываются матрицы относительных доходов и новых весов портфелей, а для построения портфелей применяется та же формула (16).

## 4 Алгоритм построения комбинированного портфеля

Построение комбинированного портфеля (*комбинированный* алгоритм) и его идеалистичной и суррогатной версий основано на применении *стандартного* алгоритма [2–5], краткая запись которого приводится в начале разд. 3. Для его запуска требуется предварительная (до применения *стандартного* алгоритма) трансформация матриц  $c$ ,  $p$  и  $\rho$  для рынка #0 и векторов  $c_X$ ,  $p_X$  и  $\rho_X$  для рынка #X и  $c_Y$ ,  $p_Y$  и  $\rho_Y$  для рынка #Y.

Схема *комбинированного* алгоритма приводится без технических подробностей. Она состоит из нескольких последовательных блоков.

1. Нахождение матрицы замещений  $A$  и построение матриц  $c_{0N}$ ,  $p_{0N}$  и  $\rho_{0N}$  операцией обнуления в матрицах  $c$ ,  $p$  и  $\rho$  всех элементов  $(i, j)$ , для которых  $a_{ij} \neq 0$ .
2. Образование матриц  $p_{XN}$  и  $p_{YN}$  обнулением всех элементов в матрице  $p$ , для которых соответственно  $a_{ij} \neq 1$  и  $a_{ij} \neq 2$ .
3. Определение векторов суммарной вероятности замещения  $p_{XM} = \{P\{M_{1;i}\}, i \in I\}$  и  $p_{YM} = \{P\{M_{2;j}\}, j \in J\}$  на основе результатов п. 2, а также векторов  $\theta_X = \{\theta_{X;i}, i \in I\}$  и  $\theta_Y = \{\theta_{Y;j}, j \in J\}$  согласно (9) и (11) соответственно.
4. Вычисление (в дополнение к  $p_{XM}$  и  $p_{YM}$ ) векторов  $c_{XM} = \{a_{X;i}\theta_{X;i}c_{X;i}, i \in I\}$ ,  $c_{YM} = \{a_{Y;j}\theta_{Y;j}c_{Y;j}, j \in J\}$  (9)–(11) и  $\rho_{XM} = \{a_{X;i}\rho_{X;i}, i \in I\}$ ,  $\rho_{YM} = \{a_{Y;j}\rho_{Y;j}, j \in J\}$  (параметры  $a_{X;i}$  (5) и  $a_{Y;j}$  (6) используются во избежание деления на нуль).
5. Преобразование матриц  $c_{0N}$ ,  $p_{0N}$  и  $\rho_{0N}$  в векторы (без изменения обозначений) и формирование комбинированной тройки векторов ( $c_F$ ,  $p_F$  и  $\rho_F$ ) размерности  $nm + n + m$  каждой операцией конкатенации:  $c_F = (c_{0N}, c_{XM}, c_{YM})$ ;  $p_F = (p_{0N}, p_{XM}, p_{YM})$ ;  $\rho_F = (\rho_{0N}, \rho_{XM}, \rho_{YM})$ .
6. Применение *стандартного* алгоритма оптимизации к тройке векторов ( $c_F$ ,  $p_F$ ,  $\rho_F$ ) для нахождения вектора  $g_F$  весовых коэффициентов оптимального комбинированного портфеля с разбиением на три подвектора:  $g_F = (g_{0F}, g_{XF}, g_{YF})$ .

*Примечание.* В наборах, получаемых в п. 5, в результате обнуления в п. 1 и п. 2 содержится, как правило, значительная доля нулевых троек. Но это не сказывается на окончательном результате п. 6, так как этим тройкам придается нулевой вес. Зато такой вариант алгоритма упрощает вычисления. Тем не менее из

соображений строгости и чистоты алгоритма, а также буквального следования представлению (12), но уже с векторами  $g^{cmb}$ ,  $g_X^{cmb}$  и  $g_Y^{cmb}$ , все подвергшиеся обнулению элементы нужно было бы удалить из наборов с запоминанием их положения для последующего выхода на представление (12).

7. Вычисление записей основных результатов инвестиции  $L = \langle A, R, y \rangle$  ( $A$  — инвестиционная сумма;  $R$  — средний доход;  $y$  — средняя доходность) и построение графиков платежных функций портфелей (11)–(14) как по частям, так и совместно (в варианте идеалистичного портфеля).
8. Построение суррогатного портфеля путем формирования стоимостного  $e^{srq}$  по формуле (15) и прогнозного  $p^{srq} (\equiv p)$  векторов размерности  $n \times m$ , образованием из них вектора относительных доходов  $p^{srq}$  той же размерности с последующим применением *стандартного* алгоритма.

## 5 Иллюстративные примеры

Рассматриваются примеры с  $X = Y = [0, 1]$ , при этом  $x, s \in X$ ,  $y, t \in Y$ . В них используется функция рискованных предпочтений инвестора  $\phi(\varepsilon) = \varepsilon^2$ ,  $\varepsilon \in [0, 1]$ , а при задании стоимостных и вероятностных характеристик существенная роль отводится двухпараметрическому бета-распределению  $Be(u, v)$  с плотностью

$$\beta(x; u, v) = \frac{x^{u-1}(1-x)^{v-1}}{B(u, v)}, \quad u, v > 0,$$

где  $B(u, v) = \int_0^1 x^{u-1}(1-x)^{v-1} dx = \Gamma(u)\Gamma(v)/\Gamma(u+v)$  и  $\Gamma(u) = \int_0^\infty x^{u-1} \exp(-x) dx$  — соответственно бета- и гамма-функции.

Для рынка #0 с помощью бета-плотности предварительно строится нормированная двумерная плотность  $f(x, y; u, v)$ :

$$\left. \begin{aligned} f'(x, y; u, v) &= \begin{cases} x^{u-1}(1-y)^{v-1}, & x \leq y; \\ y^{u-1}(1-x)^{v-1}, & x > y; \end{cases} \\ N(u, v) &= \int_0^1 \int_0^1 f'(x, y; u, v) dx dy; \\ f(x, y; u, v) &= \frac{\int_0^x \int_0^y f'(x, y; u, v) dx dy}{N(u, v)}, \end{aligned} \right\} (17)$$

а затем стоимостную и прогнозную плотности для этого рынка задаем соответственно в виде выпуклых комбинаций:

$$c(x, y) = \omega_c f(x, y; 2, 8; 1, 8) + (1 - \omega_c) \beta(x; 3; 3) \beta(y; 3; 3); \quad (18)$$

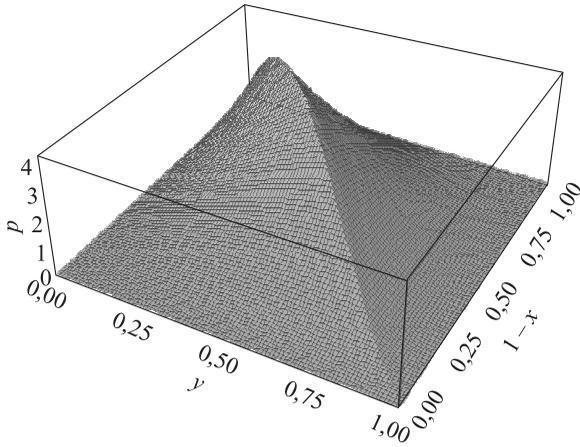


Рис. 1 Плотность  $p(1-x, y)$

$$p(x, y) = \omega_p f(x, y; 3; 2) + (1 - \omega_p) \beta(x; 3; 3) \beta(y; 3; 3),$$

$$\omega_p = 0,5, \quad \omega_c = 0,5. \quad (19)$$

(С ростом параметра  $\omega_p \in [0, 1]$  корреляция между базовыми активами увеличивается.) На рис. 1 приводится график плотности  $p(x, y)$ .

Для большей наглядности график изображается с заменой  $x \rightarrow 1 - x$ . Различие между плотностями (18) и (19) весьма незначительно, и потому график  $c(x, y)$  не приводится. Отметим лишь, что отличие графика плотности  $c(x, y)$  проявляется в меньшем угле излома на диагонали квадрата  $[0, 1] \times [0, 1]$  и снижении максимальных значений. Такое соотношение плотностей говорит о большем разбросе стоимостной плотности в сравнении с прогнозной. Подобная задача на языке финансового рынка означает *продажу волатильности*.

Для стоимостных плотностей рынков #X и #Y принимается также

$$c_X(x) = \chi_X (\omega_x c_1(x) + (1 - \omega_x) \beta(x; 2; 2)),$$

$$\omega_x = 0,9; \quad (20)$$

$$c_Y(y) = \chi_Y (\omega_y c_2(y) + (1 - \omega_y) \beta(y; 2; 2)),$$

$$\omega_y = 0,9. \quad (21)$$

При  $\chi_X = \chi_Y = 1$  эти функции приобретают все свойства плотности вероятности.

В отличие от стоимостных плотностей прогнозное распределение вероятности, как и в [6], едино для всех рассматриваемых рынков, и потому  $p_X(x) \equiv p_1(x); p_Y(y) \equiv p_2(y)$ .

Дискретизация множества X проводится при  $n = 13$ , а Y — при  $m = 12$ , порождая тем самым  $n \times m = 156$  двумерных сценариев, при этом стоимостные и прогнозные характеристики определяются для всех сценариев, одномерных и двумерных, интегрированием функций (17)–(21). Отметим лишь, что для дискретной модели при нахождении маргинальных цен и вероятностей од-

номерных сценариев достаточно ограничиться однократным суммированием по  $j \in J$  и  $i \in I$  соответственно цен и вероятностей для двумерных сценариев.

Заметим еще, что при интегрировании плотности типа бета-распределения в пределах крайних сценариев при некоторых значениях параметров из интервала  $(-1, 0)$  приходится иметь дело с интегрируемой особенностью в точке  $x = 1$ . Для обеспечения корректности счета имеет смысл выделять из плотности «чистую» особенность, допускающую аналитическое интегрирование. Остаточный интеграл особенностей уже не содержит и без труда приближенно интегрируется. Так, если  $v \in (-1, 0)$  и  $u > 0$ , то при  $0 < \tau < 1$  имеет место:

$$\int_{\tau}^1 x^u (1-x)^v dx =$$

$$= \int_{\tau}^1 (1-x)^v dx + \int_{\tau}^1 (x^u - 1)(1-x)^v dx =$$

$$= \frac{1 - \tau^{v+1}}{v+1} + \int_{\tau}^1 (x^u - 1)(1-x)^v dx.$$

По полученным вероятностям и ценам находятся относительные доходы для рынков #0, #X и #Y, необходимые для проведения операций замещения. При этом двумерные массивы трансформируются в векторы с лексикографическим упорядочением компонент.

Применением стандартного дискретного алгоритма оптимизации ко всем исходным рынкам по отдельности можно находить для них векторы весов оптимальных портфелей и доходности инвестиции. Так, для рынков #0, #X и #Y при  $\chi_X = 1$  и  $\chi_Y = 1$  они соответственно равны 0,0423358, 0,0313433 и 0,0312055. Их можно использовать в качестве ориентиров для сравнения результатов.

При построении комбинированного портфеля используется алгоритм разд. 4. С фиксацией  $\chi_X = 1$  и  $\chi_Y = 1$  находится матрица замещений A (4):

0	0	2	2	2	2	2	2	2	2	2	2
0	0	0	2	2	2	2	2	2	2	2	1
1	0	0	2	2	2	2	2	2	2	1	1
1	0	0	0	2	2	2	2	2	1	1	1
1	1	0	0	0	2	2	2	1	1	1	1
1	1	1	0	0	0	1	1	1	1	1	1
1	1	1	1	0	0	1	1	1	1	1	1
1	1	1	1	1	0	0	0	1	1	1	1
1	1	1	1	1	2	2	0	0	0	1	1
1	1	1	1	2	2	2	2	0	1	1	1
1	1	2	2	2	2	2	2	0	1	1	1
1	2	2	2	2	2	2	2	2	2	2	2

(в ней 35 нулей, 60 единиц и 61 двойка), а также векторы (5) и (6):

$$\begin{aligned} \mathbf{a}_X &= \{0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}; \\ \mathbf{a}_Y &= \{0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}. \end{aligned}$$

По матрице  $\mathbf{A}$  и распределению вероятностей на сценариях (19) (двумерных и одномерных) определяются необходимые для построения комбинированного портфеля векторы  $\theta_X$  и  $\theta_Y$  параметров рандомизации, состоящие из вероятностей переключения на одномерные части портфеля и вычисляемые по формулам (9) и (11):

$$\begin{aligned} \theta_X &= \{0; 0,0062; 0,0139; 0,0434; 0,1385; 0,2017; \\ &0,4047; 0,532; 0,539; 0,309; 0,2489; 0,0788; 0,0012\}; \\ \theta_Y &= \{0; 0,0051; 0,0353; 0,0617; 0,1669; 0,3336; \\ &0,4396; 0,4511; 0,4556; 0,1992; 0,0842; 0,1241\}. \end{aligned}$$

Результатом реализации п. 5 алгоритма оптимизации становится тройка векторов ( $c_F$ ,  $p_F$ ,  $\rho_F$ ) размерности  $nm + n + m = 181$  каждый, а п. 6 — весовой вектор  $g_F$  той же длины с сохранением всех элементов обнуления. В них первые  $nm$  элементов отвечают рынку #0, следующие  $n$  — рынку #X, последние  $m$  — рынку #Y. Разбиение вектора  $g_F$  на такие три блока дает векторы весов для трех частей комбинированного портфеля с представлением, аналогичным (12) и фактически эквивалентным ему (см. примечание к п. 6).

С точностью до  $10^{-4}$  для рынков #0, #X и #Y соответственно:

$$\begin{aligned} g_{0F} &= \{0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0,0006; 0,0007; \\ &0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0,0011; 0,0162; 0,0107; 0; 0; 0; \\ &0; 0; 0; 0; 0; 0,001; 0,0363; 0,6184; 0,0454; 0; 0; 0; 0; \\ &0; 0; 0; 0; 0,0096; 0,5743; 0,7013; 0,1704; 0; 0; 0; 0; \\ &0; 0; 0; 0; 0,0411; 0,6762; 0,7346; 0,5057; 0; 0; 0; 0; \\ &0; 0; 0; 0; 0,158; 0,8035; 0,8754; 0,5568; 0; 0; 0; 0; 0; \\ &0; 0; 0; 0; 0,6035; 0,9589; 0,9157; 0; 0; 0; 0; 0; 0; 0; \\ &0; 0; 0,6483; 1; 0,7662; 0; 0; 0; 0; 0; 0; 0; 0; 0,529; \\ &0,8348; 0,1819; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0,1073; 0; \\ &0; 0; 0; 0; 0; 0; 0; 0; 0,0005; 0; 0; 0; 0; 0; 0; 0; \\ &0; 0; 0; 0\}; \\ g_{XF} &= \{0; 0; 0; 0,0008; 0,0083; 0,034; 0,1451; 0,4798; \\ &0,2442; 0,0257; 0,0034; 0,0001; 0\}; \\ g_{YF} &= \{0; 0; 0,0001; 0,0015; 0,0151; 0,099; 0,3114; \\ &0,3862; 0,0724; 0,006; 0,0002; 0\}. \end{aligned}$$

В векторе  $g_{0F}$  содержатся 35 подвергшихся обнулению элементов, относящихся к множеству  $M_0$ . Наличие в нем прочих нулевых элементов есть результат округления; они вносят свой вклад в общий доход комбинированного портфеля, хотя и с незначительным весом.

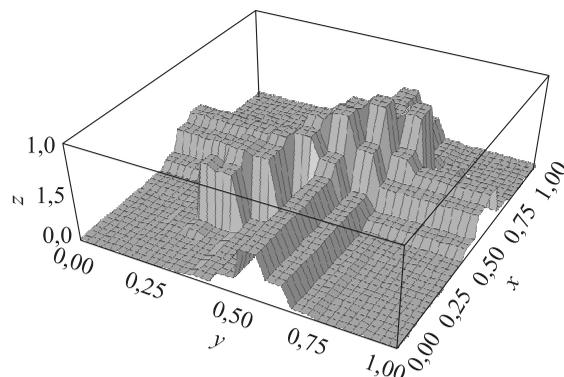


Рис. 2 Доходы идеалистичного портфеля

Запись результатов для комбинированного портфеля:

$$L^{cmb} = \langle 0,335345; 0,351642; 0,048599 \rangle,$$

а график его платежной функции изображен на рис. 2 в идеалистичной версии.

С помощью матрицы  $\mathbf{A}$  находится и вектор весовых коэффициентов для суррогатного портфеля (16) и его запись результатов:

$$\begin{aligned} g^{srg} &= \{0; 0; 0,0001; 0,0011; 0,0108; 0,0726; 0,2446; \\ &0,3317; 0,0455; 0,0035; 0,0001; 0; 0; 0,0006; 0,0007; \\ &0,0015; 0,0113; 0,0914; 0,247; 0,3342; 0,0724; 0,0036; \\ &0,0001; 0; 0; 0,0011; 0,0162; 0,0107; 0,0126; 0,076; \\ &0,253; 0,3402; 0,0473; 0,0039; 0,0001; 0,0000; 0,0007; \\ &0,001; 0,0363; 0,6184; 0,0454; 0,082; 0,2632; 0,3505; \\ &0,0503; 0,0045; 0,0008; 0,0008; 0,0071; 0,0076; 0,0096; \\ &0,5743; 0,7013; 0,1704; 0,2781; 0,3653; 0,0547; 0,007; \\ &0,0082; 0,0083; 0,0259; 0,0268; 0,0292; 0,0411; 0,6762; \\ &0,7346; 0,5057; 0,3293; 0,0606; 0,0321; 0,0336; 0,034; \\ &0,1095; 0,1092; 0,1142; 0,1223; 0,158; 0,8035; 0,8754; \\ &0,5568; 0,1328; 0,1401; 0,144; 0,1451; 0,3867; 0,3902; \\ &0,3983; 0,412; 0,4316; 0,6035; 0,9589; 0,9157; 0,4537; \\ &0,4691; 0,4775; 0,4798; 0,2442; 0,1839; 0,1886; 0,1966; \\ &0,2081; 0,223; 0,6483; 1; 0,7662; 0,235; 0,2418; 0,2438; \\ &0,022; 0,0167; 0,0178; 0,0197; 0,025; 0,0892; 0,2935; \\ &0,529; 0,8348; 0,1819; 0,0219; 0,0257; 0,0015; 0,0016; \\ &0,0018; 0,0022; 0,0141; 0,0964; 0,3114; 0,3788; 0,0668; \\ &0,1073; 0,0031; 0,0034; 0; 0; 0,0001; 0,0013; 0,0149; \\ &0,099; 0,2991; 0,3862; 0,0704; 0,0056; 0,0005; 0,0001; \\ &0; 0; 0,0001; 0,0013; 0,0151; 0,0898; 0,3006; 0,3313; \\ &0,0715; 0,006; 0,0002; 0\}; \end{aligned}$$

$$L^{srg} = \langle 0,324923; 0,340849; 0,0490131 \rangle.$$

Здесь не приводится график доходов суррогатного портфеля, так как он весьма похож на график рис. 2 (до степени смешения), хотя и не тождествен ему. Расхождения объясняются тем, что в суррогатном портфеле (в отличие от идеалистичного)



4. *Агасандян Г. А.* Континуальный критерий VaR на сценарных рынках // Информатика и её применения, 2018. Т. 12. Вып. 1. С. 32–40.
5. *Агасандян Г. А.* Континуальный критерий VaR и оптимальный портфель инвестора // Управление большими системами, 2018. Вып. 73. С. 6–26.
6. *Агасандян Г. А.* Теоретические основы оптимизации по CC-VaR на совокупности рынков // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 36–41.
7. *Кramer Г.* Математические методы статистики / Пер. с англ. — М.: Мир, 1975. 750 с. (*Cramer H.* Mathematical methods of statistics. — Princeton, NJ, USA: Princeton University Press, 1946. 575 p.)

Поступила в редакцию 21.11.19

## COMPUTATIONAL ASPECTS OF OPTIMIZATION ON CC-VaR IN A COMPLEX OF MARKETS

G. A. Agasandyan

A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The work is the direct continuation of the previous author’s investigation on using continuous VaR-criterion (CC-VaR) in a set of markets of different dimensions, which are mutually connected by their underliers. The exposition is aimed at the application of ideas and methods developed for the theoretical continuous model to discrete scenarios markets. In a typical model case of a collection of one two-dimensional market and two one-dimensional markets, a rule of constructing a combined portfolio in these markets is submitted. This rule gives a necessary and sufficient condition of portfolio optimality in the weighted composition of basis instruments. The condition is founded on misbalance in returns relative between markets with maintaining optimality on CC-VaR. The optimal combined portfolio with three components is constructed. Also, the idealistic and surrogate versions of this combined portfolio, which are useful in testing all algorithmic calculations and in graphic illustrating portfolio’s payoff functions, are adduced. The model can be extended without difficulties, theoretic anyway, on markets of greater dimensions.

**Keywords:** underlie; risk preferences function; continuous VaR-criterion; cost and forecast densities; return relative function; Newman–Pearson procedure; combined portfolio; surrogate portfolio

**DOI:** 10.14357/19922264200309

### Acknowledgments

The reported study was funded by RFBR, project No. 17-01-00816.

### References

1. Agasandyan, G. A. 2002. Optimal behavior of an investor in option market. *Joint Conference (International) on Neural Networks Proceedings*. Honolulu, HI: IEEE. 1859–1864.
2. Agasandyan, G. A. 2011. *Применение континуального критерия VaR на финансовых рынках* [Application of continuous VaR-criterion in financial markets]. Moscow: CC RAS. 299 p.
3. Agasandyan, G. A. 2015. *Континуальный критерий VaR на многомерных рынках опционов* [Continuous VaR-criterion in multidimensional option markets]. Moscow: CC RAS. 297 p.
4. Agasandyan, G. A. 2018. Континуальный критерий VaR на сценарных рынках [Continuous VaR-criterion in scenario markets]. *Информатика и её Применения — Inform. Appl.* 12(1):32–40.
5. Agasandyan, G. A. 2018. Континуальный критерий VaR и оптимальный портфель инвестора [Continuous VaR-criterion and investor’s optimal portfolio]. *Управление большими системами* [Large-Scale Systems Control] 73:6–26.
6. Agasandyan, G. A. 2018. Теоретические основы оптимизации по континуальному критерию VaR на совокупности рынков [Theoretical foundations of continuous VaR optimization in the collection of markets]. *Информатика и её Применения — Inform. Appl.* 13(4):36–41.
7. Cramer, H. 1946. *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press. 575 p.

Received October 21, 2019

### Contributor

**Agasandyan Gennady A.** (b. 1941) — Doctor of Science in physics and mathematics, leading scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; agasand17@yandex.ru

## МЕТОДЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В ЗАДАЧЕ ПОИСКА ИНСАЙДЕРА\*

А. А. Грушо<sup>1</sup>, М. И. Забейло<sup>2</sup>, Д. В. Смирнов<sup>3</sup>, Е. Е. Тимонина<sup>4</sup>, С. Я. Шоргин<sup>5</sup>

**Аннотация:** Исследованы подходы к выявлению враждебных инсайдеров организации, использующих сговор. Проблема выявления организованной группы нарушителей информационной безопасности — одна из самых сложных задач обеспечения безопасности организации. Исходное множество данных для анализа состоит из множества малых выборок, описывающих функционал информационных технологий (ИТ) организации. Это множество можно считать большими данными. Для сокращения объема исходных данных использован метод кластеризации. Это позволило эффективно использовать методы математической статистики, т. е. выявить малые выборки, несущие информацию о враждебных инсайдерах. Сложность задачи заключалась в том, чтобы как можно меньше потерять искомым малым выборкам. Найдены условия, когда в схеме серий вероятность выявления инсайдеров, использующих сговор, стремится к 1.

**Ключевые слова:** выявление организованной группы враждебных инсайдеров; малые выборки; большие данные; математическая статистика

**DOI:** 10.14357/19922264200310

### 1 Введение

Сбор ценной информации сотрудниками, использующими личные связи, — серьезная проблема информационной безопасности в государственных и коммерческих организациях. При этом часто нарушитель информационной безопасности сам имеет доступ к части ценной информации (ЦИ), но не имеет права получать ее целиком. Примером такой ситуации может служить разделение общей проблемы для ее решения в разных подразделениях организации. Тогда враждебный инсайдер (далее — инсайдер) попытается узнать недостающую ЦИ от своих друзей в других подразделениях или войти в сговор о сборе и продаже ЦИ.

Исследованию проблемы выявления инсайдеров посвящено множество научных работ.

В работе [1] приведен ряд вызовов, исследование которых авторы данного обзора считают важными, но трудными проблемами. Вызов 4 (см. [1]) связан с проблемой выявления инсайдеров, организованных с помощью преступного сговора.

Выявление инсайдеров в течение ряда лет финансировалось американским управлением иссле-

дованиями Министерства обороны США (DARPA, Defence Advanced Research Projects Agency). Основное направление этих исследований связано с выявлением аномалий в больших данных. В частности, в работе [2] разработан язык описания аномалий и представлена информация по его использованию.

В работах [3, 4] для выявления слабо выраженных аномалий, порожденных инсайдерами, предложено использование информации из нескольких информационных пространств.

Важной проблемой остается определение условий, при которых выявление инсайдера возможно. Все методы в данной области строятся на основе некоторых предположений, выполнение которых необходимо для их работоспособности. Далее рассматривается проблема выявления инсайдера в банковской сфере.

В статье предполагается, что цели инсайдера состоят в сборе и продаже ЦИ о клиентах банка. Как правило, такой информацией выступают персональные данные вместе с ЦИ о счетах и движении денежных средств. Для того чтобы эти данные могли дать постоянный доход инсайде-

\* Работа частично поддержана РФФИ (проект 18-29-03081).

<sup>1</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, grusho@yandex.ru

<sup>2</sup> Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, m.zabzhailo@yandex.ru

<sup>3</sup> ПАО Сбербанк России, Департамент кибербезопасности, dvlsmirnov@sberbank.ru

<sup>4</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, eltimon@yandex.ru

<sup>5</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru

ру, они должны собираться в достаточно большом объеме.

Для защиты от таких инсайдеров можно разделить информацию, например на персональные данные и ЦИ на счетах. При этом работать с каждым из этих блоков данных могут только разные менеджеры (сотрудники банка). Отсюда возникает задача выявления возможного сговора каких-то менеджеров, занимающихся либо персональными данными, либо информацией о счетах. Статья посвящена анализу условий выявления такого сговора.

## 2 Формальная модель сговора инсайдеров

Для исследования задачи выявления инсайдеров, использующих сговор, рассмотрим следующую модель. Обозначим через  $V$  множество клиентов, использующих сервисы банка, через  $U$  — множество менеджеров (пользователей банковской системы и реализующих ИТ). Хранилище данных, которое используют менеджеры, устроено следующим образом. Все данные заносятся в прямоугольную таблицу, столбцы которой нумеруются атрибутами  $A_0, A_1, \dots$ , а строки содержат записи данных и результатов действий с ними. Строки не изменяются, но при изменениях данных появляются новые строки с новыми данными. В отличие от традиционных реляционных баз данных записи не обязательно содержат значения всех атрибутов и поиск по таблице осуществляется по сложному ключу, зависящему от типа данных в строке. Однако каждая строка содержит данные о менеджере, осуществляющем обращение, идентификаторе экземпляра ИТ, которая связана с обращением, и санкционированием текущей транзакции от предыдущих шагов ИТ. Часть атрибутов  $\vec{A}_1$  соответствует содержанию персональных данных, нужных для авторизации транзакций, часть атрибутов  $\vec{A}_2$  соответствует ЦИ (данные счетов, переводы и др.).

Политика безопасности запрещает одному менеджеру иметь доступ одновременно к  $\vec{A}_1$  и к  $\vec{A}_2$ . Поэтому строки, соответствующие использованию атрибутов  $\vec{A}_2$ , обладают индексами, скрывающими данные, соответствующие атрибутам  $\vec{A}_1$ . Кроме того, выделены пользователи  $U_1$ , которые могут иметь доступ к информации с атрибутами  $\vec{A}_1$ , и пользователи  $U_2$ , не имеющие доступа к данным с атрибутами  $\vec{A}_1$ , но работающие с данными, соответствующими атрибутам  $\vec{A}_2$ .

Интересная для инсайдеров ЦИ может быть получена тогда и только тогда, когда известны значения атрибутов  $\vec{A}_1$  и соответствующие им значе-

ния атрибутов  $\vec{A}_2$ . Для получения такой информации необходим сговор какого-либо пользователя из множества  $U_1$  с каким-либо пользователем из множества  $U_2$ . Задача состоит в выявлении таких инсайдеров.

Пусть  $u_1 \in U_1$  и  $u_2 \in U_2$  образуют такую пару  $(u_1, u_2)$  инсайдеров, использующих сговор.

Сделаем дополнительные предположения о том, что пара  $(u_1, u_2)$  может так управлять потоком заявок клиентов на сервисы, что пара  $(u_1, u_2)$  появляется чаще, чем все остальные пары из множества  $U_1 \times U_2$ . Тогда можно пытаться использовать методы математической статистики для выявления пары  $(u_1, u_2)$ .

Однако для большого значения числа элементов в множестве  $U_1 \times U_2$  выявление пары  $(u_1, u_2)$  упирается в проблему многих малых выборок [5, 6].

Пусть статистические критерии выявления  $(u_1, u_2)$  имеют ошибки  $\alpha > 0, \beta > 0$  где  $\alpha$  — вероятность «ложной» тревоги;  $\beta$  — вероятность пропуска  $(u_1, u_2)$ . Тогда при многих малых выборках анализ «ложных» тревог становится трудоемкой задачей, а вероятность выявления пары  $(u_1, u_2)$  ограничена снизу константой (приблизительно 1/3). Поэтому для эффективного применения статистических методов необходимо снижать объем множества малых выборок. Иными словами, необходимо применять статистические методы в рамках некоторых кластеров ограниченного набора малых выборок.

В данной статье используется параметр, управляющий объемом множества малых выборок и позволяющий повышать эффективность статистического анализа. Этим параметром служит объем множества клиентов, в интересах которых проводятся транзакции, интересные для  $(u_1, u_2)$ , т.е. вместе с каждой парой  $(u, u')$ ,  $u \in U_1, u' \in U_2$ , необходимо рассматривать параметр  $v$  со значениями в  $V$ , идентифицирующий клиента, который инициирует транзакцию. Обозначим эту тройку  $(u, u')|v$ .

Инсайдеров  $(u_1, u_2)$  интересуют клиенты из множества  $V$ , продажа информации о которых имеет высокую стоимость. Вместе с тем если инсайдеров интересуют только такие клиенты (обозначим множество таких клиентов через  $V_1$ ) и их транзакции, то количество троек  $(u, u')|v$ , где  $v \in V_1$ , для анализа сокращается, что позволяет повысить эффективность статистических методов.

## 3 Анализ выявляемости сговора инсайдеров

Пусть транзакция обрабатывается случайной парой менеджеров  $(u, u')$ ,  $u \in U_1, u' \in U_2$ . Положим  $|U_1| = n_1, |U_2| = n_2, |U_1 \times U_2| = n_1 n_2 = n$ .

Тогда вероятность появления пары  $(u, u')$  равна  $1/n$ .

Если  $(u_1, u_2)$  — инсайдеры, использующие сговор, и клиент  $v$  представляет для них интерес, то вероятность того, что эта пара будет обслуживать этого клиента равна

$$P((u_1, u_2) | v) = p,$$

а в случае любой другой пары  $(u, u') \neq (u_1, u_2)$  эта вероятность равна

$$P((u, u') | v) = (1 - p) \frac{1}{n - 1}.$$

С помощью методов кластеризации можно выделить множество  $V_1$  тех клиентов, которые представляют интерес для инсайдеров, использующих сговор. Ясно, что мощность множества  $V_1$  много меньше, чем мощность множества  $V$ . Как было отмечено выше, это позволяет избежать эффекта малых выборок [6].

Пусть  $C_1$  — среднее число транзакций у клиентов из множества  $V_1$ . Тогда средний объем данных для клиентов из множества  $V_1$  равен  $|V_1|C_1 = m$ . Далее считаем  $m$  известным параметром схемы. Тогда соотношение параметров  $m$ ,  $n$  и  $p$  определяет возможность выявления инсайдеров, использующих сговор. Инсайдеры  $(u_1, u_2)$  не могут переключить на себя весь поток клиентов из множества  $V_1$ . Поэтому для данных  $\xi$ , которые они получили, можно предложить модель биномиального распределения:

$$P(\xi(u_1, u_2) = k) = \binom{m}{k} p^k (1 - p)^{m-k}.$$

Тогда для любой пары  $(u, u') \neq (u_1, u_2)$  распределение числа случаев предоставления сервисов описывается следующим образом:

$$P(\xi(u, u') = r) = \sum_{k=0}^m \binom{m}{k} p^k (1 - p)^{m-k} \times \binom{m-k}{r} \left(\frac{1}{n-1}\right)^r \left(1 - \frac{1}{n-1}\right)^{m-k-r}.$$

Поскольку числа  $m$  и  $n$  являются большими, то рассмотрение задачи выявляемости инсайдеров  $(u_1, u_2)$  в множестве  $V_1$  будет вестись в терминах асимптотических распределений вероятностей в схеме серий, т.е. в предположении, что  $m \rightarrow \infty$ ,  $n \rightarrow \infty$  и  $p \rightarrow 0$ . Задача состоит в поиске значений вероятности  $p$ , при которых вероятность выявления инсайдеров  $(u_1, u_2)$  стремится к 1.

В практическом плане интересен случай, когда  $m \rightarrow \infty$ ,  $n \rightarrow \infty$  и  $\alpha = m/n$  удовлетворяет условию  $\alpha/\ln(n) \rightarrow 0$ . Это условие соответствует преобладанию числа предоставления сервисов для

клиентов из множества  $V_1$  над числом пар менеджеров  $n_1 n_2 = n$ , которые проводят это обслуживание. В каком-то смысле множитель  $\ln n$  можно интерпретировать как число обслуживаний на одну пару менеджеров, не являющихся инсайдерами  $(u_1, u_2)$ , использующими сговор.

Рассмотрим случайную величину  $\eta_{(n)}$ , равную максимальному числу обслуживаний клиентов из множества  $V_1$  по всем парам менеджеров  $(u, u')$ . Обозначим

$$\alpha = \frac{m}{n}; \quad p_k = \frac{\alpha^k}{k!}, \quad k = 0, 1, \dots$$

В работе [7] приведена следующая теорема.

**Теорема [7].** Если  $m, n \rightarrow \infty$ ,  $\alpha/\ln(n) \rightarrow 0$  и  $r = r(\alpha, n)$  выбрано так, что  $r > \alpha$  и  $np_r \rightarrow \lambda$ , где  $\lambda$  — положительная постоянная, то

$$P\{\eta_{(n)} = r - 1\} \rightarrow e^{-\lambda}; \quad P\{\eta_{(n)} = r\} \rightarrow 1 - e^{-\lambda}.$$

Рассмотрим простой пример применения этой теоремы. Пусть

$$n \frac{\alpha^3}{6} e^{-\alpha} \rightarrow \lambda > 0,$$

т.е.  $r = r(\alpha, n) = 3$ . Тогда максимальное число обслуживаний клиентов из множества  $V_1$  по всем парам менеджеров  $(u, u') \neq (u_1, u_2)$  равно 3 с вероятностью, стремящейся к 1.

Превышение числа появлений пары  $(u_1, u_2)$ , начиная с 4, однозначно идентифицирует эту пару инсайдеров с вероятностью, стремящейся к 1. Вероятность того, что пара  $(u_1, u_2)$  в определенной выше схеме появится не больше трех раз, равна

$$(1-p)^m + mp(1-p)^{m-1} + \frac{m(m-1)}{2} p^2 (1-p)^{m-2} + \frac{m(m-1)(m-2)}{6} p^3 (1-p)^{m-3}.$$

Отсюда следует, что при  $p \leq \ln(m)/m$  пара инсайдеров  $(u_1, u_2)$ , использующих сговор, выявляется с вероятностью, стремящейся к 1.

## 4 Заключение

В статье показано, как сочетать методы кластеризации и методы математической статистики для выявления инсайдеров, использующих сговор, а именно: для эффективного применения математической статистики на множестве разнородных данных, которые состоят из множества малых выборок, необходимо приближенное выполнение соотношения между параметрами схемы.

Однако согласование параметров возможно, когда существует возможность управления объемами данных. Для такого управления можно использовать методы кластеризации.

В рассмотренной задаче можно выделить систему вложенных кластеров данных, повышающих шансы выявления инсайдеров, использующих сговор. Выделение таких кластеров позволило доказать пренебрежимо малую вероятность превышения построенного порога для числа случайных повторений встречаемости пар менеджеров, не являющихся инсайдерами, использующими сговор.

Вместе с тем при полученной отсюда оценке вероятности встречаемости инсайдеров, использующих сговор, их идентификация происходит с вероятностью, стремящейся к 1.

## Литература

1. *Gheyas I., Abdallah A.* Detection and prediction of insider threats to cyber security: A systematic literature review and meta-analysis // *Big Data Anal.*, 2016. Vol. 1. P. 1–29. doi: 10.1186/s41044-016-0006-0.
2. *Memory A., Goldberg H.G., Senator T.E.* Context-aware insider threat detection // *Workshops at 27th AAAI Conference on Artificial Intelligence*, 2013. P. 44–47. <https://pdfs.semanticscholar.org/04aa/e6d97900ba62e90b07ac682fb7bd8c2e1029.pdf>.
3. *Грушо А.А., Забейжайло М.И., Смирнов Д.В., Тимонина Е.Е.* Модель множества информационных пространств в задаче поиска инсайдера // *Информатика и её применения*, 2017. Т. 11. Вып. 4. С. 65–69.
4. *Grusho A., Grusho N., Timonina E.* Method of several information spaces for identification of anomalies // *Intelligent distributed computing XIII* / Eds. I. Kotenko, C. Badica, V. Desnitsky, D. El Baz, M. Ivanovic. — *Studies in computational intelligence ser.* — Springer, 2020. Vol. 868. P. 515–520. doi: 10.1007/978-3-030-32258-8\_60.
5. *Axelsson S.* The base-rate fallacy and the difficulty of intrusion detection // *ACM T. Inform. Syst. Ser.*, 2000. Vol. 3. Iss. 3. P. 186–205.
6. *Grusho A., Grusho N., Timonina E.* The bans in finite probability spaces and the problem of small samples // *Distributed computer and communication networks* / Eds. V. Vishnevskiy, K. Samouylov, D. Kozyrev. — *Lecture notes in computer science ser.* — Springer, 2019. Vol. 11965. P. 578–590.
7. *Колчин В.Ф., Севастьянов Б.А., Чистяков В.П.* Случайные размещения. — М.: Наука, 1976. 224 с.

Поступила в редакцию 02.06.20

---

---

## MATHEMATICAL STATISTICS IN THE TASK OF IDENTIFYING HOSTILE INSIDERS

N. A. Grusho<sup>1</sup>, M. I. Zabezhailo<sup>2</sup>, D. V. Smirnov<sup>3</sup>, E. E. Timonina<sup>1</sup>, and S. Ya. Shorgin<sup>1</sup>

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

<sup>2</sup>A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

<sup>3</sup>Sberbank of Russia, 19 Vavilov Str., Moscow 117999, Russian Federation

**Abstract:** The paper explores approaches to identifying hostile insiders of the organization using collusion. The problem of identifying the organized group of information security violators is one of the most complex tasks of ensuring the security of organization. The set of source data for analysis consists of many small samples describing the functionality of the organization’s information technologies. This set can be considered as big data. The clustering method is used to reduce the amount of source data that made it possible to use mathematical statistics efficiently, i. e., to identify small samples carrying information about hostile insiders. The difficulty of the task was to lose as little as possible the needed small samples. The conditions have been found where in the series scheme, the probability of identifying insiders using collusion tends to 1.

**Keywords:** identification of the organized group of hostile insiders; small samples; big data; mathematical statistics

**DOI:** 10.14357/19922264200310

## Acknowledgments

The work was partially supported by the Russian Foundation for Basic Research (project 18-29-03081).

## References

1. Gheyas, I., and A. Abdallah. 2016. Detection and prediction of insider threats to cyber security: A systematic literature review and meta-analysis. *Big Data Anal.* 1:1–29. doi: 10.1186/s41044-016-0006-0.
2. Memory, A., H.G. Goldberg, and T.E. Senator. 2013. Context-aware insider threat detection. *Workshops at 27th Conference on Artificial Intelligence.* 44–47. Available at: <https://pdfs.semanticscholar.org/04aa/e6d97900ba62e90b07ac682fb7bd8c2e1029.pdf> (accessed August 13, 2020).
3. Grusho, A.A., M.I. Zabezhailo, D.V. Smirnov, and E.E. Timonina. 2017. Model' mnozhestva informatsionnykh prostranstv v zadache poiska insaydera [The model of the set of information spaces in the problem of insider detection]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):65–69.
4. Grusho, A., N. Grusho, and E. Timonina. 2020. Method of several information spaces for identification of anomalies. *Intelligent distributed computing XIII.* Eds. I. Kotenko, C. Badica, V. Desnitsky, D. El Baz, and M. Ivanovic. Studies in computational intelligence ser. Springer. 868:515–520.
5. Axelsson, S. 2000. The base-rate fallacy and its implications for the difficulty of intrusion detection. *ACM T. Inform. Syst. Ser.* 3(3):186–205.
6. Grusho, A., N. Grusho, and E. Timonina. 2019. The bans in finite probability spaces and the problem of small samples. *Distributed computer and communication networks.* Eds. V.M. Vishnevskiy, K.E. Samouylov, and D.V. Kozyrev. Lecture notes in computer science ser. Springer. 11965:578–590.
7. Kolchin, V.F., B.A. Sevastyanov, and V.P. Chistyakov. 1976. *Sluchaynye razmeshcheniya* [Random allocations]. Moscow: Nauka. 224 p.

*Received June 2, 2020*

## Contributors

**Grusho Alexander A.** (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; [grusho@yandex.ru](mailto:grusho@yandex.ru)

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

**Smirnov Dmitry V.** (b. 1984) — business partner for IT security department, Sberbank of Russia, 19 Vavilov Str., Moscow 117999, Russian Federation; [dvlsmirnov@sberbank.ru](mailto:dvlsmirnov@sberbank.ru)

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

**Shorgin Sergey Ya.** (b. 1952) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; [sshorgin@ipiran.ru](mailto:sshorgin@ipiran.ru)

## ВЫЯВЛЕНИЕ АНОМАЛИЙ С ПОМОЩЬЮ МЕТАДАННЫХ\*

А. А. Грушо<sup>1</sup>, Е. Е. Тимонина<sup>2</sup>, Н. А. Грушо<sup>3</sup>, И. Ю. Терехина<sup>4</sup>

**Аннотация:** Рассматривается проблема контроля безопасности информационных технологий (ИТ) на основе данных компьютерного аудита. Эти данные представляют собой последовательность малых выборок, каждая из которых описывает передачу информации от одних преобразований к другим. Информационные технологии представляются математическими моделями в виде ориентированных ациклических графов. Такие графы, описывающие передачу данных, в статье называются метаданными (МД). В единых данных компьютерного аудита могут одновременно присутствовать следы выполнения нескольких ИТ, описываемые своими графами. Это обстоятельство затрудняет распознавание информационных потоков (ИП), которые соответствуют дугам разных графов. В работе введено понятие легальных ИП, которые соответствуют передаче данных всех выполняемых ИТ. Информационные потоки, не соответствующие выполнению действующих ИТ, называются нелегальными, или аномалиями. Такие ИП могут возникать из-за враждебной деятельности инсайдеров или из-за ошибок действий пользователей. В статье на основе МД решена задача эффективного выявления легальных ИП и аномалий.

**Ключевые слова:** информационная безопасность; информационный поток; аномалии; метаданные; системы различных представителей

**DOI:** 10.14357/19922264200311

### 1 Введение

В работе рассмотрена проблема выявления нелегальных ИП с помощью МД. Пусть распределенная информационно-вычислительная система (РИВС) реализует ряд ИТ, поддерживающих работу организации, например банка. Каждая ИТ формируется на основе бизнес-процесса, и в РИВС одновременно функционируют несколько ИТ. В работах [1, 2] модель произвольной ИТ была представлена в виде ориентированного ациклического графа (DAG — Directed Acyclic Graph). В этих моделях вершины обозначаются греческими буквами и соответствуют преобразованиям информации (блокам преобразований) вместе с входными и выходными данными. Преобразования обозначаются строчными латинскими буквами, а входные и выходные данные — прописными латинскими буквами. Дуги DAG соответствуют передаче данных блокам от предыдущих блоков, т. е. дуга выходит из выходных данных выполненного преобразования и входит во входные данные для следующего преобразования. Существуют дуги, входящие

в вершины DAG как исходные данные извне ИТ, и существуют выделенные дуги, идущие вне DAG к потребителям результатов ИТ. Все указанные дуги фиксированы и считаются легальными. Определенные таким образом модели ИТ будем называть метаданными.

Одной из важных угроз организации представляется возможность существования ИП, которые используются для утечки ценной информации из ИТ. Далее ИП, не участвующие в функционирующих ИТ, но исходящие из данных ИТ или входящие в них, будем называть нелегальными ИП, или аномалиями. В работе поиск аномальных ИП проводится исходя из противоречий с МД, определяемых с помощью DAG. Исходными данными для анализа служат результаты компьютерного аудита [3, 4], в которых отражены имена субъектов, инициировавших ИП, идентификаторы получателей ИП и время производимых действий. Таким образом, последовательность легальных ИП соответствует МД.

Проблема восстановления структуры ИП по данным логов рассматривалась в ряде научных ра-

\* Работа частично поддержана РФФИ (проект 18-07-00274).

<sup>1</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, grusho@yandex.ru

<sup>2</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, eltimon@yandex.ru

<sup>3</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, info@itake.ru

<sup>4</sup>Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, irina\_teryokhina@mail.ru

бот, например в [5, 6]. В качестве модели структур, восстанавливаемых по логам, предлагались сети Петри. Однако модели сетей Петри представляют собой сложные структуры, и поэтому их восстановление требует сложных алгоритмов. Кроме того, для сетей Петри не найдены алгоритмы иерархической декомпозиции. В то же время иерархическая декомпозиция DAG [2] позволяет отслеживать не все ИП (часть ИП невозможно отследить).

В работе рассматривается задача восстановления множества DAG, соответствующих действующим в данный период времени ИТ. Отметим связь между МД, рассматривавшимися в предыдущих работах [7–9], касавшихся контроля соединений в сети. Если ИП проходил через сеть, то его содержание могло быть зашифровано, т. е. определялись входные и выходные блоки сети, где формировались сообщения для передачи по сети, и в этих блоках использовались криптографические средства. Протокол организации сетевого обмена мог опираться на несколько удаленных субъектов, однако это учитывалось в организации МД [10].

Отметим, что при определении легальности ИП необходимо учитывать присутствие дуг в других ИТ, а также возможность кратного использования одной ИТ. В этом множестве ИП необходимо найти нелегальные ИП.

В работе рассматривается проведение анализа по поиску нелегальных ИП в режиме офлайн с использованием следующей идеи. Для восстановления всех действующих в данное время ИТ предлагается использовать метод трансверсалей [11, 12]. Трансверсали (системы различных представителей, СРП) позволяют быстро и однозначно восстанавливать DAG ИТ, выполняемых в данное время, т. е. по СРП восстанавливаются МД, в соответствии с которыми поступают данные аудита. После восстановления МД из данных аудита убирается информация о легальных ИП. Тогда оставшиеся ИП оказываются аномалиями, которые подлежат дальнейшему анализу службы безопасности.

## 2 Выявление нелегальных информационных потоков при наличии системы различных представителей

Пусть в рассматриваемых данных функционировали ИТ  $IT_1, \dots, IT_L$ . Этим технологиям соответствуют МД  $G_1, \dots, G_L$ , при этом множеству DAG  $G_1, \dots, G_L$  соответствуют множества вершин  $V_1, \dots, V_L$  и множества дуг  $E_1, \dots, E_L$ . В тех случаях,

когда в ИТ вносятся внешние данные, можно формально определить вершину  $\alpha_0$ , из которой выходит дуга в соответствующую вершину из множества  $V$ , причем эту дугу можно включить в множество  $E$ . Аналогично если предусматривается легальный ИП из  $G$ , то можно ввести формальную вершину  $\alpha_1$ , в которую из соответствующей вершины графа  $G$  направлена дуга с выходящим из  $G$  ИП.

Воспользуемся определением СРП и формулировкой теоремы Ф. Холла из работы [11]. Пусть  $E$  — множество всех дуг графов  $G_1, \dots, G_L$ ,  $P(E)$  — множество подмножеств множества  $E$ ,  $D = (e_1, \dots, e_L)$  — выборка из множества  $E$ .

**Определение.** Пусть  $L$  элементов выборки  $D$  отличны друг от друга и  $e_i \in E_i, i = 1, \dots, L$ . Тогда  $D$  называется *системой различных представителей* множеств  $E_1, \dots, E_L$ .

Условия существования СРП получены в теореме Ф. Холла [11].

**Теорема 1** [11]. *Множества  $E_1, \dots, E_L$  имеют СРП тогда и только тогда, когда для каждого  $k = 1, \dots, L$  и всех сочетаний индексов  $i_1, \dots, i_k$  выполняются условия:*

$$\left| \bigcup_{j=1}^k E_{i_j} \right| \geq k.$$

Пусть множества  $E_1, \dots, E_L$  удовлетворяют теореме 1.

**Теорема 2.** *В последовательности данных обязательно встретятся все элементы  $e_1, \dots, e_s$  из СРП тогда и только тогда, когда одновременно функционируют различные ИТ  $IT_{i_1}, \dots, IT_{i_s}$ .*

**Доказательство. Достаточность.** При анализе данных в режиме офлайн в последовательности данных все запущенные ИТ можно считать завершившимися. Тогда в каждом графе  $G_{i_j}, j = 1, \dots, s$ , должны быть пройдены все дуги. Значит, должны быть пройдены все дуги, принадлежащие СРП.

**Необходимость.** Если в последовательности данных встретился элемент СРП, то однозначно определен граф, которому принадлежит эта дуга. Это означает, что однозначно определена ИТ, моделью которого служит этот граф. Теорема доказана.

**Следствие.** Если в последовательности данных встретился элемент СРП, то в данных должны встретиться все дуги, принадлежащие соответствующему графу модели ИТ.

Отсюда следует алгоритм определения легальных ИП. Сначала в данных находятся все элементы СРП, по ним восстанавливаются соответствующие DAG и ИТ. Далее выявляются все дуги этих DAG в данных. Построенное множество ИП является легальным для всех восстановленных DAG.

Как было отмечено выше, одновременно могут функционировать несколько копий одной ИТ. В условиях существования СРП выявление легальных ИП в этом случае не выглядит сложным. Пусть одна ИТ запущена  $r$  раз, тогда элемент СРП соответствующего DAG встретится  $r$  раз. Это означает, что необходимо собрать  $r$  экземпляров данного DAG. Очевидно, что в этих графах дуги могут быть переставлены и относиться к разным экземплярам DAG, но несмотря на это совокупность восстановленных графов будет правильно определять легальные ИП.

В случае существования СРП удаление легальных ИП из последовательности данных определяет множество нелегальных ИП, анализировать которые необходимо отдельно. Отсюда следует, что СРП значительно упрощает поиск нелегальных ИП.

### 3 Определение легальных информационных потоков в случае отсутствия системы различных представителей

Рассмотрим случай, когда в системе множеств  $E_1, \dots, E_L$  нет СРП. Это означает, что существует набор подмножеств дуг  $E_{i_1}, \dots, E_{i_k}$ , не удовлетворяющих теореме 1, т. е.

$$\left| \bigcup_{j=1}^k E_{i_j} \right| < k.$$

Тогда имеет место следующая цепочка неравенств:

$$k > |E_{i_1} \cup \dots \cup E_{i_k}| \geq |E_{i_1} \cup \dots \cup E_{i_{k-1}}| \geq \dots \geq |E_{i_1}| = t > 0.$$

В этой цепочке  $k$  неравенств, из которых строгими могут быть не более  $k - t$ , остальные  $t$  неравенств являются равенствами. Пусть для некоторого  $1 < s < k$  справедливо равенство:

$$|E_{i_1} \cup \dots \cup E_{i_s}| = |E_{i_1} \cup \dots \cup E_{i_{s-1}}|.$$

Отсюда следует, что

$$E_{i_s} \subseteq E_{i_1} \cup \dots \cup E_{i_{s-1}},$$

т. е. дуги графа  $G_{i_s}$  являются частью объединения дуг других DAG.

Найденные условия позволяют реализовать другой подход к использованию СРП. Избавляясь от таких включений, разобьем множество  $E_1, \dots, E_L$

на  $M$  групп  $E_1^{(1)}, \dots, E_{L_1}^{(1)}, E_1^{(2)}, \dots, E_{L_2}^{(2)}, \dots, E_1^{(M)}, \dots, E_{L_M}^{(M)}$ , в каждой из которых есть СРП. Тогда встреча дуги в  $i$ -й СРП позволяет восстановить один из DAG этой группы. Если эта дуга встречается в нескольких СРП, то в каждой группе она восстанавливает DAG. Если дуга из данного СРП не позволила восстановить DAG, то эта дуга может принадлежать другой группе DAG. Тогда СРП этой другой группы позволит восстановить DAG, в который войдет указанная дуга. Если дуга  $i$ -й СРП не породила из данных аудита DAG этой группы и эта дуга не попала ни в один из DAG других групп, то это — нелегальный ИП.

Основной критерий завершенности выявления легальных ИП — это полное восстановление всех DAG, элементы которых возникли из разных СРП с включенными в них дугами СРП других групп. Данный метод повышает эффективность анализа данных, если групп мало. Однако следующий пример показывает, что метод работает при любом числе групп.

**Пример.** Пусть ИТ порождают графы  $G_1, \dots, G_L$  и множества дуг  $E_1, \dots, E_L$ . Пусть каждое множество дуг есть отдельная группа, тогда каждая дуга может быть взята как СРП в своей группе. Если встретилась дуга из  $E_i$  и граф  $G_i$  был восстановлен из данных, то множества ИП дуг графа  $G_i$  являются легальными. Если эта же дуга принадлежит другому графу  $G_j$  со своим СРП и граф  $G_j$  полностью восстановлен, то все его дуги также порождают легальные ИП. Если дуга не попала ни в один из восстановленных графов  $G_i$ , то она представляет нелегальный ИП. После восстановления  $G_i$  нельзя исключать дуги из рассмотрения, так как другой граф может оказаться недостроенным. Поэтому дуги, соответствующие восстановленным легальным ИП, до окончания работы алгоритма необходимо «раскрашивать».

### 4 Заключение

В множестве ИТ, модели которых представимы в виде DAG (МД), можно ввести понятие легальных и нелегальных ИП. Тогда каждая ИТ может быть представлена множеством дуг соответствующего DAG. Легальность ИП может быть проверена на основе СРП, по которым сами DAG восстанавливаются однозначно. Нелегальные ИП определяются после устранения всех легальных ИП из данных компьютерного аудита.

Этот метод может быть распространен на случай, когда в исходной информационной системе не существует СРП.

## Литература

1. Самуйлов К. Е., Чукарин А. В., Яркина Н. В. Бизнес-процессы и информационные технологии в управлении телекоммуникационными компаниями. — М.: Альпина Паблишерс, 2009. 442 с.
2. Грушо Н. А., Грушо А. А., Забейайло М. И., Тимонина Е. Е. Методы нахождения причин сбоев в информационных технологиях с помощью метаданных // Информатика и её применения, 2020. Т. 14. Вып. 2. С. 33–39.
3. Department of Defense Trusted Computer System Evaluation Criteria. DoD 5200.28-STD. — U.S. National Institute of Standards and Technology, Department of Defense, 1985. <http://csrc.nist.gov/publications/history/dod85.pdf>.
4. Грушо А. А., Применко Э. А., Тимонина Е. Е. Теоретические основы компьютерной безопасности. — М.: Академия, 2009. 272 с.
5. Aalst W., Weijters T., Maruster L. Workflow mining: Discovering process models from event logs // IEEE T. Knowl. Data En., 2004. Vol. 16. Iss. 9. P. 1128–1142.
6. Bezerra F., Weiner J. Algorithms of anomaly detection of traces in logs of process aware information systems // Inform. Syst., 2013. Vol. 38. Iss. 1. P. 33–44.
7. Grusho A., Grusho N., Timonina E. Detection of anomalies in non-numerical data // 8th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings. — Piscataway, NJ, USA: IEEE, 2016. P. 273–276.
8. Грушо А. А., Тимонина Е. Е., Шоргин С. Я. Иерархический метод порождения метаданных для управления сетевыми соединениями // Информатика и её применения, 2018. Т. 12. Вып. 2. С. 44–49.
9. Grusho A., Grusho N., Timonina E. Information flow control on the basis of meta data // Distributed computer and communication networks, 22nd International Conference, Revised Selected Papers / Eds. V. M. Vishnevskiy, K. E. Samouylov, and D. V. Kozyrev. — Lecture notes in computer science ser. — Springer, 2019, Vol. 11965. P. 548–562.
10. Grusho A. A., Timonina E. E., Shorgin S. Ya. Modelling for ensuring information security of the distributed information systems // 31th European Conference on Modelling and Simulation Proceedings. — Dudweiler, Germany: Digitaldruck Pirrot GmbH, 2017. P. 656–660.
11. Райзер Г. Дж. Комбинаторная математика / Пер. с англ. К. А. Рыбникова. — М.: Мир, 1966. 154 с. (Ryser H. J. Combinatorial mathematics. — New York, NY, USA: Wiley, 1963. 154 p.)
12. Холл М. Комбинаторика / Пер. с англ. С. А. Широковой. — М.: Мир, 1970. 424 с. (Hall M., Jr. Combinatorial theory. — New York, NY, USA: Wiley, 1967. 424 p.)

Поступила в редакцию 15.07.20

## IDENTIFYING ANOMALIES USING METADATA

A. A. Grusho<sup>1</sup>, E. E. Timonina<sup>1</sup>, N. A. Grusho<sup>1</sup>, and I. Yu. Teryokhina<sup>2</sup>

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

<sup>2</sup>Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

**Abstract:** The paper discusses the problem of information technology security control based on computer audit data. These data are the sequence of small samples, each of which describes the transmission of information from one transformation to another. Information technologies are represented by mathematical models in the form of oriented acyclic graphs. In the article, such graphs describing data transmission are called metadata. In integrated computer audit data, there may be at the same time traces of the execution of several information technologies described by their graphs. This makes it difficult to recognize information flows that correspond to arcs of different graphs. The concept of legal information flow is introduced in the paper, which corresponds to the transfer of data of all information technologies being performed. Information flows that do not correspond to the execution of existing information technologies are called illegal or anomalies. Such information flows can occur due to hostile activities of insiders or due to errors in user actions. The article solves the problem of effective identification of legal information flows and anomalies on the basis of metadata.

**Keywords:** information security; information flow; anomalies; metadata; systems of different representatives

**DOI:** 10.14357/19922264200311

## Acknowledgments

The paper was partially supported by the Russian Foundation for Basic Research (project 18-07-00274).

## References

1. Samuylov, K. E., A. V. Chukarin, and N. V. Yarkina. 2009. *Biznes-protsessy i informatsionnye tekhnologii v upravlenii telekommunikatsionnymi kompaniyami* [Business processes and information technologies in management of the telecommunication companies]. Moscow: Alpina Pabls. 442 p.
2. Grusho, N. A., A. A. Grusho, M. I. Zabezhailo, and E. E. Timonina. 2020. Metody nakhozhdeniya prichin sboev v informatsionnykh tekhnologiyakh s pomoshch'yu metadannykh [Methods of finding the causes of information technology failures by means of meta data]. *Informatika i ee Primeneniya — Inform. Appl.* 14(2):33–39.
3. DoD 5200.28-STD. 1985. Department of Defense Trusted Computer System Evaluation Criteria. Available at: <http://csrc.nist.gov/publications/history/dod85.pdf> (accessed July 14, 2020).
4. Grusho, A. A., E. A. Primenko, and E. E. Timonina. 2009. *Teoreticheskie osnovy komp'yuternoy bezopasnosti* [Theoretical bases of computer security]. Moscow: Academy. 272 p.
5. Aalst, W., T. Weijters, and L. Maruster. 2004. Workflow mining: Discovering process models from event logs. *IEEE T. Knowl. Data En.* 16(9):1128–1142.
6. Bezerra, F., and J. Weiner. 2013. Algorithms of anomaly detection of traces in logs of process aware information systems. *Inform. Syst.* 38(1):33–44.
7. Grusho, A., N. Grusho, and E. Timonina. 2016. Detection of anomalies in non-numerical data. *8th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings*. Piscataway, NJ: IEEE. 273–276.
8. Grusho, A. A., E. E. Timonina, and S. Ya. Shorgin. 2018. Ierarkhicheskiy metod porozhdeniya metadannykh dlya upravleniya setevymi soedineniyami [Hierarchical method of meta data generation for control of network connections]. *Informatika i ee Primeneniya — Inform. Appl.* 12(2):44–49.
9. 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, 22nd International Conference, Revised Selected Papers*. Eds. V. M. Vishnevskiy, K. E. Samouylov, and D. V. Kozyrev. Lecture notes in computer science ser. Springer. 11965:548–562.
10. Grusho, A. A., E. E. Timonina, and S. Ya. Shorgin. 2017. Modelling for ensuring information security of the distributed information systems. *31th European Conference on Modelling and Simulation Proceedings*. Dudweiler, Germany: Digitaldruck Pirrot GmbH. 656–660.
11. Ryser, H. J. 1963. *Combinatorial mathematics*. New York, NY: Wiley. 154 p.
12. Hall, M., Jr. 1967. *Combinatorial theory*. New York, NY: Wiley. 424 p.

Received July 15, 2020

## Contributors

**Grusho Alexander A.** (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; [grusho@yandex.ru](mailto:grusho@yandex.ru)

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

**Teryokhina Irina Yu.** (b. 1994) — PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; [irina\\_teryokhina@mail.ru](mailto:irina_teryokhina@mail.ru)

# АППРОКСИМАЦИЯ МНОЖЕСТВА ДОСТИЖИМЫХ ПОТОКОВ МНОГОПОЛЬЗОВАТЕЛЬСКОЙ СЕТИ

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

**Аннотация:** Рассматривается метод приближенного описания выпуклого многогранного множества допустимых потоков, передаваемых между всеми узлами сети одновременно. Предлагается способ построения внутреннего выпуклого аппроксимирующего каркаса. Каркас формируется на основе векторов предельно достижимых потоков между парами вершин источник–приемник. Система опорных векторов определяется точками, лежащими на внешних гранях исходного множества. Любая выпуклая комбинация базовых векторов задает допустимое распределение потоков. Разработанные алгоритмические схемы допускают распараллеливание вычислительных процедур на гетерогенных многопроцессорных комплексах. Полученное агрегированное описание можно использовать при диспетчеризации интенсивных входных информационных потоков, превышающих функциональные возможности сети.

**Ключевые слова:** многопродуктовая потоковая модель; множество допустимых потоков; внутренний опорный каркас

**DOI:** 10.14357/19922264200312

## 1 Введение

В работе рассматривается общая схема построения приближенного описания множества достижимых потоков [1] в многопользовательской сетевой системе [2, 3]. В рамках формальной математической записи системы ограничений исследуются векторы различных потоков, которые могут одновременно передаваться между всеми парами источник–приемник. Вычисляются граничные точки, которые соответствуют предельно достижимому значению потока для некоторой пары узлов при фиксированных значениях для всех остальных. Полученные наборы точек на границе множества достижимости определяют опорные векторы внутреннего каркаса. Любая выпуклая комбинация базисных векторов задает допустимое распределение потоков в исходной сети. Предложенную алгоритмическую схему можно рассматривать как многоэтапную динамическую процедуру диспетчеризации в условиях неопределенности [4, 5].

## 2 Математическая модель

Для описания многопользовательской сетевой системы используется математическая запись модели передачи многопродуктового потока [5]. Структура сети  $G(\mathbf{d})$  задается множествами  $\langle V, R, U \rangle$ : узлов (вершин) сети  $V = \{v_1, v_2, \dots$

$\dots, v_n, \dots, v_N\}$ ; неориентированных ребер  $R = \{r_1, r_2, \dots, r_k, \dots, r_E\} \subset V \times V$ . Ребро  $r_k = (v_{n_k}, v_{j_k})$  соединяет вершины  $v_{n_k}, v_{j_k}$  (инцидентно вершинам  $v_{n_k}, v_{j_k}$ ), которые для него служат концевыми. Предполагается, что в сети отсутствуют петли и сдвоенные ребра. Ребру  $r_k$  ставятся в соответствие две ориентированные дуги  $u_k$  и  $u_{k+E}$  прямого и обратного направления из множества ориентированных дуг  $U = \{u_1, u_2, \dots, u_k, \dots, u_{2E}\}$ . Дуги  $\{u_k, u_{k+E}\}$  определяют направление передачи потока по ребру  $r_k$  между концевыми вершинами  $v_{n_k}$  и  $v_{j_k}$ .

Обозначим через  $S(v_n)$  множество номеров исходящих дуг, по которым поток покидает узел  $v_n$ ;  $T(v_n)$  — множество номеров входящих дуг, по которым поток поступает в узел  $v_n$ . Состав множеств  $S(v_n)$  и  $T(v_n)$  однозначно определяется в ходе выполнения следующей процедуры. Пусть некоторое ребро  $r_k \in R$  соединяет вершины с номерами  $n$  и  $j$ , такими что  $n < j$ . Тогда ориентированная дуга  $u_k = (v_n, v_j)$  считается *исходящей* из вершины  $v_n$  и ее номер  $k$  заносится в множество  $S(v_n)$ , ориентированная дуга  $u_{k+E} = (v_j, v_n)$  — *входящей* для  $v_n$  и ее номер  $k + E$  помещается в список  $T(v_n)$ . Соответственно дуга  $u_k$  является *входящей* для  $v_j$  и ее номер  $k$  попадает в  $T(v_j)$ , а дуга  $u_{k+E}$  — *исходящей* и номер  $k + E$  вносится в список исходящих дуг  $S(v_j)$ .

Предполагается, что в сети между любой парой узлов могут передаваться потоки разных видов.

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

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

Пара узлов-корреспондентов  $p_m$  определяется упорядоченной парой вершин  $p_m = (v_{s_m}, v_{t_m})$ , где вершина с номером  $s_m$  называется источником, а с номером  $t_m$  — стоком, или приемником, потока  $m$ -го вида. В настоящей работе в сети из  $N$  узлов рассматривается  $M = N(N - 1)$  независимых, невзаимозаменяемых и равноправных потоков различных видов, которые передаются между узлами-корреспондентами из множества  $P = \{p_1, p_2, \dots, p_M\}$ .

Пусть  $x_{mk}$  — величина потока  $m$ -го вида, который передается по ребру  $r_k$ ,  $x_{mk} \geq 0$ ,  $m = \overline{1, M}$ ,  $k = \overline{1, E}$ . 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 все потоки, передаваемые по ребру  $r_k$ :

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

Обозначим через  $z_m$  величину потока  $m$ -го вида, который поступает в сеть в узле с номером  $s_m$  и покидает из узла с номером  $t_m$ .

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

$$\sum_{i \in S(v_n)} x_{mi} - \sum_{i \in T(v_n)} x_{mi} = \begin{cases} z_m, & \text{если } v_n = v_{s_m}; \\ -z_m, & \text{если } v_n = v_{t_m}; \\ 0 & \text{в остальных случаях,} \end{cases} \\ n = \overline{1, N}, \quad m = \overline{1, M}, \quad x_{mi} \geq 0, \quad z_m \geq 0. \quad (2)$$

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

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

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

### 3 Аппроксимация множества допустимых потоков

Для приближенного описания множества допустимых потоков строится внутренний опорный

каркас [2]. Каркас формируется на основе набора точек (векторов), лежащих на границе  $\mathcal{Z}(x, \mathbf{d})$ . На начальном этапе рассматривается произвольная пара узлов корреспондентов  $p_a \in P$  и определяется максимальный поток [3], который можно независимо передать в сети из вершины с номером  $s_a$  в вершину с номером  $t_a$  без учета всех остальных потоков.

**Задача 1.** Найти

$$z_a^0(1) = \max_{(\mathbf{z}, x) \in \mathcal{Z}(x, \mathbf{d})} z_a$$

при дополнительных условиях

$$z_m = 0, \quad m \neq a, \quad m = \overline{1, M}.$$

Вектор решений  $\mathbf{z}_a^0(1) = (0, 0, \dots, z_a^0(1), \dots, 0)$  определяет угловую точку выпуклого множества  $\mathcal{Z}(x, \mathbf{d})$ , которая лежит на пересечении грани  $\mathcal{Z}(x, \mathbf{d})$  с осью координат, соответствующей паре  $p_a$ . Последовательное решение задачи 1 для всех  $p_m \in P$  позволяет построить множество  $\mathbf{Z}^0(1) = \{\mathbf{z}_1^0(1), \mathbf{z}_2^0(1), \dots, \mathbf{z}_m^0(1), \dots, \mathbf{z}_M^0(1)\}$  опорных векторов  $C(1)$ -сечения 1-го слоя.

Векторы из множества  $\mathbf{Z}^0(1)$  можно также рассматривать как направляющие векторы конуса возможных направлений к внешней границе множества  $\mathcal{Z}(x, \mathbf{d})$ . Компоненты вектора  $\mathbf{z}(0) = (z_1^0(1), z_2^0(1), \dots, z_m^0(1), \dots, z_M^0(1))$  численно равны предельно достижимым потокам для каждой пары  $p_m \in P$ . Рассмотрим луч  $\mathbf{z}(\beta) = \beta \mathbf{z}(0)$ , направленный из начала координат в точку  $\mathbf{z}(0)$ . Длину шага по лучу определим как коэффициент разложения в выпуклой комбинации опорных векторов из  $\mathbf{Z}^0(1)$ .

**Задача 2.** Найти  $\beta^*$  и коэффициенты  $\gamma_m^*$ ,  $m = \overline{1, M}$ ,

$$\beta^* \mathbf{z}(0) = \sum_{m=1}^M \gamma_m^* \mathbf{z}_m^0(1) = \sum_{m=1}^M \gamma_m^* \mathbf{z}^0(1)$$

при условии

$$\sum_{m=1}^M \gamma_m^* = 1, \quad \gamma_m^* \geq 0, \quad m = \overline{1, M}.$$

Из покомпонентной записи

$$\beta^* z_m^0(1) = \gamma_m^* z_m^0(1), \quad \sum_{m=1}^M \gamma_m^* = 1, \quad \gamma_m^* \geq 0, \quad m = \overline{1, M},$$

следует, что  $\beta^* = \gamma_m^* = 1/M$ . Координаты точки  $\mathbf{z}(\beta^*)$  в множестве  $\mathcal{Z}(x, \mathbf{d})$  обозначим

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

Вектор  $\mathbf{z}^*(1) = (z_1^*(1), z_1^*(1), \dots, z_m^*(1), z_M^*(1))$  определяет распределение потоков при одновременной передаче потока между всеми парами узлов источник–приемник. Каждая компонента вектора  $\mathbf{z}^*(1)$  пропорциональна максимальному потоку для соответствующей пары  $p_m \in P$ .

На следующем этапе в множестве  $\mathcal{Z}(x, \mathbf{d})$  осуществляется поиск вектора потоков, каждая компонента которого больше или равна достигнутому значению  $z_m^*(1)$ .

**Задача 3.** Для некоторой пары узлов  $p_a$  определить максимальный поток

$$z_a^0(2) = \max_{(\mathbf{z}, x) \in \mathcal{Z}(x, \mathbf{d})} z_a$$

при дополнительных условиях

$$z_a \geq z_a^*(1); \quad z_m \geq z_m^*(1) \quad \forall p_m \in P, \quad m \neq a.$$

Решение задачи 3 — вектор

$$\mathbf{z}_a^0(2) = (z_1^*(1), z_2^*(1), \dots, z_a^0(2), \dots, z_M^*(1)),$$

который лежит на границе множества  $\mathcal{Z}(x, \mathbf{d})$ . На основе решения цепочки задач 3 последовательно для всех  $p_m \in P$  формируется множество

$$\mathbf{Z}^0(2) = \{z_1^0(2), z_2^0(2), \dots, z_m^0(2), \dots, z_M^0(2)\}$$

опорных векторов  $C(2)$ -сечения 2-го слоя или направляющих векторов конуса возможных направлений из точки с координатами  $(z_1^*(1), z_2^*(1), \dots, z_m^*(2), \dots, z_M^*(1))$ .

Любой вектор потоков  $\mathbf{z}$ , который можно представить как выпуклую комбинацию векторов из  $\mathbf{Z}^0(2)$ , является допустимым и принадлежит множеству  $\mathcal{Z}(x, \mathbf{d})$ . Вектор-луч  $\Delta \mathbf{z}(2)$  с компонентами

$$\Delta z_m(2) = z_m^0(2) - z_m^*(1), \quad m = \overline{1, M},$$

определяет возможное направление увеличения достигнутых величин потоков. При этом коэффициент  $\alpha$  — длина шага в возможном направлении при переходе из  $\mathbf{z}^*(1)$  в точку  $\mathbf{z}^*(2)$  — к новому распределению потоков:

$$\mathbf{z}^*(2) = \mathbf{z}^*(1) + \alpha \Delta \mathbf{z}(2).$$

**Задача 4.** Найти  $\alpha^*$  и коэффициенты выпуклой комбинации  $\gamma_m^*$ ,  $m = \overline{1, M}$ , такие что

$$\begin{aligned} \mathbf{z}^*(2) &= \mathbf{z}^*(1) + \alpha^* \Delta \mathbf{z}(2) = \sum_{m=1}^M \gamma_m^* z_m^0(2), \\ \sum_{m=1}^M \gamma_m^* &= 1, \quad \gamma_m^* \geq 0, \quad m = \overline{1, M}, \end{aligned}$$

при покомпонентной записи:

$$\begin{aligned} z_m^*(2) &= z_m^*(1) + \alpha^* (z_m^0(2) - z_m^*(1)) = \\ &= \gamma_m^* z_m^0(2) + \sum_{j=1}^M \gamma_j^* z_m^*(1) - \gamma_m^* z_m^*(1) = \\ &= \gamma_m^* z_m^0(2) + (1 - \gamma_m^*) z_m^*(1); \\ \sum_{m=1}^M \gamma_m^* &= 1, \quad \gamma_m^* \geq 0, \quad m = \overline{1, M}. \end{aligned}$$

Решение задачи 4 — координаты разложения  $\alpha^* = \gamma_m^* = (1/M)$ ,  $m = \overline{1, M}$ , — определяет вектор допустимых потоков  $\mathbf{z}^*(2)$  с координатами:

$$\begin{aligned} z_m^*(2) &= \frac{M-1}{M} z_m^*(1) + \frac{1}{M} z_m^0(2) = \\ &= \left( \frac{1}{M} - \frac{1}{M^2} \right) z_m^0(1) + \frac{1}{M} z_m^0(2) = \\ &= \frac{1}{M} \left[ \frac{M-1}{M} z_m^0(1) + z_m^0(2) \right], \quad m = \overline{1, M}. \end{aligned}$$

После сравнения найденных  $\mathbf{z}^*(1)$  и  $\mathbf{z}^*(2)$  формируется множество пар корреспондентов для поиска вектора  $\mathbf{z}^*(3)$ .

Пусть задано малое положительное число  $\delta > 0$ . Обозначим  $P^-(2)$  — множество пар, для которых на всех последующих этапах фиксируется достигнутый поток  $z_m^*(2)$ ,

$$P^-(2) = \{p_m \mid p_m \in P, |z_m^*(2) - z_m^*(1)| \leq \delta\},$$

$M^-(2)$  — множество индексов пар, входящих в  $P^-(2)$ ,  $M^-(2) = \{m \mid p_m \in P^-(2)\}$ ,  $\mu^-(2)$  — число элементов в  $P^-(2)$  или  $M^-(2)$ . При этом  $P^+(2)$  — множество пар, для которых на следующем, третьем, этапе определяется возможное направление увеличения потоков,

$$P^+(2) = \{p_m \mid p_m \in P, |z_m^*(2) - z_m^*(1)| > \delta\},$$

$M^+(2)$  — множество индексов пар, входящих в  $P^+(2)$ ,  $\mu^+(2) = \{m \mid p_m \in P^+(2)\}$ ,  $\mu^+(2)$  — число элементов в  $P^+(2)$  или  $M^+(2)$ .

**Задача 5.** Для некоторой пары узлов-корреспондентов  $p_a \in P^+(2)$  определить максимальный поток

$$z_a^0(3) = \max_{(\mathbf{z}, x) \in \mathcal{Z}(x, \mathbf{d})} z_a$$

при дополнительном условии  $z_a \geq z_a^*(2)$  и сохранении значений всех остальных потоков  $z_m \geq z_m^*(2)$  для всех пар  $p_m \in P$ ,  $m \neq a$ .

Вектор-решение задачи 5

$$\mathbf{z}_a^0(3) = (z_1^*(2), z_2^*(2), \dots, z_a^0(3), \dots, z_M^*(2))$$

находится на границе множества  $\mathcal{Z}(x, \mathbf{d})$ . Решение цепочки задач 5 для каждого  $p_m \in P^+(2)$  позволяет построить множество опорных векторов

$$\mathbf{Z}^0(3) = \{z_1^0(3), z_2^0(3), \dots, z_m^0(3), \dots, z_{M^+(2)}^0(3)\}$$

$C(3)$ -сечения 3-го слоя.

Компоненты вектора  $\mathbf{z}^*(3)$  определяются по следующему правилу:

$$z_j^*(3) = \begin{cases} z_j^*(2) + \frac{1}{\mu^+(2)} (z_j^0(3) - z_j^*(2)), & \text{если } j \in M^+(2); \\ z_j^*(2), & \text{если } j \in M^-(2). \end{cases}$$

Предположим, что на некотором этапе  $T$  найден вектор  $\mathbf{z}^*(T)$  и при попытке формирования множества  $P^+(T)$  оказалось, что  $P^+(T) = \emptyset$ , поскольку для всех  $p_m \in P$  выполняется неравенство:

$$|z_m^0(T) - z_m^*(T)| \leq \delta, \quad m = \overline{1, M}.$$

Вектор  $\mathbf{z}^*(T)$  лежит *вблизи* границы множества  $\mathcal{Z}(x, \mathbf{d})$  и определяет *приближенное* предельное распределение потоков, при котором нельзя увеличить ни одно из достигнутых значений  $z_m^*(T)$ , не уменьшив при этом поток хотя бы для одной пары узлов-корреспондентов.

На основе вектора  $\mathbf{z}^*(T)$  и всех векторов из  $\mathbf{Z}^0(t)$ ,  $t = \overline{1, T}$ , формируется опорный внутренний каркас:

$$\text{SiF}(\cdot) = \{\mathbf{z}^*(T), \mathbf{Z}^0(1), \mathbf{Z}^0(2), \dots, \mathbf{Z}^0(T)\}$$

(от *англ.* Support-internal-Frame — опорный внутренний каркас). SiF( $\cdot$ )-каркас можно рассматривать как приближенное описание, или аппроксимацию, множества  $\mathcal{Z}(x, \mathbf{d})$ . Любая выпуклая комбинация базисных векторов SiF( $\cdot$ )-каркаса задает допустимое распределение потоков, которые могут одновременно передаваться между всеми парами узлов-корреспондентов.

## 4 Заключение

Разработанную схему вычислений можно использовать как составную часть многоэтапной динамической процедуры диспетчеризации в условиях отсутствия данных о состоянии передающих

подсистем сети и при интенсивных входных потоках [5]. Вектор потоков  $\mathbf{z}^*(1)$ , вычисленный на первом шаге, можно трактовать как априорное допустимое распределение, полученное на основе нормативных показателей работоспособности системы. На втором шаге решается цепочка задач поиска максимального увеличения потока независимо для каждой пары корреспондентов, но при фиксированных достигнутых значениях для всех остальных. Каждая найденная точка лежит на границе множества достижимости и определяет предельно допустимые потоки. Любой набор допустимых потоков приближенно представим выпуклой комбинацией векторов предельно допустимых потоков.

Значение коэффициента разложения меньше единицы будет означать, что пропускные возможности системы не позволяют одновременно передать все запрашиваемые потоки. В результате, согласно диспетчерским правилам, будут введены ограничения на величины передаваемых потоков, а через несколько итераций вновь будут определены новые возможные распределения. Предлагаемый метод позволяет в процессе принятия решений использовать как нормативную, так и текущую информацию о состоянии сети.

## Литература

1. Лотов А. В., Поспелова И. И. Многокритериальные задачи принятия решений. — М.: Макс Пресс, 2008. 197 с.
2. Данциг Дж. Линейное программирование, его применения и обобщения / Пер. с англ. — М.: Прогресс, 1966. 589 с. (Dantzig G. Linear programming and extensions. — Princeton, NJ, USA: Princeton University Press, 1963. 600 p.)
3. Йенсен П., Барнес Д. Потокоевое программирование / Пер. с англ. — М.: Радио и связь, 1984. 392 с. (Jensen P. A., Barnes J. W. Network flow programming. — New York, NY, USA: Wiley, 1980. 408 p.)
4. Малашенко Ю. Е., Назарова И. А., Новикова Н. М. Один подход к анализу возможных структурных повреждений в многопродуктовых сетевых системах // Ж. вычисл. мат. мат. физ., 2019. Т. 59. № 9. С. 1626–1638.
5. Малашенко Ю. Е., Назарова И. А., Новикова Н. М. Экспресс-анализ и агрегированное представление множества достижимых потоков многопродуктовой сетевой системы // Изв. РАН. ТиСУ, 2019. № 6. С. 63–72.

Поступила в редакцию 23.06.20

# APPROXIMATION OF THE MULTIUSER NETWORK FEASIBLE FLOWS SET

Yu. E. Malashenko and I. A. Nazarova

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** A method for approximate description of a convex polyhedral set of feasible flows transmitted between all network nodes simultaneously is considered. A method for constructing an internal convex approximating frame is proposed. The frame is formed based on the vectors of maximum feasible flows between pairs of source–receiver vertices. The system of support vectors is determined by the points lying on the outer edges of the baseline set. Any convex combination of base vectors sets the feasible flow distribution. The developed algorithmic schemes allow parallelization of computational procedures on heterogeneous multiprocessor complexes. The resulting aggregated description can be used for dispatching intensive input information flows that exceed the network’s capability.

**Keywords:** multicommodity flow model; feasible flows set; internal support frame

**DOI:** 10.14357/19922264200312

## References

1. Lotov, A. V., and I. I. Pospelova. 2008. *Mnogokriterial’nye zadachi prinyatiya resheniy* [Multicriteria decision making tasks]. Moscow: Maks Press. 197 p.
2. Dantzig, G. 1963. *Linear programming and extensions*. Princeton, NJ: Princeton University Press. 600 p.
3. Jensen, P. A., and J. W. Barnes. 1980. *Network flow programming*. New York, NY: Wiley. 408 p.
4. Malashenko, Yu. E., I. A. Nazarova, and N. M. Novikova. 2019. An approach to the analysis of possible structural damages in multicommodity network systems. *Comp. Math. Math. Phys.* 59(9):1562–1574.
5. Malashenko, Yu. E., I. A. Nazarova, and N. M. Novikova. 2019. Express analysis and aggregated representation of the set of reachable flows for a multicommodity network system. *J. Comput. Sys. Sc. Int.* 58(6):889–897.

Received June 23, 2020

## Contributors

**Malashenko Yuri E.** (b. 1946) — Doctor of Science in physics and mathematics, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; malash09@ccas.ru

**Nazarova Irina A.** (b. 1966) — Candidate of Science (PhD) in physics and mathematics, scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; irina-nazar@yandex.ru

# БАЙЕСОВСКИЙ ПОДХОД К ПОСТРОЕНИЮ ИНДИВИДУАЛЬНОЙ ТРАЕКТОРИИ ПОЛЬЗОВАТЕЛЯ В СИСТЕМЕ ДИСТАНЦИОННОГО ОБУЧЕНИЯ\*

А. В. Босов<sup>1</sup>, Я. Г. Мартюшова<sup>2</sup>, А. В. Наумов<sup>3</sup>, А. П. Сапунова<sup>4</sup>

**Аннотация:** Рассмотрена задача формирования индивидуальной траектории пользователя системы дистанционного обучения (СДО) при смешанной форме ведения образовательной деятельности с организацией самостоятельной работы обучаемых с помощью СДО. По окончании каждого раздела курса обучения происходит разделение пользователей СДО на категории, определяемые решением задачи байесовской классификации. Для каждой категории предлагается индивидуальное задание разного уровня сложности на следующий раздел курса, и таким образом формируется индивидуальная траектория обучаемого. Настройка байесовского классификатора проводится на основе статистики работы пользователей СДО. Приводятся экспериментальные результаты решения задачи на одном из этапов обучения.

**Ключевые слова:** система дистанционного обучения; байесовский классификатор; адаптивные системы; индивидуальная траектория обучения

**DOI:** 10.14357/19922264200313

## 1 Введение

Современный тренд в обучающих системах, в том числе в СДО, — разработка методов построения индивидуальных траекторий пользователей, учитывающих статистику их работы в системе. Современная теория тестирования (item response theory, IRT) часто использует вероятностные модели для адаптации процесса тестирования под уровень знаний пользователей, учитывая специфику обучаемых [1–8]. В качестве случайных параметров в таких моделях выступают обычно уровень знаний пользователей и уровень сложности заданий [1, 4] или время ответа пользователя на задание [7, 8]. Оценивая эти параметры, система проводит самонастройку контента, адаптируя его под контингент пользователей.

Формирование индивидуальных траекторий тестирования или обучения в СДО [3, 5, 6] подразумевает решение в определенные моменты функционирования системы задачи классификации пользователей — отнесение их к различным категориям (классам) на основе статистики работы в системе. По результатам проведенной классификации разным категориям пользователей предлагаются разные траектории дальнейшего тестиро-

вания или обучения. Решение задачи классификации осуществляется, например, с использованием нейросетевых технологий [9–11] или байесовского подхода [12, 13].

В данной работе в модели, учитывающей смешанную форму обучения с проведением нескольких объективных очных рубежных контролей в течение курса, решается задача построения индивидуальной траектории пользователя в используемой для организации его самостоятельной работы СДО. Индивидуальная траектория корректируется после прохождения пользователем каждого раздела предложенного в СДО курса с помощью построения байесовского классификатора. Для определения соответствующих вероятностей используются классические частотные оценки, построенные на основе статистики работы пользователей СДО.

## 2 Задача построения индивидуальной траектории как задача байесовской классификации

В рамках рассматриваемой модели предполагается, что дистанционное обучение проводится

\* Исследование выполнено при частичной финансовой поддержке РФФИ в рамках научного проекта 18-07-00617.

<sup>1</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский авиационный институт, avbosov@ipiran.ru

<sup>2</sup> Московский авиационный институт, ma1554@mail.ru

<sup>3</sup> Московский авиационный институт, naumovav@mail.ru

<sup>4</sup> Московский авиационный институт, sap2603@mail.ru

совместно с очной формой обучения в виде организации самостоятельной работы студентов посредством использования электронного учебника для выполнения домашнего задания и организации тестирования.

Предлагаемый пользователю курс дистанционного обучения разбит на несколько разделов, которые он должен изучить для освоения данной дисциплины. После каждого раздела пользователю нужно решить тест определенного уровня сложности. Сложность заданий теста определяется экспертом или с помощью специальных алгоритмов, основанных на модели Раша [1]. Для формирования теста из существующих заданий необходимо решить задачу математического программирования, описанную в [14], параметром которой служит суммарный уровень сложности теста. Необходимо определить, какого уровня сложности тест выдавать пользователю в конце каждого раздела, чтобы он соответствовал уровню его подготовки и формировал индивидуальную траекторию обучения.

Решение о том, какой должна быть сложность выдаваемого в конце каждого раздела теста, предлагается принимать экспертно на основе того, к какой объективной категории принадлежит пользователь, например: «отличник», «хорошист» и т. д. Таким образом, на каждом этапе обучения предлагается решать задачу классификации пользователя: отнесение его к одной из категорий. Определяются категории пользователя дискретно в процессе обучения по окончании каждого раздела курса дистанционного обучения. Обозначим моменты времени, когда значение категории может меняться, через  $t = 0, \dots, T$ .

В начальный момент времени  $t = 0$  объективной категории пользователя присваивается некоторая априорная оценка, определяемая на основе предыдущей траектории обучения. Пусть  $t = \{t_1, t_2, t_3\}$  — моменты времени, в которые проводятся очные контрольные мероприятия, на основании которых происходит объективная классификация пользователей СДО с последующей адаптацией их индивидуальных траекторий обучения.

Пусть  $X$  — множество, содержащее  $m$  допустимых категорий пользователя, например  $X = \{2, 3, 4, 5\}$ , где  $m = 4$  — число категорий.

В моменты времени  $t_1, t_2$  и  $t_3$  объективная категория пользователя  $y_{t_i}$  определяется на основе результата очной контрольной работы с помощью традиционной оценки  $k_{t_i} \in Y, i = 1, 2, 3$ , т. е.  $y_{t_i} = k_{t_i}, i = 1, 2, 3$ , где  $k_{t_i} \in Y = \{0, 2, 3, 4, 5\}$ . Случай  $k_{t_i} = 0$  означает, что пользователь не писал контрольную работу в момент времени  $t_i$ . В другие моменты времени объективная категория пользо-

вателя не определена. Однако согласно формуле текущего рейтинга пользователя, предложенной в работе [15] и модифицированной под условия рассматриваемой задачи, можно определить текущую категорию рейтинга  $x_t \in X$  пользователя согласно следующей методике.

Пусть  $B_t$  — текущий (субъективный) рейтинг пользователя в момент времени  $t = 1, \dots, T$ , вычисляемый по формуле:

$$B_t = \frac{\omega_k}{5} \sum_{t_i \leq t} k_{t_i} + \omega_E \frac{\sum_{\tau=1}^t E_{\tau}^+}{\sum_{\tau=1}^t E_{\tau}} - \omega_{\delta} \left( 1 - \frac{\delta_t}{\Delta t} \right),$$

где  $E_{\tau}$  — суммарная сложность теста, выполненного в момент  $\tau$ ;  $E_{\tau}^+$  — суммарная сложность правильно решенных заданий теста, выполненного в момент  $\tau$ ;  $\Delta t$  — общее число очных занятий, прошедших к моменту времени  $t$ ;  $\delta_t$  — число посещений пользователем очных занятий, прошедших к моменту времени  $t$ .

Величины  $\omega_k, \omega_E$  и  $\omega_{\delta}$  — это весовые коэффициенты, которые выбраны так, чтобы максимальное значение текущего рейтинга пользователя в момент времени  $T$  было равно 100. Эти величины отражают важность каждого слагаемого в формуле рейтинга. Например, в [15] выбраны следующие значения:  $\omega_k = \omega_E = 25$  и  $\omega_{\delta} = 15$ . Если при вычислении текущего рейтинга требуется учитывать лишь успехи пользователя в СДО и посещение очных занятий, можно пересчитать коэффициенты, положив  $\omega_k = 0$ .

Пусть  $B_t^{\max}$  — максимальный рейтинг, который может быть набран пользователем к моменту времени  $t$ . Например, для  $t_1 \leq t < t_2$  справедливо  $B_t^{\max} = 50$ . Траектория этой верхней границы, а также несколько примеров траекторий текущего рейтинга  $B_t$  для случая  $T = 17, t_1 = 6, t_2 = 11$  и  $t_3 = 15$  приведены на рис. 1.

Текущая категория рейтинга  $x_t$  для заданного множества  $X = \{2, 3, 4, 5\}$  определяется текущим рейтингом  $B_t$ , например согласно табл. 1, предложенной в [15]. При этом границы изменения рейтинга в каждой из категорий определяются экспертами.

Таким образом, в каждый момент времени  $t$  необходимо по текущей категории рейтинга  $x_t$  принять решение об отнесении пользователя к одной из объективных категорий по результатам следующей за  $t$  контрольной работы, т. е. требуется прогнозировать оценку, которую получит пользователь на следующей контрольной работе, имея текущую категорию рейтинга  $x_t$ . Этот прогноз позволит наилучшим образом подобрать сложность предлагаемого в текущий момент времени теста, чтобы

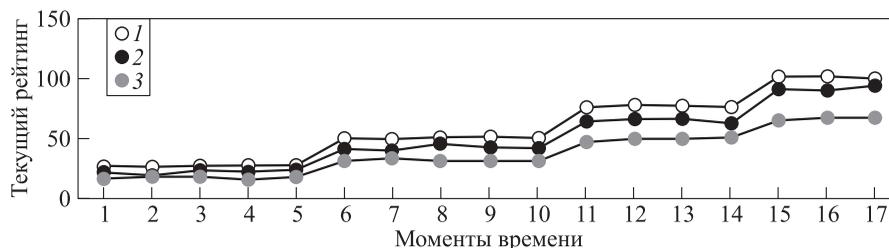


Рис. 1 Траектории текущего рейтинга: 1 — максимальный рейтинг; 2 — рейтинг хорошо успевающего студента; 3 — рейтинг студента, успевающего удовлетворительно

Таблица 1 Соответствие между категорией и рейтингом

$x_t$	$b_t = \frac{B_t}{B_t^{\max}}$
2	[0; 0,6]
3	[0,6; 0,75]
4	[0,75; 0,91]
5	[0,91; 1]

максимально способствовать повышению успеваемости пользователя. Для подбора сложности теста в этой ситуации используются различные педагогические приемы, описание которых выходит за рамки данной статьи. В данной работе для прогнозирования предлагается использовать метод байесовской классификации. Пусть  $P(y_{t_k})$  — вероятность того, что наугад выбранный пользователь в момент времени  $t_k$  проведения контрольной работы, следующей за моментом времени  $t$ , будет иметь объективную категорию рейтинга  $y_{t_k} \in Y$ ;  $P(x_t|y_{t_k})$  — условная вероятность того, что наугад выбранный пользователь в момент времени  $t$  имеет текущую категорию рейтинга, равную  $x_t \in X$ , при условии что в момент времени  $t_k$  проведения контрольной работы, следующей за моментом времени  $t$ , он будет иметь объективную категорию рейтинга  $y_{t_k} \in Y$ .

Рассмотрим алгоритм классификации  $a(\cdot) : X \rightarrow Y$ . Он ставит в соответствие пользователю с текущей категорией рейтинга  $x_t$  объективную категорию  $s$  на момент  $t_k$  следующей контрольной работы. Рассмотрим множество  $A_s = \{x_t \in X | a(x_t) = s, s \in Y\}$ . Вероятность того, что пользователь имеет текущую категорию рейтинга  $x_t \in X$  и алгоритм  $a(\cdot)$  отнесет его к объективной категории  $s$ , равна  $P(y_{t_k})P(A_s|y_{t_k})$ , где  $P(A_s|y_{t_k})$  — условная вероятность событий  $x_t \in A_s$  при условии что в момент времени  $t_k$  пользователь будет иметь объективную категорию рейтинга  $y_{t_k}$ . Каждой паре  $(y_{t_k}, s) \in Y \times Y$  поставим в соответствие величину потери  $\lambda_{y_{t_k}s}$  при отнесении пользователя к объективной категории вместо  $y_{t_k}$ . Тогда функционал

среднего риска — ожидаемая величина потери при классификации алгоритмом  $a(\cdot)$  — будет иметь вид:

$$R(a(\cdot)) = \sum_{s \in Y} \sum_{y_{t_k} \in Y} \lambda_{y_{t_k}s} P(y_{t_k}) P(A_s|y_{t_k}) .$$

Согласно теории байесовской классификации, если известны вероятности  $P(y_{t_k})$  и  $P(x_t|y_{t_k})$ , то минимум среднего риска  $R(a(\cdot))$  достигается алгоритмом [13]:

$$a(x_t) = \arg \min_{s \in Y} \sum_{y_{t_k} \in Y} \lambda_{y_{t_k}s} P(y_{t_k}) P(x_t|y_{t_k}) . \quad (1)$$

Решать задачу оптимизации (1) не представляется возможным, поскольку вероятности  $P(y_{t_k})$  и  $P(x_t|y_{t_k})$  неизвестны. Вместо этого для классификации можно использовать эмпирические оценки вероятностей  $\hat{P}(y_{t_k})$  и  $\hat{P}(x_t|y_{t_k})$ , найденные по обучающей выборке. Под обучающей выборкой понимается статистика работы пользователей СДО, обучающихся по аналогичной программе за определенный промежуток времени. Эта информация становится доступной после выполнения СДО первого курса обучения. Таким образом, указанные оценки можно найти как относительные частоты:

$$\hat{P}(y_{t_k}) = \frac{l(y_{t_k})}{l}, \quad i = 2, 3, 4, 5; \quad \hat{P}(x_t|y_{t_k}) = \frac{l_{y_{t_k}}^{x_t}}{l(y_{t_k})},$$

где  $l(y_{t_k})$  — число пользователей, имевших по результатам теста в СДО в момент  $t_k$  объективную категорию рейтинга  $y_{t_k}$ ;  $l$  — общее число пользователей СДО, обучавшихся по данной программе, по результатам обучающей выборки;  $l_{y_{t_k}}^{x_t}$  — число пользователей СДО, имевших к моменту времени текущую категорию  $x_t$ , у которых объективная категория рейтинга на момент времени  $t_k$  была равна  $y_{t_k}$ .

В силу малого числа элементов множеств  $X$  и  $Y$  задача (1) может быть решена в каждый момент времени  $t$  простым перебором. Результат решения сильно зависит от выбора величины потери  $\lambda_{y_{t_k}s}$ . Традиционно в качестве  $\lambda_{y_{t_k}s}$  выбирается модуль разности  $y_{t_k}$  и  $s$  или его квадрат. Этот

выбор приводит к формированию классификатора на основе условного математического ожидания  $\sum_{y_{t_k} \in Y} y_{t_k} P(y_{t_k} | x_t)$ , симметрично усредняющего ошибки как с завышением, так и с занижением объективного рейтинга пользователя. Однако с точки зрения процесса обучения и построения индивидуальной траектории пользователя представляется, что заметно хуже ошибочно занизить категорию пользователя, чем ошибочно ее завысить. Поэтому предлагается использовать разные штрафы для неположительных  $y_{t_k} - s$ , когда рейтинг ошибочно завывается (например, использовать модуль разности или корень), и для положительных, когда рейтинг ошибочно занижается (можно предложить  $\lambda_{y_{t_k} s} = 2^{y_{t_k} - s}$ , т.е. кратно увеличивать штраф за занижение рейтинга).

Представленный в табл. 1 способ формирования текущей категории рейтинга пользователя отражает его текущую успеваемость по классической пятибалльной шкале. Для более тонкой настройки процесса адаптации индивидуальной траектории пользователя с применением описанного классификатора можно расширить множество возможных значений текущего рейтинга пользователя  $x_t$  до  $K$  элементов, дискретизировав соответствующим образом величину  $b_t \in [0, 1]$ . Например, для  $K = 20$  текущей категории рейтинга можно присваивать значение  $k$ , т.е.  $x_t = k$ , если  $b_t$  будет удовлетворять условию  $(k - 1)0,05 < b_t \leq 0,05k$ ,  $k = 1, \dots, 20$ .

В качестве альтернативы предложенному классификатору можно рассматривать вариант

$$a(x_t) = \arg \max_{y_{t_k} \in Y} P(y_{t_k} | x_t), \quad (2)$$

т.е. оценку, обеспечивающую максимум апостериорной вероятности  $P(y_{t_k} | x_t)$  объективной категории относительно текущего рейтинга  $x_t$ . В реализации эта вероятность также может быть оценена

**Таблица 3** Оценки условного относительно  $x_8$  распределения категории  $y_{t_{10}}$

$x_8$	$y_{t_{10}}$				
	0	2	3	4	5
1	0,30	0,42	0,22	0,05	0,01
2	0,18	0,54	0,23	0,04	0,01
3	0,33	0,36	0,27	0,03	0,01
4	0,11	0,34	0,47	0,06	0,02
5	0,05	0,28	0,42	0,17	0,08
6	0,06	0,24	0,41	0,19	0,10
7	0,05	0,09	0,33	0,32	0,21
8	0,04	0,05	0,31	0,41	0,19
9	0,02	0,06	0,17	0,34	0,41
10	0,01	0,02	0,09	0,39	0,49

соответствующей частотой по обучающей выборке. Это метод, так же как и метод условного среднего, имеет недостаток симметричного учета штрафа за ошибки завышения/занижения рейтинга, что и демонстрирует представленный далее численный эксперимент.

### 3 Результаты численного эксперимента

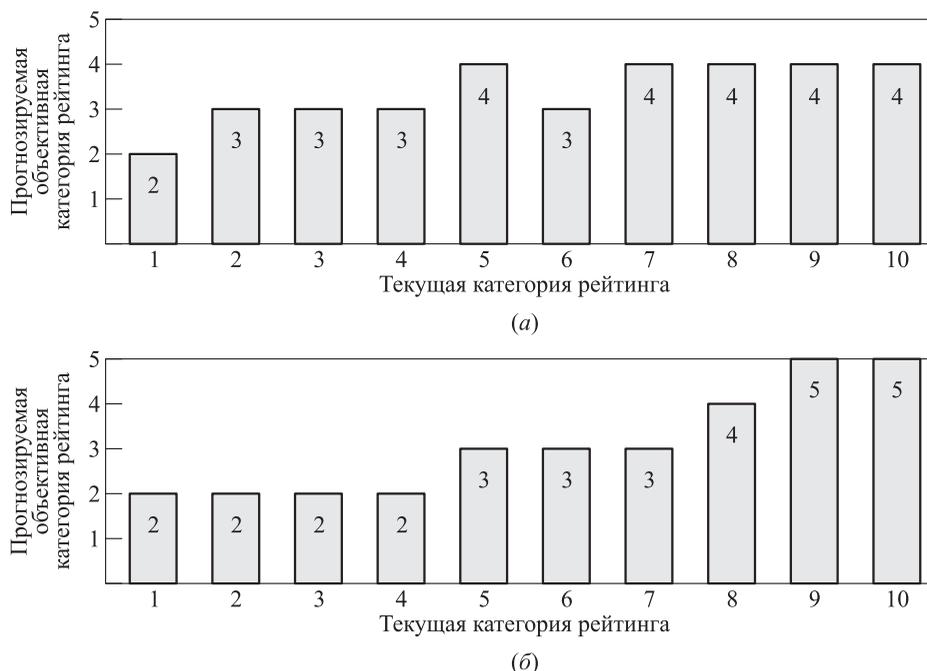
Представленная в статье методика формирования индивидуальной траектории обучения была опробована на примере использования электронного учебника по курсу «Теория вероятностей и математическая статистика» в системе дистанционного обучения CLASS.NET [16] МАИ при очной форме преподавания этой дисциплины студентам инженерных факультетов и наличии трех контрольных работ в процессе обучения. Использовались оба приведенных выше классификатора для числа дискретных значений текущей категории рейтинга  $K = 10$ . Электронный учебник содержит 14 разделов. Очные контрольные работы проводятся после 5-го, 10-го и 14-го разделов. Таким образом,  $T = 17$ . Рассмотрим решение задачи классификации перед 8-м разделом учебника. Соответствующие оценки вероятностей, необходимые для решения задач классификации, отражены в табл. 2–4.

**Таблица 2** Оценки распределения объективной категории  $y_{t_{10}}$

$y_{t_{10}}$	$\hat{P}(y_{t_{10}})$
0	0,09
2	0,11
3	0,38
4	0,28
5	0,14

**Таблица 4** Оценки условного относительно  $y_{t_{10}}$  распределения категории  $x_8$

$y_{t_{10}}$	$x_8$				
	0	2	3	4	5
1	0,11	0,23	0,01	0,01	0
2	0,15	0,09	0,03	0,01	0
3	0,14	0,17	0,02	0,03	0
4	0,09	0,13	0,02	0,02	0,01
5	0,11	0,09	0,10	0,10	0,01
6	0,12	0,05	0,13	0,09	0,01
7	0,09	0,03	0,23	0,14	0,13
8	0,06	0,05	0,24	0,26	0,24
9	0,07	0,09	0,13	0,21	0,31
10	0,06	0,07	0,09	0,13	0,29



**Рис. 2** Результаты несимметричной байесовской классификации (а) и классификации по максимуму апостериорной вероятности (б)

Эти оценки получены по обучающей выборке объемом 1500 элементов, составленной из наблюдений за схожим контингентом пользователей СДО, обучавшихся по аналогичной программе в предыдущем семестре.

Результаты классификации с использованием формулы (1) для случая

$$\lambda_{y_{t_k} s} = \begin{cases} s - y_{t_k}, & y_{t_k} - s \leq 0, \\ 2^{y_{t_k} - s}, & y_{t_k} - s > 0, \end{cases}$$

отражены на диаграмме рис. 2, а, а с использованием формулы (2) — на диаграмме рис. 2, б. По горизонтальной оси расположены значения текущей категории рейтинга  $x_8$ , а по вертикальной оси — объективная категория рейтинга по результатам классификации. Для приведенных исходных данных задачи (1) и (2) имеют единственные решения для каждого значения  $x_8$ .

Результаты использования байесовского классификатора (см. рис. 2, а) выглядят более сбалансированно и в среднем завышенно по сравнению с использованием классификатора, максимизирующего условные вероятности (см. рис. 2, б). Использование весовых коэффициентов  $\lambda_{y_{t_k} s}$  добавляет гибкости инструменту байесовской классификации.

Предложенный байесовский классификатор был использован для построения индивидуальных

траекторий обучения и адаптации функционирования СДО в осеннем семестре 2019 г. в педагогическом эксперименте с использованием 5 экспериментальных групп и 12 контрольных.

Успеваемость в экспериментальных группах по итогам семестра оказалась выше, чем в контрольных группах, в среднем на 10%–15%. При этом наблюдался рост успеваемости в экспериментальных группах по сравнению с итогами их обучения в предыдущем семестре.

## 4 Заключение

В работе предложен адаптивный способ формирования индивидуальной траектории пользователя СДО, используемой при очной форме ведения занятий. Индивидуальность траектории достигается последовательным изменением суммарной сложности теста в СДО в дискретные моменты времени на основе решения задачи классификации пользователя — оценки категории, соответствующей уровню знаний. Адаптивность процесса происходит за счет проведения в некоторые из указанных дискретных моментов времени, соответствующих разделам курса дистанционного обучения, очных контрольных работ, которые позволяют оценить объективную категорию пользователя с последующим использованием этой информации в процессе поэтапной классификации. Результаты про-

веденного численного эксперимента на примере использования СДО МАИ CLASS.NET в процессе преподавания физико-математических дисциплин показали эффективность предложенной методики построения индивидуальной траектории обучения.

## Литература

1. *Rasch G.* Probabilistic models for some intelligence and attainment tests. — Chicago, IL, USA: The University of Chicago Press, 1980. 224 p.
2. *Van der Linden W. J., Scrams D. J., Schnipke D. L., et al.* Using response-time constraints to control for differential speededness in computerized adaptive testing // *Appl. Psych. Meas.*, 1999. Vol. 23. Iss. 3. P. 195–210.
3. *Куравский Л. С., Мармалюк П. А., Алхимов В. И., Юрьев Г. А.* Новый подход к построению интеллектуальных и компетентностных тестов // *Моделирование и анализ данных*, 2013. № 1. С. 4–28.
4. *Кибзун А. И., Иноземцев А. О.* Оценивание уровней сложности тестов на основе метода максимального правдоподобия // *Автоматика и телемеханика*, 2014. № 4. С. 20–37.
5. *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.
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. Iss. 48. P. 2369–2380.
7. *Наумов А. В., Мхитарян Г. А.* О задаче вероятностной оптимизации для ограниченного по времени тестирования // *Автоматика и телемеханика*, 2016. № 9. С. 124–135.
8. *Наумов А. В., Мхитарян Г. А., Черыгова Е. Е.* Стохастическая постановка задачи формирования теста заданного уровня сложности с минимизацией количества времени выполнения // *Вестник компьютерных и информационных технологий*, 2019. № 2. С. 37–46.
9. *Каллан Р.* Основные концепции нейронных сетей / Пер. с англ. — М.: Вильямс, 2001. 287 с. (*Callan R.* The essence of neural networks. — Prentice Hall Europe, 1999. 232 p.)
10. *Воробьев Е. В., Пучков Е. В.* Классификация текстов с помощью сверточных нейронных сетей // *Молодой исследователь Дона*, 2017. № 6(9). С. 2–7. [http://mid-journal.ru/upload/iblock/8ed/1.-vorobev\\_-puchkov.pdf](http://mid-journal.ru/upload/iblock/8ed/1.-vorobev_-puchkov.pdf).
11. *Микрюков А. А., Бабаиш А. В., Сизов В. А.* Классификация событий в системах обеспечения информационной безопасности на основе нейросетевых технологий // *Открытое образование*, 2019. Т. 23. № 1. С. 57–63.
12. *Дьяконов А. Г.* Методы решения задач классификации с категориальными признаками // *Прикладная математика и информатика. Труды факультета Вычислительной математики и кибернетики МГУ имени М. В. Ломоносова*, 2014. № 46. С. 103–127.
13. *Воронцов К. В.* Лекции по алгоритмическим композициям, 2015. 45 с. <http://www.ccas.ru/voron/download/Composition.pdf>.
14. *Наумов А. В., Иноземцев А. О.* Алгоритм формирования индивидуальных заданий в системах дистанционного обучения // *Вестник компьютерных и информационных технологий*, 2013. № 6. С. 46–51.
15. *Мартюшова Я. Г., Лыкова Н. М.* Организация рефлексивно-оценочной деятельности студентов университетов средствами электронного учебника // *Психолого-педагогические исследования*, 2018. Т. 10. № 2. С. 125–134. doi: 10.17759/psyedu.2018100211.
16. *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.

Поступила в редакцию 01.06.20

## BAYESIAN APPROACH TO THE CONSTRUCTION OF AN INDIVIDUAL USER TRAJECTORY IN THE SYSTEM OF DISTANCE LEARNING

A. V. Bosov<sup>1,2</sup>, Ya. G. Martyushova<sup>2</sup>, A. V. Naumov<sup>2</sup>, and A. P. Sapunova<sup>2</sup>

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

<sup>2</sup>Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation

**Abstract:** The paper considers the task of forming an individual user path for a distance learning system (LMS) with a mixed form of conducting educational activities with organization of independent work of students using LMS. At the end of each section of the training course, the LMS users are divided into categories determined by

the solution of the Bayesian classification problem. For each category, an individual task of a different level of complexity is proposed for the next section of the course, thus forming the individual trajectory of the student. The Bayesian classifier is set up based on statistics of work of the users of LMS. The experimental results of solving the problem at one of the stages of training are presented.

**Keywords:** distance learning system; Bayesian classifier; adaptive systems; individual learning path

**DOI:** 10.14357/19922264200313

## Acknowledgments

The reported study was partially funded by RFBR, project No. 18-07-00617.

## References

1. Rasch, G. 1980. *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press. 224 p.
2. Van der Linden, W. J., D. J. Scrams, and D. L. Schnipke. 1999. Using response-time constraints to control for differential speededness in computerized adaptive testing. *Appl. Psych. Meas.* 23(3):195–210.
3. Kuravsky, L. S., P. A. Marmalyuk, V. I. Alkhimov, and G. A. Yuryev. 2013. Novyy podkhod k postroeniuyu intellektual'nykh i kompetentnostnykh testov [A new approach to constructing intellectual and competence-based tests]. *Modelling Data Analysis* 1:4–28.
4. Kibzun, A. I., and A. O. Inozemtsev. 2014. Using the maximum likelihood method to estimate test complexity levels. *Automat. Rem. Contr.* 75(4):607–621.
5. Kuravsky, L. S., P. A. Marmalyuk, G. A. Yuryev, P. N. Dumin, and A. S. Panfilova. 2015. Veroyatnostnoe modelirovanie protsessa vypolneniya testovykh zadaniy na osnove modifitsirovannoy funktsii Rasha [Probabilistic modeling of the test tasks based on the modified Rasch function]. *Voprosy psikhologii* [Psychology Issues] 4:109–118.
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. 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.
8. Naumov, A. V., G. A. Mkhitarian, and E. E. Cherygova. 2019. Stokhasticheskaya postanovka zadachi formirovaniya testa zadannogo urovnya slozhnosti s minimizatsiyey kvantili vremeni vypolneniya [Stochastic statement of the problem of generating tests with defined complexity with the minimization of quantile of test passing time]. *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Herald of Computer and Information Technologies] 2:37–46.
9. Callan, R. 1999. *The essence of neural networks*. Prentice Hall Europe. 232 p.
10. Vorob'ev, E. V., and E. V. Puchkov. 2017. Klassifikatsiya tekstov s pomoshch'yu svertochnykh neyronnykh setey [Classification of texts using convolutional neural networks]. *Molodoy issledovatel' Dona* [Young Researcher of the Don] 6(9):2–7. Available at: [http://mid-journal.ru/upload/iblock/8ed/1.-vorobev\\_-puchkov.pdf](http://mid-journal.ru/upload/iblock/8ed/1.-vorobev_-puchkov.pdf) (accessed July 23, 2020).
11. Mikryukov, A. A., A. V. Babash, and V. A. Sizov. 2019. Klassifikatsiya sobytii v sistemakh obespecheniya informatsionnoy bezopasnosti na osnove neyrosetevykh tekhnologiy [Classification of events in information security systems based on neural networks]. *Otkrytoe obrazovanie* [Open Education] 23(1):57–63.
12. D'yakonov, A. G. 2015. Solution methods for classification problems with categorical attributes. *Computational Mathematics Modeling* 26(3):408–428.
13. Vorontsov, K. V. 2015. Lektsii po algoritmicheskim kompozitsiyam [Lectures on algorithmic compositions]. Available at: <http://www.ccas.ru/voron/download/Composition.pdf> (accessed July 23, 2020).
14. Naumov, A. V., and A. O. Inozemtsev. 2013. Algoritm formirovaniya individual'nykh zadaniy v sistemakh distantsionnogo obucheniya [The algorithm for generating the individual tasks in distance learning]. *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Herald of Computer and Information Technologies] 6:46–51.
15. Martyushova, Ya. G., and N. M. Lykova. 2018. Organizatsiya refleksivno-otsenочноy deyatel'nosti studentov universitetov sredstvami elektronnoy uchebnika [Organization of reflexive-evaluative activity of university students by using the learning management system]. *Psikhologo-pedagogicheskie issledovaniya* [Psychological-Educational Studies] 10(2):125–134. doi: 10.17759/psyedu.2018100211.
16. 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.

Received June 1, 2020

## Contributors

**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; Moscow State Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; AVBosov@ipiran.ru

**Martyushova Yanina G.** (b. 1970) — Candidate of Science in pedagogy, senior lecturer, Department of Probability Theory and Computer Simulations, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; ma1554@mail.ru

**Naumov Andrey V.** (b. 1966) — Doctor of Science in physics and mathematics, professor, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; naumovav@mail.ru

**Sapunova Anastasia P.** (b. 1998) — Master student, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125933, Russian Federation; sap2603@mail.ru

# АНАЛИЗ МЕХАНИЗМОВ НАРЕЗКИ СЕТИ С УЧЕТОМ ГАРАНТИЙ ДЛЯ РАЗЛИЧНЫХ ТИПОВ ТРАФИКА\*

К. А. Агеев<sup>1</sup>, Э. С. Сопин<sup>2</sup>, Н. В. Яркина<sup>3</sup>, К. Е. Самуйлов<sup>4</sup>, С. Я. Шоргин<sup>5</sup>

**Аннотация:** Нарезка радиоресурсов сети (network slicing) — это одна из ключевых возможностей современных сетей, позволяющая нескольким виртуальным мобильным операторам использовать ресурсы одной базовой станции. Это дает возможность операторам, владельцам ресурсов, предоставлять в аренду и управлять несколькими выделенными логическими сетями с определенной функциональностью, реализуемой поверх общей инфраструктуры. Каждая из этих логических сетей называется слайсом сети и может быть адаптирована для обеспечения определенного поведения системы, чтобы наилучшим образом поддерживать определенные показатели качества услуг. В работе построена модель механизма нарезки радиоресурсов, распределяющего ресурс по слайсам, и проведен анализ этой модели методом имитационного моделирования.

**Ключевые слова:** имитационное моделирование; система массового обслуживания; ограниченные ресурсы; нарезка сети

**DOI:** 10.14357/19922264200314

## 1 Введение

Нарезка радиоресурсов сети (англ. *network slicing*) дает возможность оператору мобильной связи предоставлять выделенные логические сети в аренду виртуальным сетевым операторам в виде сетевых слайсов с функциями, специфичными для клиента. Слайс сети, который охватывает все сегменты сетевой инфраструктуры, может быть выделен для конкретных видов услуг нескольким виртуальным операторам, предоставляющим схожие услуги, либо отдельно для каждого виртуального оператора [1, 2].

Для каждого слайса сети выделяются ресурсы (например, виртуализированные сетевые функции, пропускная способность сети и др.), и ошибки или неисправность, возникающие в одном слайсе, не влияют на обеспечение показателей качества обслуживания QoS (Quality of Service) в других слайсах; иными словами, гарантируется изоляция слайса для обеспечения гарантированного качества обслуживания. При этом алгоритм нарезки радиоресурсов должен обеспечивать эффективное использование ресурсов соты с учетом гарантированного объема ресурсов, выделенного для каждого слайса [3].

Данная тематика в последнее время привлекает повышенное внимание исследователей. В [4] представлена гибкая модель нарезки сети радиодоступа (Radio Access Network, RAN). Основные цели заключаются в определении уровня изоляции производительности между операторами виртуальных сетей (Virtual Network Operator, VNO), которые выступают в качестве арендаторов сети, с тем чтобы гарантировать, что их соглашения об уровне обслуживания (Service Level Agreement, SLA) не будут затронуты изменением различных параметров сети, и в то же время оптимизировать использование инфраструктуры RAN путем динамического распределения радиоресурсов между различными сегментами справедливым образом.

В [5] также рассматривается система управления виртуальными радиоресурсами (Virtual radio resource management, VRRM), которая обеспечивает оптимальное использование виртуализированных ресурсов поставщика инфраструктуры между несколькими операторами виртуальной сети. В статье представлена архитектура инструмента моделирования VRRM в терминах систем массового обслуживания (СМО). С помощью разработанного ин-

\* Исследование выполнено при поддержке Программы РУДН «5-100» и при частичной финансовой поддержке РФФИ в рамках научных проектов № 19-07-00933 и № 19-37-90147.

<sup>1</sup> Российский университет дружбы народов, ageev-ka@rudn.ru

<sup>2</sup> Российский университет дружбы народов; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sopin-es@rudn.ru

<sup>3</sup> Российский университет дружбы народов, yarkina-nv@rudn.ru

<sup>4</sup> Российский университет дружбы народов; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, samouylov-ke@rudn.university

<sup>5</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru

струмента проводится анализ практического сценария с тремя поставщиками и различными типами SLA и исследуются показатели производительности при изменении нагрузки на трафик и SLA.

В работах [6, 7] рассматривается теоретическая основа для многооператорного планирования (Multi-Operator Scheduling, MOS). Благодаря динамической адаптации к каналу и нагрузке централизованный подход максимизирует спектральную эффективность для нескольких операторов с полным контролем над гарантиями совместного использования.

В данной работе рассматривается сценарий функционирования одной соты беспроводной сети связи, в которой активировано  $S$  слайсов, модуль нарезки делит между ними  $C$  единиц радиоресурсов. Каждый слайс предоставляет пользователям услугу связи, предполагающую непрерывную передачу данных с определенным выделенным ресурсом, скоростью передачи, не менее  $a_s \geq 0$  и не более  $b_s \geq a_s, s \in S$ . При этом скорость передачи является переменной: в каждый момент времени она пересчитывается и зависит от числа активных сессий в каждом слайсе. Предполагается, что ресурс в рамках одного слайса распределяется поровну между пользователями. В работе описана модель в виде СМО, предложен алгоритм разделения радиоресурсов, описана работа средства имитационного моделирования, проведен численный эксперимент и анализ полученных результатов.

## 2 Математическая модель

Пусть в многолинейную СМО поступает  $S$  потоков заявок, соответствующих запросам на передачу данных от пользователей  $S$  различных слайсов. Потоки являются пуассоновскими с интенсивностями  $\lambda_s, s \in S$ . Объемы заявок — независимые случайные величины, распределенные по экспоненциальному закону с параметрами  $1/\mu_s, s \in S$ , а скорость обслуживания заявки определяется объемом выделенного ей ресурса.

Пусть общий объем ресурсов СМО для обслуживания заявок равен  $C$ . Количество ресурсов, выделяемых заявке, принятой на обслуживание, зависит от состояния системы и может варьироваться в диапазоне  $[a_s, b_s], s \in S$ . При этом после каждого поступления либо ухода заявки происходит перераспределение ресурсов между слайсами.

Определим случайный процесс  $X(t) = \{m_1(t), m_2(t), \dots, m_S(t)\}$ , где  $m_s, s \in S$ , — число заявок в слайсе в момент времени  $t$ , причем

$$m_s \in \left\{ 0, 1, \dots, \left\lfloor \frac{C}{a_s} \right\rfloor \right\}, s \in S.$$

Тогда пространство возможных состояний процесса имеет вид:

$$X = \left\{ (m_1, m_2, \dots, m_S), \sum_{s=1}^S m_s a_s \leq C \right\}.$$

Обозначим через  $r_s, s \in S$ , количество выделенного ресурса одной заявке в слайсе  $s, s \in S$ . Тогда интенсивности обслуживания заявок соответствующих слайсов определяются как  $r_s \mu_s, s \in S$ .

Для обеспечения изоляции слайсов обозначим через  $\bar{R}_s$  объем ресурсов, который гарантированно выделен слайсу  $s, s \in S$ , а через  $\bar{M}_s = \bar{R}_s / a_s$  — число заявок, которое гарантированно может быть принято в слайсе  $s, s \in S$ . Слайсы, число заявок в которых превышает гарантированное значение, будем называть нарушителями. В случае нехватки ресурсов и наличия слайсов-нарушителей поступившая заявка другого слайса может вытеснить одну или несколько заявок слайсов-нарушителей. В случае нехватки ресурсов и отсутствия нарушителей, а также в случае когда все слайсы нарушают, поступающая заявка будет сброшена.

Рассмотрим подробнее возможные события при поступлении сессий в состоянии  $(m_1, m_2, \dots, m_S) \in X$ . Пусть в систему поступает заявка слайса  $s, s \in S$ . Тогда возможны следующие случаи:

- $(m_1, m_2, \dots, m_s + 1, \dots, m_S) \in X$ , т.е. предоставление минимального количества ресурса для поступающей заявки возможно; в этом случае заявка встает на обслуживание, а случайный процесс переходит в состояние  $(m_1, m_2, \dots, m_s + 1, \dots, m_S)$ ;
- $(m_1, m_2, \dots, m_s + 1, \dots, m_S) \notin X$ , т.е. не гарантируется предоставление минимально требуемого количества ресурса, тогда:
  - если  $m_s < \bar{M}_s$  и  $m_k > \bar{M}_k, k \in S, k \neq s$ , то выполняется освобождение ресурсов слайса-нарушителя. Алгоритм сброса заявок приведен в разд. 3;
  - в остальных случаях поступающая заявка будет сброшена.

Таким образом, множество состояний сброса заявок при поступлении:

$$D_s = \left\{ (m_1, m_2, \dots, m_S) : \right. \\ \left. : ((m_1, m_2, \dots, m_s + 1, \dots, m_S) \notin X) \cap \right. \\ \left. \cap ((m_s \geq \bar{M}_s) \cup (m_k \leq \bar{M}_k, k \in S, k \neq s)) \right\}.$$

Множество состояний прерывания обслуживания:

$$B_k = \{ (m_1, m_2, \dots, m_s) : \\ : ((m_1, m_2, \dots, m_s + 1, \dots, m_s) \neq X) \cap \\ \cap ((m_s < \overline{M}_s) \cup (m_k > \overline{M}_k, k \in S, k \neq s)) \} .$$

### 3 Алгоритм выбора сбрасываемой заявки

Для выбора сбрасываемой заявки при наличии нескольких слайсов-нарушителей введем понятие веса слайса, показывающего, насколько сильно слайс нарушает границы других слайсов. Ниже приведены три возможные формулы для вычисления весов:

$$w_s^{(1)} = \begin{cases} 1, & m_s \leq \overline{M}_s; \\ \frac{1}{m_s - \overline{M}_s + 1}, & m_s > \overline{M}_s; \end{cases}$$

$$w_s^{(2)} = \begin{cases} 1, & m_s \leq \overline{M}_s; \\ \frac{\overline{M}_s}{m_s}, & m_s > \overline{M}_s; \end{cases}$$

$$w_s^{(3)} = \begin{cases} 1, & m_s \leq \overline{M}_s; \\ w = const < 1, & m_s > \overline{M}_s. \end{cases}$$

Освобождение ресурсов требуется в случае, если заявка поступает в слайс, который не является нарушителем, при этом она не может быть принята в систему из-за нехватки свободного ресурса:

$$\{ ((m_1, m_2, \dots, m_s + 1, \dots, m_s) \notin X) : \\ : (m_s + 1) a_s \leq R_s \} .$$

**Шаг 1:** ищем слайс с наименьшим весом, т. е.  $s^* \in S: w_{s^*} = \min\{w_r, r \in S\}$ .

**Шаг 2:** сбрасываем заявку найденного слайса  $(m_1, m_2, \dots, m_{s^*} - 1, \dots, m_S)$ .

Шаги 1 и 2 выполняются, пока не будет удовлетворено условие  $C - \sum_{i=1}^S m_i a_i \geq a_s$ . Отметим, что в случае, когда у нескольких слайсов вес минимален, выбор слайса для сброса заявки происходит случайным образом.

### 4 Распределение ресурсов

В состоянии избытка ресурсов

$$X_0 = \left( (m_1, m_2, \dots, m_s + 1, \dots, m_S) : \sum_{s=1}^S m_s b_s \leq C \right)$$

всем заявкам во всех слайсах будет выделен максимальный объем ресурса  $b_s$ .

В состоянии ограниченных ресурсов  $X_1 = X \setminus X_0$  для справедливого и эффективного распределения ресурсов решается задача оптимизации. Пусть скорости передачи данных в слайсах имеют функцию полезности  $U_s(r_s) = \ln(r_s)$ . Тогда задача выглядит следующим образом:

$$\sum_{s \in S} w_s(m_s) m_s U_s(r_s) \rightarrow \max, \quad s \in S;$$

$$\sum_{s \in S} m_s r_s = C; \quad (1)$$

$$P = \{ \mathbf{r} \in \mathbf{R}^S : a_s \leq r_s \leq b_s, s \in S \}, \quad (2)$$

где вектор  $\mathbf{r}$  имеет вид  $\mathbf{r} = (r_1, r_2, \dots, r_S)$ . Решение задачи выполняется с помощью приближенного метода проецирования градиента, в котором итеративная процедура поиска максимума описывается соотношениями:

$$\mathbf{d}_k = \mathbf{P} \nabla f(\mathbf{x}_k);$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \tau_k \mathbf{d}_k;$$

$$\tau_k > 0 : \mathbf{x}_{k+1} \in P,$$

где вектор  $\mathbf{x}$  является решением, а  $\mathbf{P}$  — матрица проецирования на гиперплоскость (1):

$$\mathbf{P} = \mathbf{I} - \mathbf{m} (\mathbf{m} \mathbf{m}^T)^{-1} \mathbf{m} = \mathbf{I} - \frac{1}{\sum_{s \in S} m_s^2} \mathbf{m} \mathbf{m}^T .$$

Здесь вектор  $\mathbf{m} = (m_1, m_2, \dots, m_S)$  — текущее состояние системы. Градиент функции полезности представляет собой вектор-столбец:

$$\nabla f(\mathbf{x}_k) = \left( \frac{w_s m_s}{r_s} \right)_{s \in S} .$$

Длину шага  $\tau_k$  выбираем таким образом, чтобы не выйти за пределы области  $P$ , задаваемой прямыми ограничениями задачи (2). В качестве начального приближения  $\mathbf{x}_0$  удобно взять точку пересечения диагонали координатного параллелограмма  $P$ , соединяющей точки  $(a_1, a_2, \dots, a_S)$  и  $(b_1, b_2, \dots, b_S)$  с гиперплоскостью (1). Данная точка находится при решении системы линейных уравнений:

$$(b_s - a_s)x_1 - (b_1 - a_1)x_s = a_1 b_s - a_s b_1;$$

$$(b_s - a_s)x_2 - (b_2 - a_2)x_s = a_2 b_s - a_s b_2;$$

$$\dots$$

$$(b_s - a_s)x_{s-1} - (b_{s-1} - a_{s-1})x_s =$$

$$= a_{s-1} b_s - a_s b_{s-1};$$

$$m_1 x_1 + \dots + m_s x_s = C .$$

Итеративная процедура обеспечивает движение по гиперплоскости (1) в направлении возрастания функции полезности до границы области  $P$ , где и будет найдено решение.

## 5 Описание работы средства имитационного моделирования

Общая схема работы имитатора изображена на рис. 1. На шаге инициализации входных параметров задаются начальные параметры:  $\lambda_s, \mu_s, \bar{M}_s, a_s, b_s, s \in S, C, \max$  — максимальное число принятых заявок в системе. Затем выполняется запуск имитационного моделирования. Далее определяется ближайшее событие:

- (1) если это поступление, то выполняется проверка на достаточность ресурсов:

- (1.1) если ресурсов хватает, то добавляется заявка, пересчитываются ресурсы, и определяется продолжение моделирования;

- (1.2) если ресурсов не хватает, то проверяется, является ли слайс нарушителем:

- (1.2.1) если является, то входящая заявка считается заблокированной;

- (1.2.2) если не является, то запускается процесс поиска нарушителя и освобождение ресурсов, после чего происходит добавление заявки, пересчитываются ресурсы и определяется продолжение моделирования;

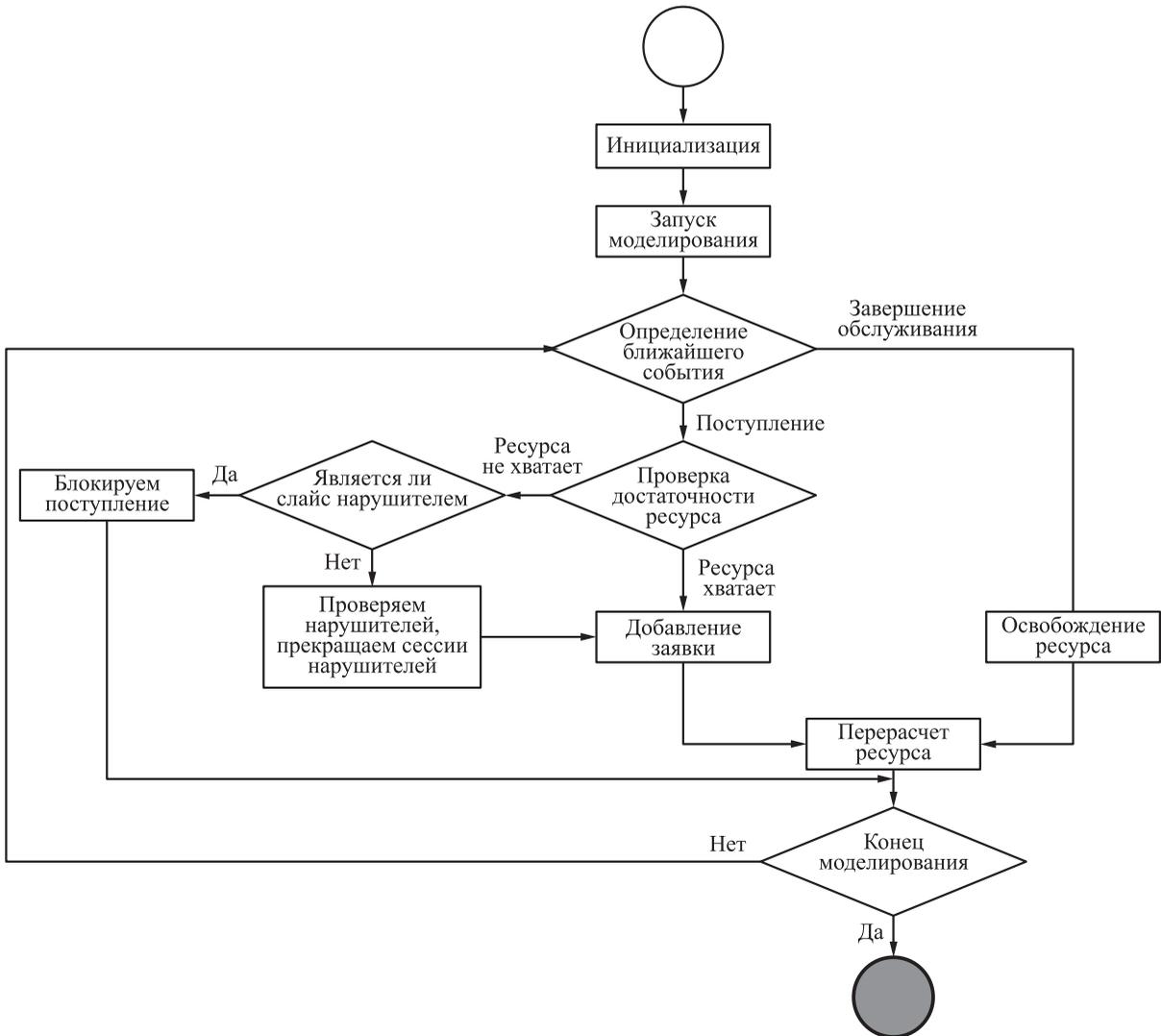


Рис. 1 Схема работы инструмента имитационного моделирования

- (2) если ближайшее событие — это обслуживание, то выполняется освобождение и перерасчет ресурсов.

## 6 Численный эксперимент

Для численного эксперимента рассматривается диапазон 10 МГц в сети LTE (long-term evolution). Минимально допустимая скорость в таком диапазоне составляет 0,75 Мбит/с ( $a_s = 0,75, s \in S$ ), максимальная — 79,9 Мбит/с ( $b_s = 79,9, s \in S$ ). Максимальный объем ресурса, который может одновременно предоставлять базовая станция в диапазоне 10 МГц, — 450 Мбит/с, для расчетов используются несколько значений объема ресурсов:

$$C = [325, 350, 375, 400, 425, 450].$$

Данный ресурс будет разделен на трех ( $S = 3$ ) виртуальных операторов с гарантированным выделенным ресурсом  $\bar{R}_1 = 225, \bar{R}_2 = 150$  и  $\bar{R}_3 = 75$ . Интенсивности поступлений  $\lambda_1 = \lambda_2 = \lambda_3 = 30$ , интенсивности обслуживания  $\mu_1 = \mu_2 = \mu_3 = 30$  оставим равными для всех операторов.

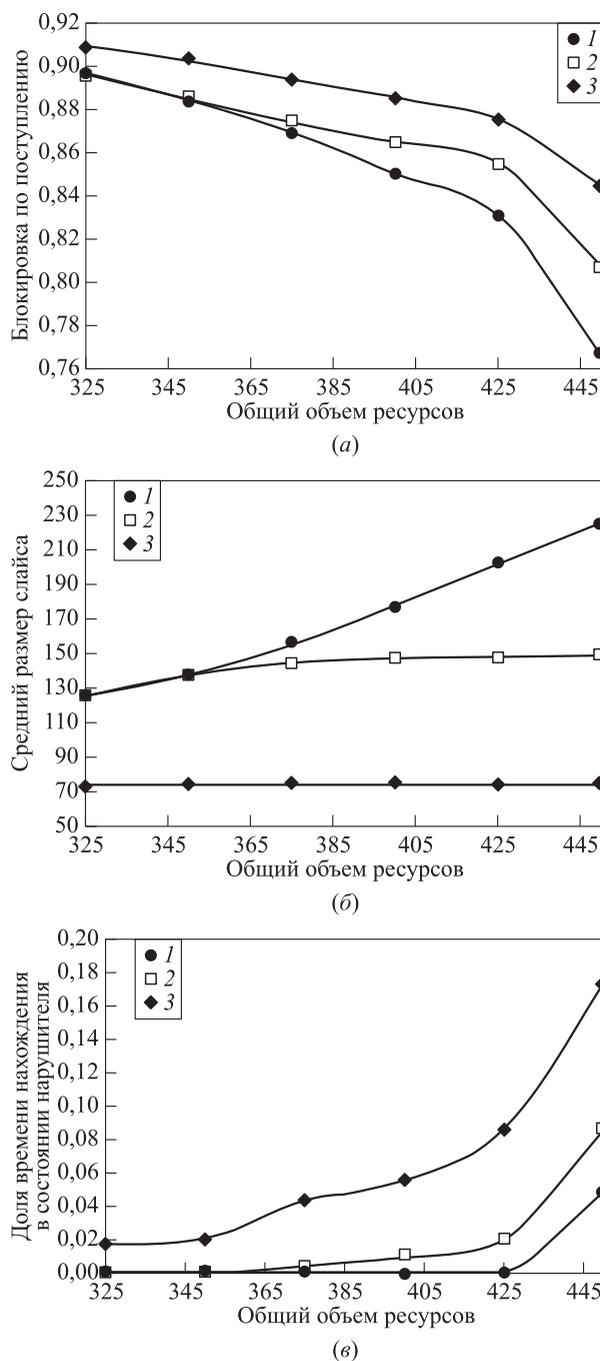
На рис. 2, а видно, что когда ресурсов не хватает для обеспечения гарантированной скорости для всех операторов, то и вероятность блокировки выше. Далее при увеличении ресурса видно, что для оператора, которому предоставлено больше гарантированных ресурсов, уменьшение вероятности блокировки происходит быстрее.

На рис. 2, б видно, что для оператора, которому предоставляется больший гарантированный ресурс, увеличивается средний размер слайса, в то время как для оператора, которому гарантируется меньший ресурс, этот параметр меняется незначительно. Это связано со способом решения оптимального распределения ресурса, описанного в данной работе.

На рис. 2, в отображен график, показывающий изменение доли времени, когда оператор является нарушителем. В связи с тем что общий размер ресурса растет, гарантированный объем остается неизменным, а интенсивности поступления — фиксированные для каждого оператора, можно наблюдать резкое увеличение значений этого параметра для оператора 3.

## 7 Заключение

В работе описана СМО с ограниченными ресурсами с распределением ресурсов в зависимости от веса слайса сети. Предложен и реализован алгоритм освобождения ресурсов слайсами-нарушителями. Предложен и реализован алгоритм распределения



**Рис. 2** Вероятность блокировки (а), средний размер слайса (б) и доля времени, когда слайс находится в состоянии нарушителя (в) в зависимости от объема ресурсов: 1 — VNO 1; 2 — VNO 2; 3 — VNO 3

ресурсов на основе весов слайсов. Разработано средство имитационного моделирования механизма распределения ресурсов. Проведен численный эксперимент для трех слайсов. В дальнейшем планируется разработать модель, в которой перераспределение слайсов происходит не при каждом из

менении состояния процесса, а при выполнении некоторого критерия.

## Литература

1. 3GPP TS 23.501 V15.4.0. System architecture for the 5G System, 2018. [http://www.3gpp.org/ftp//Specs/archive/23\\_series/23.501/23501-f40.zip](http://www.3gpp.org/ftp//Specs/archive/23_series/23.501/23501-f40.zip).
2. ITU-T Rec. Y.3101. Requirements of the IMT-2020 network, 2018. [https://www.itu.int/rec/dologin\\_pub.asp?lang=e&id=T-REC-Y.3101-201801-1!!!PDF-E&type=items](https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.3101-201801-1!!!PDF-E&type=items).
3. Pérez-Romero J., Sallent O., Ferrús R., Agustí R. On the configuration of radio resource management in a sliced RAN // IEEE/IFIP Network Operations and Management Symposium. — IEEE, 2018. P. 1–6. doi: 10.1109/NOMS.2018.8406280.
4. Rouzbehani B., Correia L. M., Caeiro L. An SLA-based method for radio resource slicing and allocation in virtual RANs // IRACON 7th MC and Technical Meeting. — Cartagena, Spain, 2018. Cost Action 15104 TD(18)07034. P. 1–7.
5. Ageev K., Garibyan A., Golskaya A., Gaidamaka Yu., Sopin E., Samouylov K., Correia L. Modelling of virtual radio resources slicing in 5G networks // Information technologies and mathematical modelling. Queueing theory and applications / Eds. A. Dudin, A. Nazarov, A. Moiseev. — Communications in computer and information science ser. — Cham: Springer, 2019. Vol. 1109. P. 150–161. doi: 10.1007/978-3-030-33388-1\_13.
6. Malanchini I., Valentin S., Aydin O. An analysis of generalized resource sharing for multiple operators in cellular networks // 25th Annual Symposium (International) on Personal, Indoor, and Mobile Radio Communication. — IEEE, 2014. P. 1157–1162. doi: 10.1109/PIMRC.2014.7136342.
7. Malanchini I., Valentin S., Aydin O. Wireless resource sharing for multiple operators: Generalization, fairness, and the value of prediction // Comput. Netw., 2016. Vol. 100. P. 110–123. doi: 10.1016/j.comnet.2016.02.014.

Поступила в редакцию 15.07.20

## ANALYSIS OF THE NETWORK SLICING MECHANISMS WITH GUARANTEED ALLOCATED RESOURCES FOR VARIOUS TRAFFIC TYPES

K. A. Ageev<sup>1</sup>, E. S. Sopin<sup>1,2</sup>, N. V. Yarkina<sup>1</sup>, K. E. Samouylov<sup>1,2</sup>, and S. Ya. Shorgin<sup>2</sup>

<sup>1</sup>Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

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

**Abstract:** Network slicing is one of the key capabilities of modern networks, allowing several virtual mobile operators to use the physical resources of one base station. This allows operators and resource owners (tenants) to lease and manage several dedicated logical networks with specific functionality working on top of a common infrastructure. Each of these logical networks is called a network slice and can be adapted to provide certain system behavior to maintain a specified level of quality of service indicators. The paper describes the developed mathematical framework of the network slicing mechanisms and analyzes it by means of extensive simulations.

**Keywords:** simulation modeling; queuing system; limited resources; network slicing

**DOI:** 10.14357/19922264200314

## Acknowledgments

The reported study was funded by the “RUDN University Program 5-100” and in part by RFBR, projects Nos. 19-07-00933 and 19-37-90147.

## References

1. 3GPP TS 23.501 V15.4.0. 2018. System architecture for the 5G System. Available at: [http://www.3gpp.org/ftp//Specs/archive/23\\_series/23.501/23501-f40.zip](http://www.3gpp.org/ftp//Specs/archive/23_series/23.501/23501-f40.zip) (accessed June 15, 2020).
2. ITU-T Rec. Y.3101. 2018. Requirements of the IMT-2020 network. Available at: [https://www.itu.int/rec/dologin\\_pub.asp?lang=e&id=T-REC-Y.3101-201801-1!!!PDF-E&type=items](https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.3101-201801-1!!!PDF-E&type=items) (accessed June 15, 2020).
3. Pérez-Romero, J., O. Sallent, R. Ferrús, and R. Agustí. 2018. On the configuration of radio resource manage-

- ment in a sliced RAN. *IEEE/IFIP Network Operations and Management Symposium*. IEEE. 1–6. doi: 10.1109/NOMS.2018.8406280.
4. Rouzbehani, B., L. M. Correia, and L. Caeiro. 2018. An SLA-based method for radio resource slicing and allocation in virtual RANs. *IRACON 7th MC and Technical Meeting*. Cartagena. TD(18)07034. 1–7.
  5. Ageev, K., A. Garibyan, A. Golskaya, Yu. Gaidamaka, E. Sopin, K. Samouylov, and L. Correia. 2019. Modelling of virtual radio resources slicing in 5G networks. *Information technologies and mathematical modelling. Queueing theory and applications*. Eds. A. Dudin, A. Nazarov, and A. Moiseev. Communications in computer and information science ser. Cham: Springer. 1109:150–161.
  6. Malanchini, I., S. Valentin, and O. Aydin. 2014. An analysis of generalized resource sharing for multiple operators in cellular networks. *25th Annual Symposium (International) on Personal, Indoor, and Mobile Radio Communication*. IEEE. 1157–1162. doi: 10.1109/PIMRC.2014.7136342.
  7. Malanchini, I., S. Valentin, and O. Aydin. 2016. Wireless resource sharing for multiple operators: Generalization, fairness, and the value of prediction. *Comput. Netw.* 100:110–123. doi: 10.1016/j.comnet.2016.02.014.

Received July 15, 2020

## Contributors

**Ageev Kirill A.** (b. 1993) — PhD student, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; ageev-ka@rudn.ru

**Sopin Eduard S.** (b. 1987) — Candidate of Science 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 Sciences and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation; sopin-es@rudn.ru

**Yarkina Natalia V.** (b. 1979) — Candidate of Science in physics and mathematics, associate professor, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; yarkina-nv@rudn.ru

**Samouylov Konstantin E.** (b. 1955) — Doctor of Science in technology, professor, Head of Department, 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 Sciences and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation; samuylov\_ke@rudn.university

**Shorgin Sergey Ya.** (b. 1952) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sshorgin@ipiran.ru

# О КОНЦЕПЦИИ СТОХАСТИЧЕСКОЙ МОДЕЛИ С УПРАВЛЕНИЕМ В МОМЕНТЫ ВЫХОДА ПРОЦЕССА НА ГРАНИЦУ ЗАДАННОГО ПОДМНОЖЕСТВА МНОЖЕСТВА СОСТОЯНИЙ

П. В. Шнурков<sup>1</sup>, Д. А. Новиков<sup>2</sup>

**Аннотация:** Работа посвящена созданию и анализу общей концепции специальной стохастической модели с управлениями. Основная особенность модели заключается в том, что управляющие воздействия осуществляются в моменты времени, когда случайный процесс, описывающий исследуемую систему, достигает границы некоторого заданного подмножества множества состояний. Само управляющее воздействие заключается в переводе процесса из граничного в одно из внутренних состояний заданного подмножества. При этом внутренние состояния интерпретируются как допустимые, а граничные — как недопустимые. Управляющие воздействия описываются набором дискретных вероятностных распределений, зависящих от номера граничного состояния. Такой набор определяет стратегию управления. Проблема оптимального управления формализуется как задача нахождения стратегии управления, доставляющей глобальный экстремум некоторому стационарному стоимостному показателю эффективности, который по своему экономическому содержанию представляет собой среднюю удельную прибыль, возникающую при длительной эволюции системы. Поставленную проблему оптимального управления предлагается называть задачей о настройке. Отмечается, что данная стохастическая модель и соответствующая задача о настройке могут быть использованы для исследования многих реальных явлений, происходящих в экономических и технических системах. В качестве примера такого явления рассматривается проведение интервенций на валютном рынке Российской Федерации.

**Ключевые слова:** управление в стохастических системах; марковские управляемые процессы; полумарковские управляемые процессы; стохастическая задача о настройке

**DOI:** 10.14357/19922264200315

## 1 Введение

При анализе многих процессов стохастического характера, происходящих в технических и экономических системах, наблюдается следующее явление. Основной случайный процесс, описывающий исследуемую систему, в некоторые моменты времени выходит из заданного подмножества состояний, которые считаются допустимыми. Для того чтобы вернуть процесс в подмножество допустимых состояний, осуществляется внешнее воздействие, которое можно считать управлением в рассматриваемой стохастической модели. В результате управляющего воздействия процесс возвращается в одно из состояний допустимого множества. При этом параметром управления, или решением, служит само состояние (или номер этого состояния), в которое будет переведен процесс в результате внешнего воздействия. Предполагается, что состояние, в которое переводится процесс, определяется как результат случайного эксперимента в соответствии с некоторым вероятностным распределением, которое можно назвать управляющим. После перевода основного случайного процесса в одно из

допустимых состояний этот процесс вновь начинает эволюционировать независимо от прошлого; его поведение будет зависеть только от состояния, в которое он был переведен в результате внешнего воздействия (управления). В некоторый момент времени основной процесс вновь покидает множество допустимых состояний, после чего осуществляется внешнее управляющее воздействие, которое организуется по указанным выше правилам. Проблема оптимизации управления заключается в нахождении управляющих вероятностных распределений, доставляющих экстремум некоторому показателю эффективности управления.

Изложенные выше идеи, основанные на особенностях функционирования реальных технических и экономических систем, образуют общую концепцию стохастической модели с управлениями, осуществляемыми в моменты выходов основного процесса на границу заданного подмножества множества состояний. Создание соответствующей математической модели и решение возникающей при этом задачи оптимального управления является актуальной проблемой прикладной математики. В дальнейшем будем называть такую теоретическую

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

<sup>2</sup>Национальный исследовательский университет «Высшая школа экономики», even.he@yandex.ru

проблему задачей о настройке, имея в виду аналогию с настройкой некоторого важного параметра технической системы.

В качестве примера реальной экономической системы, для описания которой можно использовать предлагаемую стохастическую модель с управлениями, происходящими в моменты выходов процесса на границу заданного подмножества множества состояний, рассмотрим валютный рынок Российской Федерации и так называемые интервенции, проводившиеся на этом рынке Центральным банком РФ в течение ряда лет.

## 2 Общее описание явления интервенции в экономике

Для создания модели необходимо точное описание реального явления. В связи с этим начнем исследование с научного определения понятия интервенции и некоторых связанных с ним понятий, принятых в экономической теории.

«Интервенция — экономическое воздействие одного субъекта на дела и действия другого, проводимое посредством проникновения в сферу этих действий, вложения и размещения в ней собственных денежных средств.

Обычно интервенционные операции проводятся центральными банками, казначейством путем массовой продажи или скупки валюты, ценных бумаг, предоставления кредитов в целях нормализации состояния финансовой системы» [1].

Как следует из данного определения, по своему содержанию интервенции подразделяются на те, основное действие которых заключается в продаже какого-либо товара, валюты или ценных бумаг, и те, основным действием которых является закупка таких товаров или ценных бумаг. Эти действия и составляют сущность явления, называемого интервенцией.

Приведем еще одно определение интервенции. При проведении закупочной интервенции осуществляется покупка товара, в случае если его цена опускается ниже установленной (интервенционной), чтобы повысить рыночную цену. Товарная интервенция проводится на внутреннем рынке, если внутренняя цена превышает интервенционную цену [2].

Понятия закупочной и товарной интервенций, проводимых на рынке продукции сельского хозяйства Российской Федерации, определены не только как экономические категории, но также и законодательно. Приведем описание указанных понятий, следуя опубликованному тексту статьи 14 Федерального закона «О развитии сельского хозяйства»:

«1. Государственные закупочные интервенции, товарные интервенции проводятся в целях стабилизации цен на рынке сельскохозяйственной продукции, сырья и продовольствия и поддержания уровня доходов сельскохозяйственных товаропроизводителей.

2. Государственные закупочные интервенции проводятся при снижении цен на реализуемую сельскохозяйственную продукцию ниже минимальных расчетных цен путем закупки, в том числе на биржевых торгах, у сельскохозяйственных товаропроизводителей произведенной ими сельскохозяйственной продукции или путем проведения залоговых операций в отношении данной продукции.

3. Государственные товарные интервенции проводятся при росте цен на реализуемую сельскохозяйственную продукцию свыше максимальных расчетных цен путем продажи закупленной сельскохозяйственной продукции, в том числе на биржевых торгах» [3].

Итак, в данном разделе на основе авторитетных источников определена сущность экономического явления, называемого интервенцией. Эта сущность заключается в целенаправленном воздействии на рынок какого-либо товара, валюты или ценных бумаг в форме их продажи или закупки. Проведение закупочной интервенции связывается с моментом времени, когда рыночная цена на некоторый продукт опускается ниже заданного минимального уровня. Товарные интервенции должны осуществляться в моменты времени, когда рыночная цена продукта становится выше заданного максимального уровня. Отметим, что указанные признаки, т. е. достижение ценой минимального и максимального допустимого уровня, считаются основными причинами проведения интервенций. Как следует из приведенных выше описаний, цель проведения интервенции состоит в достижении такого состояния рассматриваемой экономической системы (товарного или валютного рынка), в котором значение некоторого основного параметра, а именно: цены соответствующего товара или валюты, находилось бы в заданных допустимых пределах. Математическая модель, описывающая интервенции, должна прежде всего отражать указанные основные особенности этого реального явления.

## 3 Примеры интервенций на валютном рынке Российской Федерации

Приведем конкретные примеры закупочных и товарных интервенций, проводившихся Цент-

ральным банком на валютном рынке Российской Федерации в течение 2010–2012 гг. Основной случайный процесс, управление которым осуществлялось при помощи интервенций, представлял собой стоимость так называемой бивалютной корзины, включающей в себя доллар и евро в определенной пропорции.

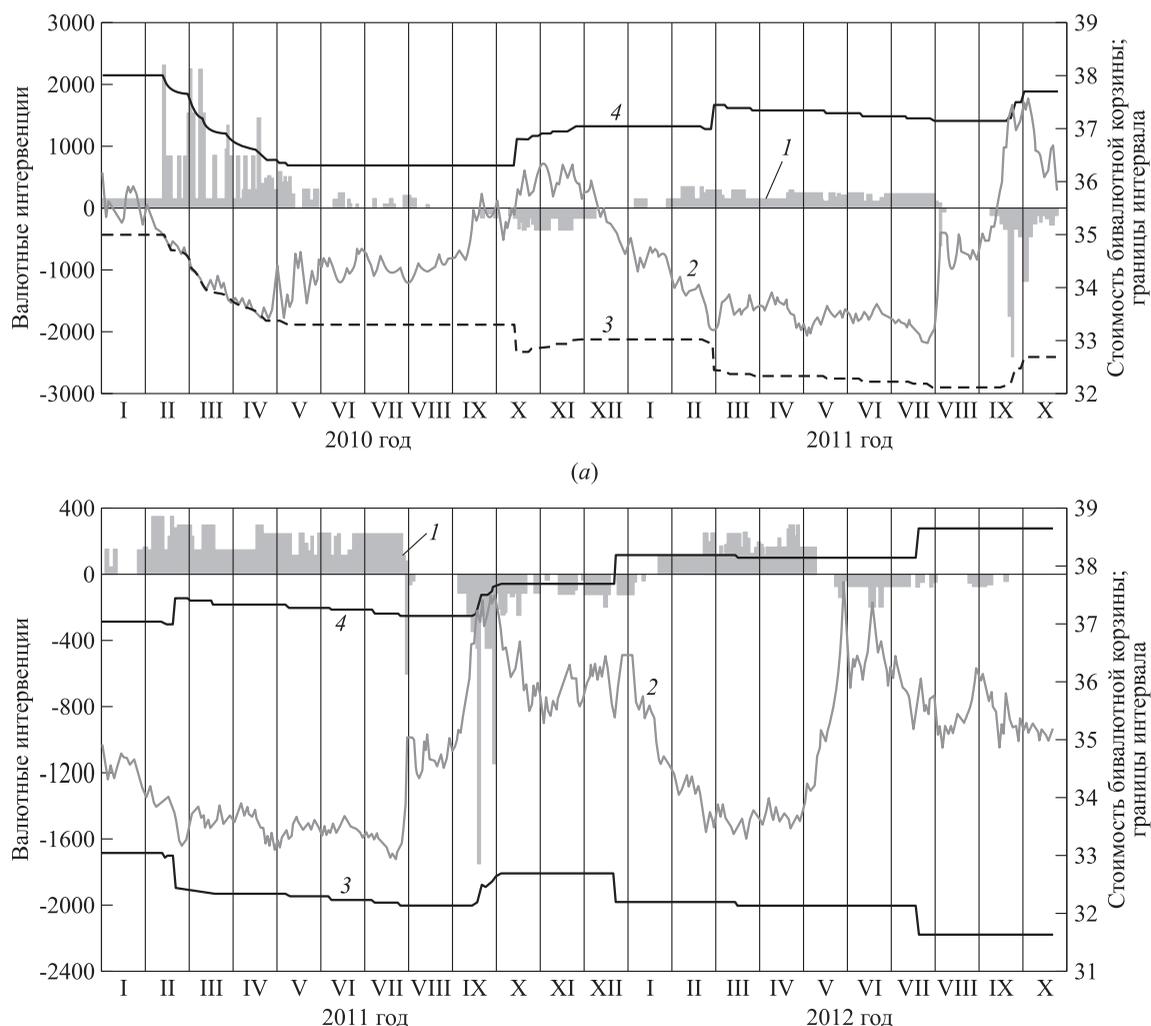
В 2011 г. политика Центрального банка была направлена на сглаживание курса рубля. В начале года проводились закупочные интервенции для замедления снижения стоимости бивалютной корзины.

С августа по сентябрь наблюдался отток частного капитала из России, в результате чего спрос на иностранную валюту на внутреннем валютном рынке значительно возрос и превысил предложение. В ответ на сложившуюся ситуацию Центральный банк принял меры в виде товарных интервенций. Таким образом, путем продажи иностранной валю-

ты Центральный банк обеспечил снижение стоимости бивалютной корзины для стабилизации ситуации на внутреннем валютном рынке.

В январе–апреле 2012 г. на внутреннем валютном рынке наблюдалось устойчивое превышение предложения иностранной валюты над спросом на нее, что создавало предпосылки для укрепления рубля и обусловило проведение Центральным банком операций по покупке иностранной валюты на внутреннем валютном рынке.

В мае–июне 2012 г. на мировых валютных рынках наблюдалось увеличение спроса на доллар США, сопровождавшееся ослаблением российского рубля. В этих условиях в рамках действующего механизма курсовой политики Центральным банком в конце мая осуществлялась продажа иностранной валюты на внутреннем валютном рынке, что позволило сгладить колебания курса рубля.



**Рис. 1** Интервенции Центрального банка на внутреннем валютном рынке (1) и динамика рублевой стоимости бивалютной корзины (2 — стоимость бивалютной корзины, рассчитанная по официальным курсам рублей; 3 и 4 — нижняя и верхняя границы операционного интервала): (а) 2010–2011 гг.; (б) 2011–2012 гг.

Иллюстрации, демонстрирующие эволюцию стоимости бивалютной корзины и объемы интервенций Центрального банка в течение 2010–2012 гг., приведены на рис. 1.

#### 4 Конструктивное описание стохастической управляемой модели

Перейдем к построению стохастической модели с управлениями, происходящими в моменты выходов процесса на границу заданного подмножества множества состояний. В данном разделе будет приведено конструктивное описание эволюции соответствующих моделей с дискретным и непрерывным временем. Для сохранения последовательности изложения будем предполагать, что рассматриваемая система имеет экономический характер, хотя, как уже отмечалось во введении, аналогичные явления имеют место и в технических системах.

В рамках данного исследования под экономической системой будем понимать какой-либо товарный или финансовый рынок, на котором объективно действуют случайные факторы. На этом рынке формируется некоторый основной параметр, изменяющийся во времени. Этот параметр представляет собой случайный процесс, который будет считаться математической моделью функционирования исследуемой системы. Состояние этого процесса в произвольный момент времени будет характеризовать состояние системы. В системе товарного рынка (например, рынка зерновых культур) таким параметром выступает текущая цена единицы соответствующего товара. В системе финансового рынка таким параметром может служить текущая стоимость какой-либо иностранной валюты. На финансовых рынках России используется понятие бивалютной корзины, в которую входят доллар и евро. Для таких рынков параметром состояния можно считать цену этой корзины.

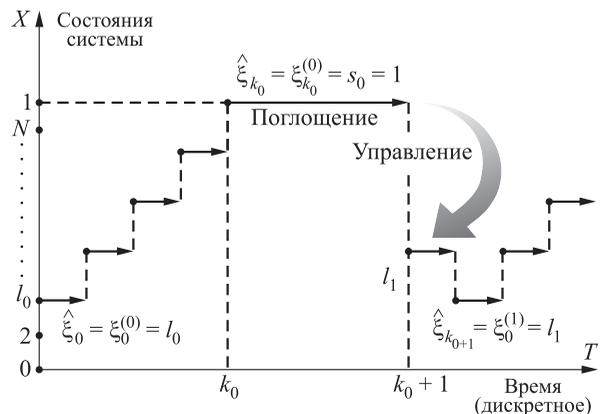
Эволюция этого процесса в целом состоит из двух основных этапов, последовательно сменяющих друг друга и образующих повторяющийся цикл. На первом этапе эволюция проходит без внешнего воздействия, по внутренним закономерностям данной рыночной системы. В множестве возможных значений процесса выделяется заданное подмножество допустимых значений. В период пребывания процесса в этом допустимом подмножестве внешние воздействия не используются. В моменты, когда процесс (основной параметр) выходит на границу или за пределы допустимого подмножества значений, осуществляется внешнее

воздействие, которое и представляет собой интервенцию. Проведение такого воздействия образует второй этап эволюции процесса. Цель этого воздействия — возвращение значения процесса в одно из состояний допустимого подмножества. Такое внешнее воздействие является управлением в данной вероятностной модели. Математическая проблема оптимального управления заключается в нахождении такого вероятностного распределения управляющего воздействия, которое доставляет глобальный экстремум некоторому стоимостному показателю эффективности функционирования данной экономической системы.

Введем важное предположение, связанное с особенностями предлагаемой стохастической модели. Математическое содержание этого предположения заключается в том, что основной случайный процесс, описывающий функционирование исследуемой системы, обладает марковским свойством в моменты достижения граничных состояний, а также в моменты перехода из граничных состояний во внутренние. В соответствии с идеей марковского свойства данное предположение означает, что начиная с каждого такого момента времени процесс продолжает эволюционировать независимо от прошлого, а его вероятностные характеристики зависят только от состояния в указанный момент. Заметим, что такое предположение выполняется для многих технических и экономических систем.

Иллюстрацией приведенного описания эволюции модели служит рис. 2, на котором изображена некоторая траектория основного процесса и возможные управляющие воздействия.

Перейдем к непосредственному описанию стохастических процессов, характеризующих упомя-



**Рис. 2** Возможная траектория случайного процесса  $\{\hat{\xi}_k\}$ , представляющего собой стохастическую модель поведения основного параметра

нутую экономическую систему. Однако предварительно примем определенные соглашения об обозначениях состояний.

Предположим, что множество состояний рассматриваемого основного процесса конечно:  $X = \{0, 1, 2, \dots, N\}$ . Для удобства применения результатов теории поглощающих цепей Маркова [4] проведем нумерацию состояний таким образом, что состояния  $\{0\}$  и  $\{1\}$  будут граничными, недопустимыми и поглощающими, а остальные состояния  $\{2, 3, \dots, N\}$  — внутренними, допустимыми и невозвратными.

Начнем с описания стохастической модели с дискретным временем. Предположим, что задана последовательность независимых поглощающих марковских цепей

$$\xi^{(n)} = \left\{ \xi_k^{(n)} \right\}_{k=0}^{\infty}, \quad n = 0, 1, 2, \dots,$$

с множеством состояний  $X$  и одинаковыми матрицами переходных вероятностей. В начальный момент времени  $t_0 = 0$  процесс стартует из некоторого допустимого состояния  $l_0 \in \{2, 3, \dots, N\}$ . Эволюция процесса описывается поглощающей марковской цепью  $\left\{ \xi_k^{(0)} \right\}_{k=0}^{\infty}$ , в которой граничные состояния являются поглощающими, а внутренние допустимые состояния — невозвратными. Как известно из теории марковских процессов, в некоторый конечный момент времени  $k_0$  процесс попадает в одно из поглощающих состояний. В примере, который иллюстрируется рис. 2,  $\xi_{k_0}^{(0)} = 1$ . После этого проводится внешнее воздействие (управление), в результате которого процесс переводится в некоторое внутреннее допустимое состояние  $l_1 \in \{2, 3, \dots, N\}$  с вероятностью  $\alpha_{l_1}^{(1)}$ ;  $\sum_{l=2}^N \alpha_l^{(1)} = 1$ . Аналогичное внешнее воздействие, осуществляемое при поглощении в состоянии 0, т. е. когда  $\xi_{k_0}^{(0)} = 0$ , описывается дискретным вероятностным распределением  $(\alpha_l^{(0)}, l = 2, 3, \dots, N)$ . После такого воздействия процесс независимо от прошлого начинает эволюционировать из состояния  $l_1$  по траектории новой поглощающей марковской цепи  $\left\{ \xi_k^{(1)} \right\}_{k=0}^{\infty}$ , вероятностные характеристики которой совпадают с характеристиками цепи  $\left\{ \xi_k^{(0)} \right\}_{k=0}^{\infty}$ . В момент достижения одного из граничных поглощающих состояний вновь проводится внешнее управляющее воздействие, которое описывается теми же дискретными распределениями вероятностей

$$\alpha^{(0)} = (\alpha_l^{(0)}, l = 2, 3, \dots, N);$$

$$\alpha^{(1)} = (\alpha_l^{(1)}, l = 2, 3, \dots, N).$$

Дальнейшая эволюция процесса осуществляется аналогично.

Перейдем к модели с непрерывным временем. Возьмем за основу этой модели полумарковский случайный процесс с дискретным множеством состояний. Общая теория таких процессов изложена в работе [5]. Из современных изданий укажем на работу [6].

Модели управляемых полумарковских процессов подробно рассмотрены в классических работах [7, 8].

Пусть  $\xi^{(n)}(t)$ ,  $n = 0, 1, 2, \dots$ , — последовательность независимых полумарковских процессов с поглощением и с одинаковыми вероятностными характеристиками. Обозначим через

$$X = \{0, 1, \dots, N\}, \quad N < \infty,$$

множество состояний данных процессов.

В дальнейшем будем предполагать, что для полумарковских процессов  $\xi^{(n)}(t)$ ,  $n = 0, 1, 2, \dots$ , заданы необходимые исходные вероятностные характеристики. Различные формы задания таких характеристик приведены в [5]. В частности, могут быть заданы полумарковские функции  $Q_{ij}(t)$ ,  $i, j \in X$ , которые представляют собой совместные распределения переходов вложенных цепей Маркова и длительностей пребывания процесса в различных состояниях.

Введем также две системы независимых неотрицательных случайных величин  $\{\Delta_n^{(0)}, n = 0, 1, 2, \dots\}$  и  $\{\Delta_n^{(1)}, n = 0, 1, 2, \dots\}$ , распределения которых для каждого фиксированного  $n$  будут зависеть от дополнительного условия, связанного с данной моделью.

Установим, что в структуре предлагаемой модели полумарковские процессы  $\xi^{(n)}(t)$ ,  $n = 0, 1, 2, \dots$ , будут описывать эволюцию исходной системы в течение периодов времени между управляющими воздействиями. Каждый процесс  $\xi^{(n)}(t)$  начинает эволюционировать в одном из допустимых состояний  $i \in \{2, 3, \dots, N\}$ . Через конечное время после начала эволюции процесс  $\xi^{(n)}(t)$  с вероятностью, равной 1, оказывается в одном из граничных поглощающих состояний  $\{0\}$  или  $\{1\}$ . После поглощения процесса  $\xi^{(n)}(t)$  модель подвергается так называемому управляющему воздействию, которое заключается в переходе из граничного состояния в одно из допустимых (внутренних) состояний. Вероятности перехода из граничного состояния 0 во внутренние допустимые состояния задаются вектором

$$\alpha^{(0)} = (\alpha_k^{(0)}, k = 2, 3, \dots, N), \quad \sum_{k=2}^N \alpha_k^{(0)} = 1,$$

вероятности перехода из граничного состояния 1 во внутренние допустимые состояния задаются вектором

$$\alpha^{(1)} = (\alpha_k^{(1)}, k = 2, 3, \dots, N), \sum_{k=2}^N \alpha_k^{(1)} = 1.$$

Распределения вероятностей  $\alpha^{(0)}$  и  $\alpha^{(1)}$  описывают внешние управляющие воздействия. Время данного перехода является случайной величиной  $\Delta_n^{(0)}$ , если переход происходит из состояния  $\{0\}$ , и случайной величиной  $\Delta_n^{(1)}$ , если переход осуществляется из состояния  $\{1\}$ . При этом распределение случайной величины  $\Delta_n^{(0)}$  или  $\Delta_n^{(1)}$  может, вообще говоря, зависеть от номера того допустимого состояния  $k \in \{2, 3, \dots, N\}$ , в которое осуществляется переход. После перехода в допустимое состояние  $k \in \{2, 3, \dots, N\}$  эволюция системы будет проходить независимо от прошлого (при фиксированном начальном состоянии  $k$ ) и описываться полумарковским процессом  $\xi^{(n+1)}(t)$ , вероятностные характеристики которого совпадают с соответствующими характеристиками процесса  $\xi^{(n)}(t)$ .

Задача оптимального управления в обеих версиях описанной модели будет формально ставиться как экстремальная задача на множестве пар дискретных вероятностных распределений  $(\alpha^{(0)}, \alpha^{(1)})$  по отношению к некоторому стационарному стоимостному показателю эффективности, имеющему смысл средней удельной прибыли.

Можно доказать [9, 10], что по своей аналитической форме указанный показатель эффективности представляет собой дробно-линейный интегральный функционал, заданный на множестве пар дискретных вероятностных распределений  $(\alpha^{(0)}, \alpha^{(1)})$ . Задача безусловного экстремума таких функционалов решена в работах [11, 12]. На основании указанных результатов можно утверждать, что решения поставленных задач оптимального управления достигаются на вырожденных вероятностных распределениях. Таким образом, решения о проведении управляющих воздействий должны приниматься детерминированно. Сами оптимальные решения определяются точкой максимума заданной функции двух дискретных переменных.

## 5 Основные особенности стохастической управляемой модели

Выделим несколько ключевых особенностей построенной стохастической управляемой модели.

1. Наличие в множестве состояний некоторого заданного подмножества так называемых допус-

тимых состояний и конечного подмножества состояний, которые можно условно называть внешними, или граничными по отношению к допустимым. При этом множество допустимых состояний необязательно должно быть дискретным

2. Выполнение некоторого общего условия, связанного с поведением исходного случайного процесса  $\xi^{(n)}(t)$  (непрерывное время) или  $\xi^{(n)}$  (дискретное время). Сущность этого условия состоит в том, что данный случайный процесс, начиная эволюцию из произвольного допустимого состояния, должен за конечное время достигать одного из внешних (граничных) состояний с вероятностью, равной единице.
3. Выполнение марковского свойства в моменты перехода во внешние (граничные) состояния, а также в моменты выхода из внешних состояний и перехода в одно из состояний допустимого подмножества.
4. Наличие возможности аналитического определения вероятностных характеристик, описывающих переход из произвольного фиксированного допустимого состояния в одно из внешних состояний (аналог вероятностей поглощения).
5. Наличие возможности аналитического определения математического ожидания времени, прошедшего от момента выхода из произвольного фиксированного допустимого состояния до попадания в одно из внешних состояний, и математического ожидания дохода, накопленного за указанное время.

Заметим, что условия 1–3 из сформулированного набора условий связаны с основными стохастическими свойствами модели. Условия 4 и 5 призваны обеспечить возможность нахождения аналитических представлений для необходимых вероятностных характеристик основного процесса, связанных со временем его пребывания в множестве допустимых состояний.

Указанные условия можно использовать при построении аналогичных моделей с управлениями, осуществляемыми в моменты выхода процесса на границу заданного подмножества множества состояний. В таких моделях могут использоваться иные виды основных процессов, описывающих эволюцию системы без управляющего воздействия. При этом можно поставить и решить соответствующие задачи о настройке.

## 6 Заключение

Формулировки теоретических утверждений, определяющих решения задач управления в опи-

санных выше моделях с дискретным и непрерывным временем, приведены в [9]. В работе [10] приведен полный анализ задачи управления с дискретным временем, включая доказательства основных утверждений.

## Литература

1. *Лопатников Л. И.* Экономико-математический словарь: Словарь современной экономической науки. — М.: Дело, 2003. 520 с.
2. OECD — Organisation for Economic Co-operation and Development: OECD Glossary of Statistical Terms (OECD Glossaries). — OECD Publishing, 2008. 288 p.
3. О развитии сельского хозяйства: Федеральный закон от 29.12.2006 № 264-ФЗ (ред. от 12.02.2015 с изм. и доп., вступ. в силу с 13.08.2015). <http://www.kremlin.ru/acts/bank/24837>.
4. *Кемени Дж. Дж., Снелл Дж. Л.* Конечные цепи Маркова / Пер. с англ.; под ред. А. А. Юшкевича. — М.: Наука, 1970. 272 с. (*Kemeny J. G., Snell J. L.* Finite Markov chains. — London: D. van Nostrand Co. Ltd., 1960. 210 p.)
5. *Королюк В. С., Турбин А. Ф.* Полумарковские процессы и их приложения. — Киев: Наукова думка, 1976. 184 с.
6. *Janssen J., Manca R.* Applied semi-Markov processes. — New York, NY, USA: Springer, 2006. 309 p.
7. *Джевелл В.* Управляемые полумарковские процессы // Кибернетич. сборник. Нов. серия. — М.: Мир, 1967. Вып. 4. С. 97–134. (*Jewell W. S.* Markov-renewal programming. I, II // Oper. Res., 1963. Vol. 11. No. 6. P. 938–971.)
8. *Майн Х., Осаки С.* Марковские процессы принятия решений / Пер. с англ. — М.: Наука, 1977. 176 с. (*Mine H., Osaki S.* Markovian decision processes. — New York, NY, USA: Elsevier, 1970. 142 p.)
9. *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). arXiv.org, 2017. arXiv: 1709.03442v1. 16 p.
10. *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. arXiv.org, 2018. arXiv: 1811.10993. 15 p.
11. *Шнурков П. В.* О решении задачи безусловного экстремума для дробно-линейного интегрального функционала на множестве вероятностных мер // Докл. Акад. наук, 2016. Т. 470. № 4. С. 387–392.
12. *Шнурков П. В., Горшенин А. К., Белоусов В. В.* Аналитическое решение задачи оптимального управления полумарковским процессом с конечным множеством состояний // Информатика и её применения, 2016. Т. 10. Вып. 4. С. 72–88.

Поступила в редакцию 15.07.20

---

# 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

P. V. Shnurkov and D. A. Novikov

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

**Abstract:** The work is devoted to the creation and analysis of the general concept of a special stochastic model with controls. The main feature of the model is that the control actions are carried out at times when a stochastic process describing the system under research reaches the boundary of a given subset of the set of states. The control action itself consists in transferring the process from the boundary to one of the internal states of a given subset. In this case, the internal states are interpreted as acceptable and the boundary ones as unacceptable. Control actions are described by a set of discrete probability distributions depending on the boundary state number. Such a set defines a control strategy. The problem of optimal control is formalized as the problem of finding a control strategy that delivers a global extremum to a certain stationary cost-effectiveness indicator, which in terms of its economic content represents the average specific profit arising from a long evolution of the system. The posed problem of optimal control is proposed to be solved by the tuning problem. The paper notes that this stochastic model and the corresponding setup problem can be used to study many real phenomena occurring in economic and technical systems. As an example of such a real phenomenon, interventions in the foreign exchange market of the Russian Federation are considered.

**Keywords:** control in stochastic systems; Markov controlled processes; semi-Markov controlled processes; stochastic tuning problem

**DOI:** 10.14357/19922264200315

## References

1. Lopatnikov, L. I. 2003. *Ekonomiko-matematicheskii slovar': slovar' sovremennoy ekonomicheskoy nauki* [Economic and mathematical dictionary: Modern economic science dictionary]. Moscow: Delo. 520 p.
2. OECD — Organisation for Economic Co-operation and Development. 2008. OECD glossary of statistical terms. OECD Publishing. 288 p. Available at: <https://stats.oecd.org/glossary/> (accessed July 30, 2020).
3. 264-FZ. 2015. O razvitii sel'skogo khozyaystva: federal'nyy zakon [About the development of agriculture: The Russian Federal Law]. Available at: <http://www.kremlin.ru/acts/bank/24837> (accessed July 30, 2020).
4. Kemeny, J., and J. Snell. 1976. *Finite Markov chains*. Prentice-Hall. 232 p.
5. Korolyuk, V. S., and A. F. Turbin. 1976. *Polumarkovskie protsessy i ikh prilozheniya* [Semi-Markov processes and their applications]. Kiev: Naukova dumka. 184 p.
6. Janssen, J., and R. Manca. 2006. *Applied semi-Markov processes*. New York, NY: Springer. 309 p.
7. Jewell, W. S. 1963. Markov-renewal programming. I, II. *Oper. Res.* 11(6):938–971.
8. Mine, H., and S. Osaki. 1970. *Markovian decision processes*. New York, NY: Elsevier. 142 p.
9. 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). arXiv.org. 16 p. Available at: <https://arxiv.org/abs/1709.03442v1> (accessed July 30, 2020).
10. 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. arXiv.org. 15 p. Available at: <https://arxiv.org/abs/1811.10993> (accessed July 30, 2020).
11. Shnurkov, P. V. 2016. Solution of the unconditional extremum problem for a linear fractional integral functional on a set of probability measures. *Dokl. Math.* 94(2):550–554.
12. Shnurkov, P. V., A. K. Gorshenin, and V. V. Belousov. 2016. Analiticheskoe reshenie zadachi optimal'nogo upravleniya polumarkovskim protsessom s konechnym mnozhestvom sostoyaniy [An analytic solution of the optimal control problem for a semi-Markov process with a finite set of states]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):72–88.

Received July 15, 2020

## 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)

**Novikov Daniil A.** (b. 1993) — PhD student, National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation; [even.he@yandex.ru](mailto:even.he@yandex.ru)

# ИЗВЛЕЧЕНИЕ ОПТИМИЗАЦИОННЫХ МОДЕЛЕЙ ИЗ ДАННЫХ

В. И. Донской<sup>1</sup>

**Аннотация:** Изложены основные принципы, методы и алгоритмы, представляющие новую информационную технологию извлечения оптимизационных математических моделей из данных (ИОМД). Эта технология позволяет автоматически строить математические модели планирования и управления на основе использования массивов прецедентов (наблюдений) над объектами управления и внешней средой, что дает возможность решать задачи интеллектуального управления и определять целесообразное поведение экономических и других объектов в сложных средах. Технология ИОМД позволяет получать объективные модели управления, отражающие реально существующие связи, цели, ограничения и процессы. В этом заключается ее главное преимущество по сравнению с традиционным, субъективным подходом к управлению. Разработаны линейные и нелинейные алгоритмы синтеза моделей по прецедентной информации.

**Ключевые слова:** машинное обучение; извлечение моделей из данных; оптимизация; нейронные сети; градиентные методы

**DOI:** 10.14357/19922264200316

## 1 Введение

В данной работе изложены математические модели и алгоритмы, составляющие основной аппарат *информационной технологии извлечения оптимизационных моделей из эмпирических данных* (обучающих выборок) в рамках парадигмы *неклассического информационного моделирования*, сложившейся в научных работах Ю. И. Журавлева, К. В. Рудакова, Вл. Д. Мазурова, а также их учеников и последователей [1–5], к которым относится и автор данной статьи. Исследования в указанном направлении предполагают применение методов машинного обучения для построения целевых функций и ограничений по достоверным обучающим выборкам — эмпирической информации, отражающей регулярные явления, объекты и процессы [6–10].

При таком подходе *построенные оптимизационные модели оказываются согласованными с реальной информацией о моделируемых объектах* в отличие от субъективных математических моделей, которые обычно предлагаются экспертами. Точные решения задач, соответствующих субъективным моделям, вследствие субъективности экспертов могут на практике давать ошибки, превышающие ошибки принятия решений на основе согласованных моделей, извлеченных из объективных эмпирических данных.

Будем обозначать  $X^n = \mathcal{X}_1 \times \dots \times \mathcal{X}_i \times \dots \times \mathcal{X}_n$  ограниченную область в  $\mathbb{R}^n$ , называемую пространством признаков размерности  $n$ ;  $X$  или  $\tilde{x} = (x_1, \dots, x_i, \dots, x_n)$  — произвольную точку в пространстве признаков, служащую описанием

допустимого объекта. Каждая координата  $x_i$  описания объекта  $\tilde{x}$  принадлежит зафиксированному *ограниченному множеству допустимых значений*:  $m_i \leq x_i \leq M_i$ . Будем далее обозначать  $\mathfrak{T}_{\text{Opt}} = \{(\tilde{a}_j, \gamma_j, y_j)\}_{j=1}^l$  *стандартную достоверную эмпирическую выборку для задачи обучения* — извлечения модели на основе *частичной информации о некотором существующем, но неизвестном скалярном критерии*  $F : X^n \rightarrow \mathbb{R}$  и *неизвестных ограничениях, которые формально могут быть представлены в виде*  $\Omega(\tilde{x}) = 0$ . Полагается, что множество  $X^n$  разбито на два класса: класс  $\mathcal{L}_0$ , состоящий из точек, заведомо удовлетворяющих некоторой системе ограничений задачи наилучшего выбора, и класс  $\mathcal{L}_1$ , содержащий точки, заведомо не удовлетворяющие этой системе ограничений. Тогда  $\Omega : X^n \rightarrow \{0; 1\}$  — *характеристическая функция ограничений, которая частично задана обучающей выборкой*;  $\Omega(\tilde{x}) = 0 \Leftrightarrow \tilde{x} \in \mathcal{L}_0$ ;  $\Omega(\tilde{x}) = 1 \Leftrightarrow \tilde{x} \in \mathcal{L}_1$ . Будем полагать, что в стандартной обучающей выборке  $\mathfrak{T}_{\text{Opt}}$  содержится достоверная информация  $\gamma_j = \Omega(\tilde{a}_j)$ ,  $\gamma_j \in \{0; 1\}$ ;  $y_j = F(\tilde{a}_j)$ ,  $\tilde{a}_j = (a_{j1}, \dots, a_{jn}) \in X^n$ .

В процессе обучения следует построить алгоритм, позволяющий выбрать такое решение  $\hat{\tilde{x}}^*$ , которое удовлетворяло бы ограничениям ( $\Omega(\hat{\tilde{x}}^*) = 0$ ) и при этом значение  $F(\hat{\tilde{x}}^*)$  было бы как можно большим (или меньшим — по смыслу задачи).

*В рассматриваемых задачах критерий  $F$  и ограничения (характеристическая функция  $\Omega$ ) не заданы точно. Они отражены в наборе данных  $\mathfrak{T}_{\text{Opt}}$  и являются частично заданными.*

<sup>1</sup>Крымский федеральный университет им. В. И. Вернадского, vidonskoy@mail.ru

Постановка задачи состоит в следующем. Требуется, используя частичную начальную информацию  $\mathfrak{T}_{\text{Opt}}$ , синтезировать модель и, соответственно, алгоритм, позволяющий выбрать решение  $\hat{x}^*$ , как можно более близкое к оптимальному решению  $\tilde{x}^*$ , определяемому неизвестными, но существующими истинными объектами  $F$  и  $\Omega$ , которые должны быть аппроксимированы в процессе машинного обучения. Если скалярный критерий и характеристическая функция ограничений аппроксимируются независимо друг от друга отдельными алгоритмами, вычисляющими как можно более точные в каком-либо смысле приближения  $\hat{F}$  и  $\hat{\Omega}$ , то *восстановленная по обучающей выборке задача (извлеченная математическая модель) нахождения наилучшего решения имеет следующий вид:*

$$\max(\min) \hat{F}(\tilde{x}) : \hat{\Omega}(\tilde{x}) = 0 \wedge \tilde{x} \in X^n .$$

Полученная в результате машинного обучения пара функций  $\langle \hat{F}, \hat{\Omega} \rangle$  называется *эмпирической информационно-математической моделью*.

## 2 Извлечение линейных оптимизационных моделей

Может оказаться, что знания о проблемной области заведомо определяют целесообразность извлечения линейной модели с вещественными переменными. Тогда можно сразу переходить к решению задачи, применяя методы, изложенные в данном разделе. Если же исходной информации о линейности нет, то сначала следует проверить: *не противоречат ли данные выборки  $\mathfrak{T}_{\text{Opt}}$  гипотезе о линейности неизвестных объектов  $F$  и  $\Omega$ , породивших эту выборку*. Гипотезу о линейности целевой функции можно проверить при помощи алгоритма, представленного в работе [10]. Гипотеза о линейности ограничений проверяется в процессе построения отделяющих гиперплоскостей алгоритмом 1. Если данные противоречат гипотезе линейности,

то следует применять методы извлечения нелинейных моделей, в частности основанные на обучении нейронных сетей.

В случае извлечения линейной оптимизационной модели с вещественными переменными из данных обучающей выборки  $\mathfrak{T}_{\text{Opt}} = \{(X_k, \gamma_k, y_k)\}_{k=1}^l$  неизвестными полагаются и линейная целевая функция

$$f(X) = \langle \Lambda, X \rangle = \lambda_1 x_1 + \dots + \lambda_i x_i + \dots + \lambda_n x_n + \lambda_0 ,$$

и линейные ограничения

$$AX \leq B; \quad A = [a_{ji}]_{m \times n} ,$$

$$X = [x_1, \dots, x_i, \dots, x_n]^T, \quad B = [b_1, \dots, b_j, \dots, b_m]^T .$$

В результате синтеза модели путем машинного обучения предполагается получить аппроксимации  $\hat{A}$  и  $\hat{B}$  матрицы  $A$  и вектора  $B$ . Значения  $\gamma_k$  выделяют точки обучающей выборки, соответствующие допустимым решениям, т. е. заведомо удовлетворяющие неравенствам  $AX \leq B$ , среди остальных точек выборки, не являющихся допустимыми. Иначе говоря, исходная информация представлена в форме

$$f(X_k) = y_k; \quad (\gamma_k = 0) \Leftrightarrow AX_k \leq B;$$

$$(\gamma_k = 1) \Leftrightarrow AX_k > B; \quad k \in \{1, \dots, l\} .$$

Обозначим

$$W_0 = \{X_k : \gamma_k = 0\}; \quad W_1 = \{X_k : \gamma_k = 1\} .$$

*Построение линейных отделяющих гиперплоскостей* проводится при помощи обучающей процедуры линейной коррекции Розенблатта–Новикова (ПЛКР) [11]:

$$\Lambda_{t+1} = \begin{cases} \Lambda_t, & \text{если } (\langle \Lambda_t, X_t \rangle > 0) \wedge (X_t \in W_1) \vee \\ & \vee (\langle \Lambda_t, X_t \rangle \leq 0) \wedge (X_t \in W_0); \\ \Lambda_t + cX_t, & \text{если } (\langle \Lambda_t, X_t \rangle \leq 0) \wedge (X_t \in W_1); \\ \Lambda_t - cX_t, & \text{если } (\langle \Lambda_t, X_t \rangle > 0) \wedge (X_t \in W_0), \end{cases}$$

### Алгоритм 1 Поиск исходного набора отделяющих гиперплоскостей

- 
- 1: Положить пустым множество отобранных гиперплоскостей:  $\mathfrak{L} := \emptyset$ .  
Очистить пометки всех точек множества  $W_1$ .  $i := 1$ .
  - 2: Выбрать из множества  $W_1$  непомеченную точку  $X$ , ближайшую к множеству  $W_0$ , и пометить ее.
  - 3: Используя ПЛКР, построить гиперплоскость  $\mathcal{L}_i$ , отделяющую эту точку  $X$  от всех точек множества  $W_0$ .  $i := i + 1$ .
  - 4: Пометить все такие непомеченные точки множества  $W_1$ , которые (как и точка  $X$ ) оказались отделенными гиперплоскостью  $\mathcal{L}_i$  от множества  $W_0$ .
  - 5: Запомнить построенную гиперплоскость:  $\mathfrak{L} := \mathfrak{L} \cup \mathcal{L}_i$ .
  - 6: Если в множестве  $W_1$  остались непомеченные точки, перейти к 2.
  - 7: Конец алгоритма. Построен набор гиперплоскостей  $\mathfrak{L} := \{\mathcal{L}_1, \dots, \mathcal{L}_s\}$ , где  $s = i - 1$ .
-

где  $0 < c \leq 1$ . К точкам обучающей выборки, предъявляемым на шагах  $t = 0, 1, 2, \dots$ , добавляется дополнительная координата с единичным значением:

$$X_t = (x_1^t, \dots, x_n^t, 1);$$

$$\Lambda_t = \langle \lambda_1^t, \dots, \lambda_n^t, \lambda_{n+1}^t \rangle; \Lambda_0 = (0, \dots, 0, 0).$$

Если выпуклые оболочки конечных множеств  $W_0$  и  $W_1$  не пересекаются, то найдется такой единичный вектор  $\Lambda^*$  и такое положительное вещественное число  $\rho$ , что  $\langle \Lambda^*, X \rangle < -\rho$  для  $X \in W_0$  и  $\langle \Lambda^*, X \rangle > \rho$  для  $X \in W_1$  [11].

При заведомой линейности модели каждая точка множества  $W_1$  может быть отделена некоторой гиперплоскостью от всех точек множества  $W_0$ . Идея синтеза аппроксимации области допустимых решений состоит в нахождении исходного набора гиперплоскостей, в совокупности отделяющих все точки множества  $W_1$  от множества  $W_0$ , с последующим выделением из этого набора минимального числа гиперплоскостей, обеспечивающих такое же совокупное отделение.

После выполнения алгоритма 1 осуществляется выбор минимального подмножества гиперплоскостей из построенного множества  $\mathcal{L}$ , которого достаточно для линейного разделения множеств  $W_0$  и  $W_1$ .

Введем логические значения  $\beta_{p,j} = 1$ , если гиперплоскость  $\mathcal{L}_j$  отделяет точку  $X_p \in W_1$  от всех точек множества  $W_0$ ;  $p = \overline{1, |W_1|}$ ,  $j = \overline{1, s}$ . Условие отделимости точки  $X_p$  хотя бы одной гиперплоскостью имеет вид:

$$\beta_{p,1}\mathcal{L}_1 \vee \dots \vee \beta_{p,j}\mathcal{L}_j \vee \dots \vee \beta_{l,s}\mathcal{L}_s. \quad (1)$$

Символы  $\mathcal{L}_j$ ,  $j = \overline{1, s}$ , в выражении (1) понимаются как *формальные логические переменные*, обозначающие соответствующие гиперплоскости. Если  $\beta_{p,j} = 1$ , то гиперплоскость  $\mathcal{L}_j$  включается в набор гиперплоскостей, отделяющих точку  $X_p$ , а если  $\beta_{p,j} = 0$ , то не включается. Формула

$$\bigwedge_{p=\overline{1, |W_1|}} (\beta_{p,1}\mathcal{L}_1 \vee \dots \vee \beta_{p,j}\mathcal{L}_j \vee \dots \vee \beta_{l,s}\mathcal{L}_s) \quad (2)$$

выражает требование отделимости каждой точки множества  $W_1$  от всех точек множества  $W_0$ . Выполняя логическое перемножение в формуле (2), а затем — все возможные поглощения, можно получить все варианты наборов отделяющих гиперплоскостей и выбрать из них кратчайший набор.

**Пример 1.** Пусть  $n = 2$  и 19 точек представлены в табл. 1 ( $W_0$ ) и 2 ( $W_1$ ).

Множество из семи прямых  $\mathcal{L}_1 - \mathcal{L}_7$ , отделяющих каждая по построению одну точку с номером

**Таблица 1** Множество точек  $W_0$

$W_0$	$x_1$	$x_2$
1	2	10
2	2	9
3	3	5
4	4	9
5	4	8
6	5	2
7	6	9
8	6	6
9	9	8
10	9	5
11	11	4
12	12	2

**Таблица 2** Множество точек  $W_1$

$W_1$	$x_1$	$x_2$
13	1	12
14	3	14
15	5	11
16	8	14
17	12	10
18	12	6
19	17	5

**Таблица 3** Построенное исходное множество прямых

№ точки $p$	Отделяющая эту точку прямая	Все отделенные точки
13	$\mathcal{L}_1 : 15x_1 - 3x_2 + 2 = 0$	13
14	$\mathcal{L}_2 : 6x_1 - 3x_2 + 19 = 0$	13–14
15	$\mathcal{L}_3 : -5x_2 + 52 = 0$	13–16
16	$\mathcal{L}_4 : -8x_1 - 4x_2 + 115 = 0$	16–19
17	$\mathcal{L}_5 : -19x_1 - 2x_2 + 244 = 0$	17–19
18	$\mathcal{L}_6 : -36x_1 - 2x_2 + 440 = 0$	17–19
19	$\mathcal{L}_7 : -18x_1 - 26x_2 + 170 = 0$	19

$p \in \{13, \dots, 19\}$  (в первом столбце таблицы) и, как оказалось после построения, некоторые другие точки множества  $W_1$  (третий столбец), представлено в табл. 3. Это множество избыточно. Логическое выражение, которое позволяет найти все возможные несократимые наборы отделяющих прямых, выписывается непосредственно по табл. 4 и имеет вид:

$$(\mathcal{L}_1 \vee \mathcal{L}_2 \vee \mathcal{L}_3) \wedge (\mathcal{L}_2 \vee \mathcal{L}_3) \wedge \mathcal{L}_3 \wedge (\mathcal{L}_3 \vee \mathcal{L}_4) \wedge$$

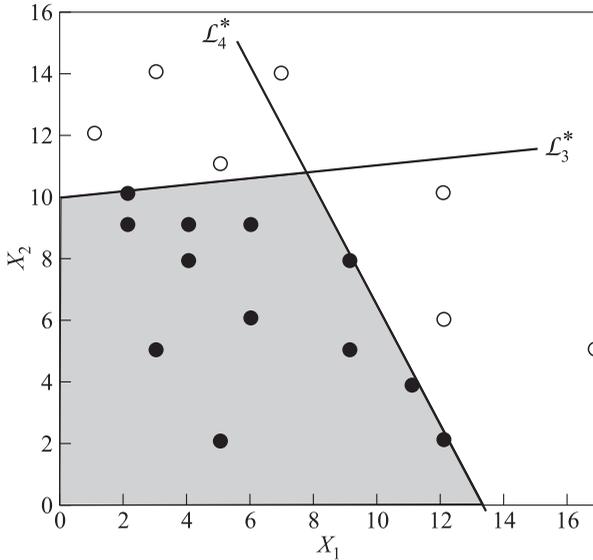
$$\wedge (\mathcal{L}_4 \vee \mathcal{L}_5 \vee \mathcal{L}_6) \wedge (\mathcal{L}_4 \vee \mathcal{L}_6) \wedge (\mathcal{L}_4 \vee \mathcal{L}_5 \vee \mathcal{L}_6 \vee \mathcal{L}_7) =$$

$$= \mathcal{L}_3\mathcal{L}_4 \vee \mathcal{L}_3\mathcal{L}_6.$$

Пара прямых  $\mathcal{L}_3\mathcal{L}_4$  более предпочтительна, поскольку прямая  $\mathcal{L}_4$  отделяет 4 точки, а прямая  $\mathcal{L}_6$  — три точки. Прямая  $\mathcal{L}_3$  отделяет точки  $\{X_{13}, X_{14}, X_{15}, X_{16}\} = \Omega_3$  от множества  $W_0$ . Вновь используя процедуру ПЛКР, находим уточненную

**Таблица 4** Логическая таблица отделимости множества  $W_0$

Прямые	Точки						
	13	14	15	16	17	18	19
$\mathcal{L}_1$	1						
$\mathcal{L}_2$	1	1					
$\mathcal{L}_3$	1	1	1	1			
$\mathcal{L}_4$				1		1	1
$\mathcal{L}_5$					1		1
$\mathcal{L}_6$					1	1	1
$\mathcal{L}_7$							1



**Рис. 1** Построенные отделители

прямую, разделяющую множества  $\Omega_3$  и  $W_0$ :  $\mathcal{L}_3^*$ :  $2x_1 - 17x_2 + 167 = 0$ . Аналогично находим  $\mathcal{L}_4^*$ :  $-12x_1 - 6x_2 + 158 = 0$ . Построенная область допустимых решений представлена на рис. 1.

*Начальная обучающая информация не противоречит гипотезе линейности ограничений, если ПЛКР обеспечивает нахождение отделяющего множества гиперплоскостей. Если число шагов коррекции превышает заданную величину, указанная линейная гипотеза отвергается.*

**Построение линейной целевой функции**

$$\hat{f}(X) = \langle \hat{\Lambda}, X \rangle = \hat{\lambda}_1 x_1 + \dots + \hat{\lambda}_i x_i + \dots + \hat{\lambda}_n x_n + \hat{\lambda}_0$$

по обучающей выборке  $\mathfrak{T}_{\text{Opt}} = \{(X_k, \gamma_k, y_k)\}_{k=1}^l$  несложно осуществить по методу наименьших квадратов. Но более предпочтительной для этой цели представляется линейная SVM (support vector machine) регрессия [12], применение которой сводится к оптимизационной задаче

$$\frac{1}{2} \langle \hat{\Lambda}, \hat{\Lambda} \rangle + C \sum_{k=1}^l (\xi_k^+ + \xi_k^-) \rightarrow \min_{\hat{\Lambda}, \hat{\lambda}_0, \hat{\xi}_k^+, \hat{\xi}_k^-};$$

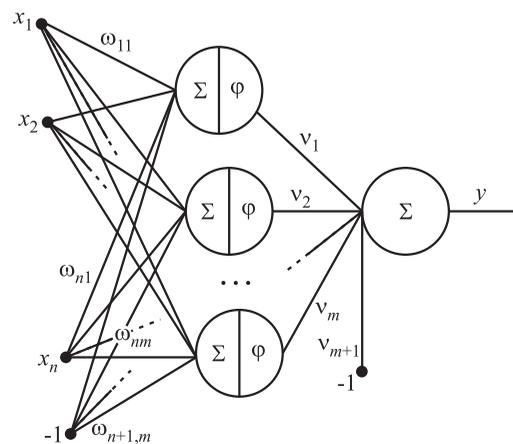
$$y_k - \varepsilon - \xi_k^- \leq \langle \hat{\Lambda}, X_k \rangle + \hat{\lambda}_0 \leq y_k + \varepsilon + \xi_k^+; \\ \xi_k^+ \geq 0; \xi_k^- \geq 0; k = 1, \dots, l,$$

где  $C$  — параметр регуляризации;  $\xi_k^+$  и  $\xi_k^-$  — дополнительные переменные, характеризующие величину ошибки в точках  $X_k$ ;  $\varepsilon$  — параметр задачи. SVM-регрессия обеспечивает минимизацию нормы вектора  $\hat{\Lambda}$ , что приводит к уменьшению верхней оценки колмогоровской сложности [8] синтезируемой модели и увеличению оценки неслучайности ее извлечения из данных.

### 3 Извлечение нелинейных оптимизационных моделей

Если имеется априорная информация о нелинейности искомой модели или в результате анализа обучающей информации гипотеза линейности отвергается, то полезным инструментом для извлечения эмпирических информационных моделей становятся нейронные сети. Известно, что нейронная сеть даже с одним скрытым слоем позволяет аппроксимировать с необходимой точностью широчайший класс функций, включая нелинейности, вообще говоря, любого вида [13].

Будем полагать, что по прецедентам, описывающим неизвестную целевую функцию  $F$ , строится аппроксимирующая ее нейронная сеть  $NN1$  с одним скрытым слоем и функцией активации  $\varphi(z) = (1 + \exp(-z))^{-1}$  (рис. 2), реализующая функцию  $\hat{F}_{NN1}(\tilde{x}) = \varphi(\tilde{x})$ . А по прецедентам, описывающим ограничения, обучается какая-нибудь другая классифицирующая нейронная сеть  $NN2$ , аппроксимирующая характеристическую функцию ограничений  $\Omega(\tilde{x})$ :  $\hat{\Omega}_{NN2}(\tilde{x})$ . Если сети  $NN1$  и  $NN2$



**Рис. 2** Сеть с одним скрытым слоем

**Алгоритм 2** Поиск условного (локального) минимума по  $\hat{F}_{NN1}$  и  $\hat{\Omega}_{NN2}(\tilde{x})$

Вход:	обучающая выборка $\mathfrak{X}_{\text{Opt}} = \{(\tilde{a}_j, \gamma_j, y_j)\}_{j=1}^l$ и нейронные аппроксимации $\hat{F}_{NN1}$ и $\hat{\Omega}_{NN2}(\tilde{x})$ .
Выход:	$\tilde{x}^* : \hat{F}_{NN1}(\tilde{x}^*) \approx \min \hat{F}_{NN1}(\tilde{x}) / \hat{\Omega}_{NN2}(\tilde{x}) = 0$ — точка экстремума эмпирической информационной модели и значение $y^*$ функции $\hat{F}_{NN1}$ в этой точке.
1:	Взять из обучающей выборки точку $\tilde{x}_0 = \tilde{a}_{j^*} : y_{j^*} = \min_j y_j$ в качестве начального приближения и вычислить $\hat{F}_{NN1}(\tilde{x}_0)$ .
2:	Выбрать начальное значение $\eta_0$ и значение $\varepsilon$ .
3:	Выбрать скорость уменьшения шага $\delta: 0,8 < \delta < 1$ .
4:	<b>for</b> $t := 1, 2, 3, \dots$ <b>do</b>
5:	$\tilde{x}_t := \tilde{x}_{t-1} - \eta_{t-1} \text{grad } \hat{F}_{NN1}(\tilde{x}_{t-1});$
6:	$\eta_t := \eta_{t-1} \cdot \delta;$
7:	<b>if</b> $\hat{\Omega}_{NN2}(\tilde{x}_t) = 1$ <b>then goto</b> 10;
8:	<b>if</b> $\ \hat{F}_{NN1}(\tilde{x}_t) - \hat{F}_{NN1}(\tilde{x}_{t-1})\  < \varepsilon$ <b>then goto</b> 11;
9:	<b>end for</b> ;
10:	$\tilde{x}^* := \tilde{x}_{t-1}; y^* := \hat{F}_{NN1}(\tilde{x}_{t-1});$ <b>stop</b> .
11:	$\tilde{x}^* := \tilde{x}_t; y^* := \hat{F}_{NN1}(\tilde{x}_t);$ <b>stop</b> .

обучены, то получена нейросетевая эмпирическая модель  $\langle \hat{F}_{NN1}(\tilde{x}), \hat{\Omega}_{NN2}(\tilde{x}) \rangle$  и можно перейти к нахождению определяемого ею экстремального решения:

$$\hat{F}_{NN1}(\tilde{x}) = v_{m+1} + \sum_{j=1}^m v_j \varphi \left( \omega_{n+1,j} + \sum_{i=1}^n \omega_{ij} x_i \right).$$

Обозначим  $A(\tilde{\omega}, \tilde{x}) = \omega_{n+1,j} + \sum_{i=1}^n \omega_{ij} x_i$ . Тогда

$$\begin{aligned} \frac{\partial \hat{F}_{NN1}(\tilde{x})}{\partial x_i} &= \\ &= \sum_{j=1}^m v_j \omega_{ij} \varphi(A(\tilde{\omega}, \tilde{x})) (1 - \varphi(A(\tilde{\omega}, \tilde{x}))), \quad i = \overline{1, n}; \end{aligned}$$

$$\begin{aligned} \text{grad } \hat{F}_{NN1} &= \\ &= \left( \frac{\partial \hat{F}_{NN1}(\tilde{x})}{\partial x_1}, \frac{\partial \hat{F}_{NN1}(\tilde{x})}{\partial x_2}, \dots, \frac{\partial \hat{F}_{NN1}(\tilde{x})}{\partial x_n} \right). \end{aligned}$$

Нахождение экстремума эмпирической информационной модели  $\langle \hat{F}_{NN1}, \hat{\Omega}_{NN2} \rangle$  можно осуществить градиентным алгоритмом 2. В общем случае функция  $\hat{F}_{NN1}$  может оказаться многоэкстремальной, и тогда этот алгоритм будет отыскивать локальный экстремум.

**При использовании полносвязных многослойных нейронных сетей** [14] изложенный метод принципиально не изменяется: требуется только умение вычислять градиент аппроксимирующей нейронной функции. Покажем, как вычислять этот градиент, полагая функции активации логистическими:  $\varphi(z) = (1 + e^{-z})^{-1}; \varphi'(z) = \varphi(z)(1 - \varphi(z))$ .

Будем использовать следующие обозначения:

$v_j^l$  — взвешенная сумма всех входов нейрона  $j$  слоя  $l$ , называемая его *индуцированным локальным полем* [14];

$l$  — номер слоя,  $0 \leq l \leq L$ ; слой с номером 0 — добавленный для облегчения понимания хода вычислений градиента сети слой входов; выходной слой имеет номер  $L$ ;

$m_1, m_2, \dots, m_{L-1}$  — число нейронов в скрытых слоях  $1, 2, \dots, L-1$ ;

$v_L$  — суммарный вход выходного нейрона;

$\tilde{x} = (x_1, \dots, x_i, \dots, x_n)$  — вход нейронной сети;

$y_j^l = \varphi(v_j^l)$  — выход нейрона  $j$  слоя  $l$ ;  $Y = Y(\tilde{x})$  — выход сети.

Для нахождения градиента функции, реализованной обученной нейронной сетью, будем использовать рекуррентную схему, которая лежит в основе алгоритма обучения нейронной сети методом обратного распространения ошибки [14].

Локальный градиент выходного нейрона определяется выражением:

$$\begin{aligned} \delta_L^{(L)} &= \frac{\partial Y}{\partial v_L} = \varphi'(v_L) = \\ &= \varphi(v_L)(1 - \varphi(v_L)) = Y(1 - Y), \end{aligned} \quad (3)$$

где верхний индекс в скобках обозначает номер слоя сети. Локальный градиент нейрона  $j$  скрытого слоя с номером  $l$ :

$$\begin{aligned} \delta_j^{(l)} &= \varphi'(v_j^l) \sum_k \delta_k^{(l+1)} \omega_{jk} = \\ &= y_j^{(l)} (1 - y_j^{(l)}) \sum_k \delta_k^{(l+1)} \omega_{jk}, \end{aligned} \quad (4)$$

где сумма берется по всем номерам нейронов слоя, непосредственно следующего за слоем, в котором

содержится нейрон  $j$ . Здесь и далее в формулах величины  $\omega_{jk}$  понимаются как веса от  $j$ -го нейрона слоя  $l$  к нейрону с порядковым номером  $k$  слоя  $l + 1$ :

$$\begin{aligned} \delta_j^{(1)} &= \varphi'(v_j^1) \sum_k \delta_k^{(2)} \omega_{jk} = y_j^{(1)} (1 - y_j^{(1)}) \sum_k \delta_k^{(2)} \omega_{jk}; \\ \delta_i^{(0)} &= \varphi_0'(x_i) \sum_k \delta_k^{(1)} \omega_{ik} = \sum_k \delta_k^{(1)} \omega_{ik}; \\ \frac{\partial Y}{\partial x_i} &= \delta_i^{(0)}; \quad \text{grad } Y = (\delta_1^{(0)}, \dots, \delta_n^{(0)}). \end{aligned} \quad (5)$$

Как и при вычислениях по методу обратного распространения ошибки, для заданного входа  $\tilde{x}$  сначала в прямом направлении от входа к выходу сети вычисляются локальные поля и выходы всех нейронов. Затем в обратном направлении, начиная с уравнения (3), рекуррентно выполняются вычисления локальных градиентов по формуле (4) и завершаются вычислением градиента по формулам (5).

#### 4 Эвристический подход к поиску глобального экстремума

Поиск глобального экстремума задачи, определенной моделью  $(\hat{F}_{NN1}, \hat{\Omega}_{NN2})$ , основан на повторяющемся применении алгоритма 2 нахождения локального экстремума, начиная с различных исходных точек области допустимых решений. Предварительно осуществляется кластеризация точек обучающей выборки  $\mathcal{T}_{\text{Opt}} = \{(\tilde{a}_j, \gamma_j, y_j)\}_{j=1}^l$  по

значениями целевой функции  $y_j$ . Для этой цели используется алгоритм построения решающего дерева, определяющего разбиение области допустимых значений переменных-признаков [8] на гиперпараллелепипеды [15]. В области, соответствующей каждому такому гиперпараллелепипеду, значения целевой функции  $y_j$  принадлежат зафиксированному полуинтервалу. Число таких полуинтервалов и есть число классов в предварительной задаче кластеризации. После построения классифицирующего дерева каждый класс получает логическое описание в терминах пороговых предикатов вида  $[x_i \leq b]$ ,  $b \in \mathbb{R}$ ,  $i \in \{1, \dots, n\}$ .

Необходимо из содержательных соображений, определяемых проблемной областью, определить две константы  $m$  и  $M$  такие, что  $m < y_j < M$  для всех  $j = \overline{1, l}$ . Затем отрезок  $[m, M]$  разбить на  $k$  равных сегментов  $[m, m + \Delta], \dots, [m + k\Delta, M]$ , где  $\Delta = (M - m)/k$ . Если точка  $\tilde{a}_j$  обучающей выборки попадает в сегмент с номером  $q$ ,  $q = \overline{1, k}$ , то эта точка считается относящейся к классу  $\mathcal{K}_q$ . Таким образом, в каждом из полученных классов окажутся точки, значения целевой функции в которых отличаются не более чем на  $\Delta$ .

Обозначим  $\bar{y}_q = m + q\Delta - \Delta/2$  середину  $q$ -го промежутка. Каждая концевая вершина дерева содержит значение  $\bar{y}_q$ , соответствующее номеру класса  $q$ , и специальную пометку-флажок, используемую для запоминания просмотренных листьев при поиске глобального экстремума.

Обоснование применения классифицирующего дерева при поиске глобального экстремума в составе алгоритма 3 состоит в следующем.

**Алгоритм 3** Эвристический поиск условного глобального минимума эмпирической информационной модели  $(\hat{F}_{NN1}, \hat{\Omega}_{NN2}(\tilde{x}))$

Вход:	обучающая выборка $\mathcal{T}_{\text{Opt}} = \{(\tilde{a}_j, \gamma_j, y_j)\}_{j=1}^l$ ; нейронные аппроксимации $\hat{F}_{NN1}$ и $\hat{\Omega}_{NN2}(\tilde{x})$ ; алгоритм 1 поиска локального минимума в качестве используемой внутренней процедуры; дерево кластеризации точек по значениям целевой функции в качестве второй используемой внутренней процедуры; наибольшее допустимое суммарное число шагов локального поиска $S$ .
Выход:	$\tilde{x}^* : \hat{F}_{NN1}(\tilde{x}^*) \approx \min \hat{F}_{NN1}(\tilde{x})/\hat{\Omega}_{NN2}(\tilde{x}) = 0$ — точка, являющаяся приближением к глобальному экстремуму эмпирической модели и значение $y^*$ функции $\hat{F}_{NN1}$ в этой точке.
1:	Очистить пометки просмотра всех листьев дерева (сброс всех флажков в ноль).
2:	Взять из обучающей выборки точку $\tilde{x}_0 = \tilde{a}_{j^*} : y_{j^*} = \min_j y_j$ в качестве начального приближения и выполнить алгоритм 1.
3:	Запомнить найденную точку $\tilde{x}^*$ и значение локального минимума $y^*$ в этой точке.
4:	Пометить лист дерева, в который попала точка $\tilde{x}^*$ (установить флажок в единицу).
5:	Выбрать непометенный лист дерева с наименьшим значением $\bar{y}_q$ и попавшую в него точку обучающей выборки с наименьшим значением целевой функции.
6:	Применить к выбранной точке алгоритм 1, получая экстремум $y^\circ$ в точке $\tilde{x}^\circ$ .
7:	Пометить выбранный лист как просмотренный (установить флажок в единицу).
8:	<b>if</b> $y^\circ < y^*$ <b>then</b> $\{y^* := y^\circ; \tilde{x}^* := \tilde{x}^\circ\}$ ;
9:	Если все еще есть непометенные листья дерева и число шагов локального поиска не превысило заданной величины $S$ , то перейти на 5.

Таблица 5 Выборка для обучения нейронной сети

№ предприятия	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$y$
1	1404,5	345,9	170	0,313	0,061	112,5
2	1709,8	431,9	225	0,285	0,067	113,7
3	1808,7	886,2	238	0,398	0,089	193,0
4	1437,1	484,2	181	0,322	0,076	125,0
5	1496,1	724,6	177	0,367	0,085	173,4
6	1034,3	200,7	179	0,206	0,039	81,4
7	1335,0	317,6	162	0,314	0,049	106,4
8	1256,1	156,1	159	0,187	0,038	72,6
9	1581,4	364,3	255	0,319	0,054	110,7
10	1826,5	554,2	275	0,338	0,083	146,3
11	1697,7	387,7	251	0,219	0,064	112,9
12	1294,6	302,5	154	0,214	0,044	105,9
13	1174,7	483,9	215	0,324	0,075	134,5
14	1180,9	220,1	165	0,217	0,039	91,4
15	1319,0	243,6	156	0,207	0,040	98,4
16	1460,0	347,3	202	0,316	0,053	107,6
17	1478,3	313,5	186	0,211	0,044	102,3

Примечание: Стоимостные показатели  $x_1$ ,  $x_2$  и  $y$  в тыс. долл. США приведены за один месяц.

1. Число точек  $l$  в обучающей выборке, как правило, многократно превышает число листьев дерева классификации, и вместо перебора всех  $l$  точек для инициализации локального поиска осуществляется повторение локального поиска не более  $\mu \ll l$  раз — каждый раз с начальной точкой, взятой из сегмента, соответствующего выбранному листу дерева.
2. Листьям дерева соответствуют разные, в том числе значительно удаленные друг от друга, сегменты области глобального поиска, что расширяет поиск, в то время как произвольный упорядоченный перебор точек обучающей выборки для инициализации поиска может привести к вычислениям в окрестности одного и того же локального минимума.

**Пример 2.** Чтобы проиллюстрировать подход к синтезу нейронной нелинейной оптимизационной модели, была использована обучающая выборка по 17 производственным предприятиям (табл. 5). Наблюдения, содержащиеся в табл. 5, соответствовали только допустимым точкам ( $\Omega(\tilde{x}) = 1$ ) и описывались пятью признаками, обозначаемыми далее как  $x_1$ – $x_5$ :  $x_1$  — объем валовой продукции в стоимостном выражении;  $x_2$  — стоимость основных фондов;  $x_3$  — число работников;  $x_4$  — коэффициент специализации<sup>1</sup>;  $x_5$  — коэффициент маркетинга<sup>2</sup>. Целевым атрибутом  $y$  выступала прибыль.

Использовалась нейронная сеть с одним скрытым слоем (см. рис. 2)<sup>3</sup> и сигмоидными активационными функциями. Ошибка нейронной сети оценивалась как

$$E(\tilde{w}, \tilde{V}) = \frac{1}{2} \sum_{j=1}^{17} \left( f_{NN}(X_j^{\text{sample}}) - y_j^{\text{sample}} \right)^2,$$

где  $f_{NN}(X_j^{\text{sample}})$  — выход нейронной сети на примере с номером  $j$ :

$$f_{NN}(X_j^{\text{sample}}) = -V_o + \sum_{q=1}^m V_q \varphi \left( -w_{0,q} + \sum_{i=1}^n w_{i,q} x_i \right);$$

$n = 5$  — число переменных;  $m = 6$  — число нейронов в скрытом слое;  $\tilde{w} = \{w_{i,q}, i = \overline{0, n}; q = \overline{1, m}\}$  — входные веса сети;  $X_j^{\text{sample}} = (x_1^j, \dots, x_n^j)$  —  $j$ -я строка (пример) таблицы обучения;  $V_q, q = \overline{1, m}$ , — выходные веса скрытого слоя сети;  $-V_o$  — смещение;  $y_j^{\text{sample}}$  — значение прибыли на  $j$ -м примере обучающей выборки. Для вычислений были использованы следующие формулы:

$$\eta_k = \varphi \left( -w_{0,k} + \sum_{s=1}^m w_{s,k} x_s \right);$$

$$\frac{\partial f_{NN}}{\partial w_{i,q}} = \sum_{k=1}^m V_k x_i \eta_k (1 - \eta_k), \quad i = \overline{1, n}; \quad q = \overline{1, m};$$

<sup>1</sup> Отношение стоимости готовой продукции профильного направления производства к стоимости всей готовой продукции, выпускаемой предприятием за месяц.

<sup>2</sup> Линейная свертка коэффициентов, определяемых значениями маркетинговых индикаторов.

<sup>3</sup> Доказано, что сети такого класса позволяют получать требуемую точность при обучении регрессии [13].

$$\frac{\partial f_{NN}}{\partial w_{0,q}} = - \sum_{k=1}^m V_k \eta_k (1 - \eta_k), \quad q = \overline{1, m};$$

$$\frac{\partial f_{NN}}{\partial V_k} = \eta_k, \quad k = \overline{1, m};$$

$$\frac{\partial f_{NN}}{\partial V_0} = -1;$$

$$\frac{\partial E(\tilde{w}, \tilde{V})}{\partial w_{i,q}} = \left( f_{NN}(X_j^{\text{sample}}) - y_j^{\text{sample}} \right) \frac{\partial f_{NN}}{\partial w_{i,q}};$$

$$\frac{\partial E(\tilde{w}, \tilde{V})}{\partial V_k} = \left( f_{NN}(X_j^{\text{sample}}) - y_j^{\text{sample}} \right) \eta_k.$$

Для обучения был использован градиентный спуск:

$$w_{i,q}^t := w_{i,q}^{t-1} - \gamma_t \frac{\partial E(\tilde{w}, \tilde{V})}{\partial w_{i,q}};$$

$$V_q^t := V_q^{t-1} - \gamma_t \frac{\partial E(\tilde{w}, \tilde{V})}{\partial V_k},$$

где  $\gamma_t, t = 1, 2, \dots$ , — уменьшающаяся с ростом  $t$  скорость обучения. В ходе 94 эпох обучения наблюдалось монотонное уменьшение средней ошибки  $(1/l) \sum_{j=1}^{17} \left( f_{NN}(X_j^{\text{sample}}) - y_j^{\text{sample}} \right)^2$  от величины 1,0878 до 0,0277. После того как нейронная сеть была обучена, решалась задача нахождения максимума реализованной ею функции прибыли при следующих искусственно введенных ограничениях, позволяющих задать область допустимых решений: значения признаков-факторов ограничивались сегментами

$$0,9 \min_{j=1,17} x_{j,i}^{\text{sample}} \leq x_i \leq 1,1 \max_{j=1,17} x_{j,i}^{\text{sample}}.$$

Для градиентного подъема использовались формулы:

$$\frac{\partial f_{NN}}{\partial x_i} = \sum_{q=1}^m V_q \eta_q (1 - \eta_q) w_{i,q};$$

$$x_i^t := x_i^{t-1} + \gamma_t \Delta y^t \frac{\partial f_{NN}}{\partial x_i},$$

где  $\Delta y^t$  — изменение значения на выходе нейронной сети на шаге  $t$ . Было выполнено 250 шагов оптимизационного градиентного поиска и найдено экстремальное значение прибыли  $Y = 195,54$  при значениях аргументов  $x_1 = 1096,56, x_2 = 934,67, x_3 = 199,94, x_4 = 0,392$  и  $x_5 = 0,0371$ . Заметим, что максимальное значение прибыли в обучающей выборке  $\left( \max_{j=\overline{1,17}} y_j^{\text{sample}} \right)$  равно 193,0.

## 5 Заключение

Исследования в области применения методов машинного обучения и распознавания для синтеза

оптимизационных моделей из данных с целью построения интеллектуализированных систем управления позволили получить результаты в виде ряда готовых алгоритмов и программ, обеспечивающих синтез таких моделей и их обоснование [8–10]. Изложенные в данной статье результаты расширяют указанные исследования и обобщают их до уровня *новой информационной технологии* — ИОМД, в англоязычной версии — *Building Optimization Models from Data* (BOMD). В рамках технологии ИОМД созданы все необходимые алгоритмы синтеза оптимизационных моделей, и они могут быть широко использованы на практике, позволяя расширить методы искусственного интеллекта, применяемые в цифровой экономике и других прикладных областях.

Ограниченный объем статьи не позволил включить в изложенный материал методы оценивания надежности моделей, извлеченных из данных. Подход к такому оцениванию, основанный на теории колмогоровской алгоритмической сложности, изложен в работе [8].

## Литература

1. Мазуров Вл. Д. Применение методов теории распознавания образов в оптимальном планировании и управлении // Тр. I Всесоюз. конф. по оптимальному планированию и управлению народным хозяйством. — М.: ЦЭМИ, 1971. С. 1–49.
2. Журавлёв Ю. И. Экстремальные алгоритмы в математических моделях для задач распознавания и классификации // Докл. Акад. наук СССР, 1976. Т. 231. № 3. С. 532–535.
3. Ерёмин И. И., Мазуров Вл. Д. Нестационарные процессы математического программирования. — М.: Наука, 1979. 288 с.
4. Рудаков К. В. Об алгебраической теории универсальных и локальных ограничений для задач классификации // Распознавание, классификация, прогноз. — М.: Наука, 1989. С. 176–201.
5. Воронцов К. В. Аддитивная регуляризация тематических моделей коллекций текстовых документов // Докл. Акад. наук, 2014. Т. 456. № 3. С. 268–271.
6. Махина Г. А. Нейросетевой подход к задачам слабоопределенной оптимизации // Искусственный интеллект, 2000. № 2. С. 145–148.
7. Таратынова Н. Ю. Построение оптимизационной модели по прецедентной начальной информации как задача нелинейной регрессии // Искусственный интеллект, 2006. № 2. С. 238–241.
8. Донской В. И. Синтез согласованных линейных оптимизационных моделей по прецедентной информации: подход на основе колмогоровской сложности // Таврический вестник информатики и математики, 2012. № 1. С. 13–25.

9. Донской В. И. Извлечение оптимизационных моделей из данных: подход на основе решающих деревьев и лесов // Таврический вестник информатики и математики, 2017. № 4. С. 59–86.
10. Donskoy V. I. A synthesis of pseudo-boolean empirical models by precedential information // Вестник Южно-Уральского государственного университета. Сер. Математическое моделирование и программирование, 2018. Т. 11. №. 2. С. 96–107.
11. Nilsson N. J. Learning machines. — New York, NY, USA: McGraw-Hill, 1965. 137 p.
12. Rokach L., Scholkopf B. A tutorial on support vector regression // Stat. Comput., 2004. Vol. 14. No. 4. P. 199–222.
13. Hornik K., Stinchcombe M., White H. Universal approximation of an unknown mapping and derivatives using multilayer feedforward networks // Neural Networks, 1990. Vol. 3. No. 1. P. 551–560.
14. Haykin S. Neural networks and learning machines. — New York, NY, USA: Prentice Hall, 2008. 906 p.
15. Breiman L., Friedman J. H., Olshen R., Stone C. J. Classification and regression trees. — New York, NY, USA: Chapman and Hall, 1984. 369 p.

Поступила в редакцию 20.05.19

## OPTIMIZATION MODELS EXTRACTION FROM DATA

V. I. Donskoy

V. I. Vernadsky Crimean Federal University, 4 Vernadsky Av., Simferopol 295007, Russian Federation

**Abstract:** The basic principles, methods and algorithms representing a new information technology for building optimization mathematical models from data (BOMD) are presented. This technology allows one to automatically build mathematical models of planning and control on the basis of use of precedents (observations) over objects that gives the chance to solve the problems of intellectual control and to define expedient behavior of economic and other objects in difficult environments. The BOMD technology allows one to obtain objective control models that reflect real-life relationships, goals, constraints, and processes. This is its main advantage over the traditional, subjective approach to control. Linear and nonlinear algorithms for synthesis of models based on precedent information are developed.

**Keywords:** machine learning; model extraction from data; optimization; neural networks; gradient methods

**DOI:** 10.14357/19922264200316

### References

1. Mazurov, V. I. D. 1971. Primenenie metodov teorii raspoznavaniya obrazov v optimal'nom planirovanii i upravlenii [Application of pattern recognition methods in optimal planning and management]. *Tr. I Vsesoyuzn. konf. po optimal'nomu planirovaniyu i upravleniyu narodnym khozyaystvom* [1st All-Union Conference on Optimal Planning and Management of the National Economy Proceedings]. Moscow. 1–49.
2. Zhuravlev, Yu. I. 1976. Ekstremal'nye algoritmy v matematicheskikh modelyakh dlya zadach raspoznavaniya i klassifikatsii [Extremal algorithms in mathematical models for discrimination and classification problems]. *Dokl. Akad. Nauk SSSR* 231(3):532–535.
3. Eremin, I. I., and V. D. Mazurov. 1979. *Nestatsionarnye protsessy matematicheskogo programmirovaniya* [Non-stationary processes of mathematical programming]. Moscow: Nauka. 288 p.
4. Rudakov, K. V. 1989. Ob algebraicheskoy teorii universal'nykh i lokal'nykh ogranicheniy dlya zadach klassifikatsii [On the algebraic theory of universal and local constraints for classification problems]. *Raspoznavanie, klassifikatsiya, prognoz* [Recognition, classification, forecast]. Moscow: Nauka. 176–201.
5. Vorontsov, K. V. 2014. Additive regularization for topic models of text collections. *Dokl. Math.* 89(3):301–304.
6. Makhina, G. A. 2000. Neyrosetevoy podkhod k zadacham slaboopredelennoy optimizatsii [Neural network approach to weakly defined optimization problems]. *Iskusstvennyy intellekt* [Artificial Intelligence] 2:145–148.
7. Taratynova, N. Yu. 2006. Postroenie optimizatsionnoy modeli po pretседentnoy nachal'noy informatsii kak zadacha nelineynoy regressii [Optimization model construction by precedents as a nonlinear regression problem]. *Iskusstvennyy intellekt* [Artificial Intelligence] 2:238–241.
8. Donskoy, V. I. 2012. Sintez soglasovannykh lineynykh optimizatsionnykh modeley po pretседentnoy informatsii: podkhod na osnove Kolmogorovskoy slozhnosti [Synthesis of coordinated linear optimization models according to precedential information: An approach based on Kolmogorov complexity]. *Tavricheskiy vestnik informatiki i matematiki* [Taurida J. Computer Science Theory Mathematics] 2:13–25.
9. Donskoy, V. I. 2017. Izvlechenie optimizatsionnykh modeley iz dannykh: podkhod na osnove reshayushchikh derev'ev i lesov [Extracting optimization models from data: An approach based on decision trees and forests]. *Tav-*

- richeskiy vestnik informatiki i matematiki* [Taurida J. Computer Science Theory Mathematics] 4:59–86.
10. Donskoy, V. I. 2018. A synthesis of pseudo-boolean empirical models by precedential information. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Ser. Matematicheskoe modelirovanie i programmirovaniye* [Bull. South Ural State University. Ser. Mathematical Modelling Programming Computer Software] 11(2):96–107.
  11. Nilsson, N.J. 1965. *Learning machines*. New York, NY: McGraw-Hill. 137 p.
  12. Rokach, L., and B. Scholkopf. 2004. A tutorial on support vector regression. *Stat. Comput.* 14(4):199–222.
  13. Hornik, K., M. Stinchcombe, and H. White. 1990. Universal approximation of an unknown mapping and derivatives using multilayer feedforward networks. *Neural Networks* 3(1):551–560.
  14. Haykin, S. 2008. *Neural networks and learning machines*. New York, NY: Prentice Hall. 906 p.
  15. Breiman, L., J. H. Friedman, R. Olshen, and C.J. Stone. 1984. *Classification and regression trees*. New York, NY: Chapman and Hall. 369 p.

Received May 20, 2019

## Contributor

**Donskoy Vladimir I.** (b. 1948) — Doctor of Science in physics and mathematics, professor, V. I. Vernadsky Crimean Federal University, 4 Vernadsky Av., Simferopol 295007, Russian Federation; vidonskoy@mail.ru

# ПРОБЛЕМНО-ОРИЕНТИРОВАННАЯ ВЕРИФИКАЦИЯ ПОЛНОТЫ ТЕМПОРАЛЬНЫХ ОНТОЛОГИЙ И ЗАПОЛНЕНИЕ ПОНЯТИЙНЫХ ЛАКУН\*

И. М. Зацман<sup>1</sup>

**Аннотация:** Предлагается подход к проверке полноты онтологии и заполнению обнаруженных в ней понятийных лакун с применением следующих симбиотических информационных процессов: целенаправленное извлечение нового знания из данных, его представление в онтологии и применение для решения некоторой проблемы. В процессе ее решения осуществляется верификация полноты онтологии, регистрируются и заполняются ее лакуны, выявленные при решении именно этой проблемы. Разделяются личностный, коллективный и конвенциональный уровни представления знания в онтологии. Этот подход позволяет обнаружить лакуны на конвенциональном уровне онтологии и заполнить их на ее личностном и/или коллективном уровне, если для извлечения нового знания доступны его потенциальные источники. Цель статьи — рассмотреть модель перечисленных симбиотических процессов. Разработанная модель представляет собой обобщенную блок-схему реализации предлагаемого подхода. Блок-схема служит основой компьютеризации симбиотических процессов. Описание модели иллюстрируется примером обнаружения и заполнения понятийных лакун в лингвистической типологии концептами нового знания, извлекаемого из текстовых данных.

**Ключевые слова:** трехуровневое представление знания; темпоральная онтология; понятийная лакуна; генерация нового знания; симбиотические информационные процессы

**DOI:** 10.14357/19922264200317

## 1 Введение

Разработка принципов создания информационных систем, обеспечивающих симбиоз процессов извлечения из текстовых данных нового знания о языке, его представления в лингвистической типологии и его применения, является одной из задач проекта по гранту РФФИ, который в настоящее время выполняется в Институте проблем информатики ФИЦ ИУ РАН. Отметим, что лингвистические типологии как частный случай онтологий принадлежат левой части их линейного спектра (см. рис. 2 в [1]), где расположены глоссарии и тезаурусы.

В результате выполнения проекта был разработан подход к проверке полноты лингвистической типологии и применению симбиотических процессов для заполнения на ее личностном и коллективном уровнях тех понятийных лакун, которые обнаружены на конвенциональном уровне. Для решения задач компьютерной лингвистики на основе этого подхода была предложена модель для описания проблемно-ориентированной верификации полноты типологии, обнаружения и регистрации лакун, а также их заполнения концептами ново-

го знания, извлекаемого из текстовых данных [2, 3]. В статье предлагаемая модель описывается в обобщенной форме, без привязки к лингвистическим типологиям и задачам компьютерной лингвистики, что позволяет применять ее в других предметных областях как инструмент проблемно-ориентированной верификации полноты онтологий и заполнения их понятийных лакун.

Обобщенная модель позиционируется как развитие следующих трех моделей:

- (1) спиральная модель генерации нового знания, предложенная Нонака [4–6];
- (2) модель Вежбицкого–Накамори (далее — WN-модель) [7–9];
- (3) модель Ниссена [10, с. 36].

В спиральной модели генерируемое новое знание подразделяется на два вида и, соответственно, на два уровня представления (личностный и коллективный). В WN-модели к этим двум уровням добавлен третий уровень представления конвенционального знания. В модели Ниссена коллективный уровень делится на два подуровня: новое зна-

\* Исследование выполнено в Институте проблем информатики ФИЦ ИУ РАН при финансовой поддержке РФФИ в рамках научного проекта № 18-07-00192.

<sup>1</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, izatsman@yandex.ru

ние группы и организации. Предлагаемая в статье модель дает возможность учесть степень социализации нового знания в коллективе, изменение которой фиксируется с единичным шагом.

Кроме уровней представления знания в этих моделях для описания процессов генерации нового знания явно или неявно используются, как правило, три разнородных измерения пространства<sup>1</sup>. При этом процессы генерации охватывают объекты двух сред: ментальной, включающей концепты знания человека, и информационной, содержащей перцептивные формы их представления, например в виде слов естественного языка (как сочетаний графических символов).

Во всех трех перечисленных моделях в явном виде не специфицированы источники нового знания, не эксплицирована цель его генерации и не даны алгоритмы извлечения концептов нового знания. В предлагаемой модели цель эксплицируется следующим образом: обнаружение понятийных лакун на конвенциональном уровне онтологии и их заполнение на ее личностном и коллективном уровнях концептами, извлеченными из специфицированного множества потенциальных источников нового знания. Поэтому для нее предлагается название — *модель обнаружения и заполнения лакун* (ОЗЛ-модель, англоязычное название — *finding and filling model*).

Цель статьи — описать ОЗЛ-модель как обобщенную блок-схему на основе трех моделей (спиральной, WN и Ниссена), развивающую их по пяти позициям:

- (1) целенаправленность процессов обнаружения лакун в онтологии и их заполнения, ориентированных на решение явно сформулированной проблемы;
- (2) симбиоз следующих четырех процессов: (а) проблемно-ориентированное применение онтологии; (б) регистрация понятийных лакун, препятствующих решению проблемы, с помощью тегов неполноты; (в) извлечение концептов нового знания для заполнения лакун; (г) представление новых концептов рубриками *понятий онтологии* на ее личностном и/или коллективном уровне;
- (3) специфицирование потенциальных источников концептов нового знания для обеспечения проблемно-ориентированной верификации полноты онтологии, обнаружения и заполнения ее понятийных лакун;

<sup>1</sup>Отметим, что термин «пространство» здесь не является математическим. Авторами статей [7–10] пространство трактуется как сочетание некоторого числа осей (которые могут быть и номинативными шкалами), используемых для описания процессов генерации нового знания. Примеры таких сочетаний приводятся далее в статье.

- (4) расширение числа сред до трех, включая цифровую среду компьютеров и сетей, ментальную и информационную среды;
- (5) использование количественных осей для двух из трех измерений пространства для описания процессов обнаружения и заполнения лакун, а также для определения числовых характеристик этих процессов и их динамики.

До начала выполнения четырех процессов, перечисленных выше в п. 2, заранее не известно, где и когда в онтологии будут обнаружены лакуны, из каких именно источников будут извлечены концепты нового знания для их заполнения. По этой причине элементы специфицированного массива данных названы *потенциальными источниками концептов нового знания*. В проведенных экспериментах размеры предварительно специфицированного массива текстовых данных как минимум на порядок превышали те текстовые фрагменты массива, из которых потом действительно были извлечены концепты нового знания. В одном из экспериментов только 59 из 2503 аннотированных источников (2,36%) включали теги неполноты и из них были извлечены четыре концепта нового знания [11]. При этом процесс их извлечения можно охарактеризовать, скорее всего, как *вероятностно предсказуемый*.

## 2 Исторический контекст создания модели обнаружения и заполнения лакун

Прежде чем дать описание ОЗЛ-модели, приведем краткий обзор спиральной модели, на основе которой были созданы WN-модель и модель Ниссена. В настоящее время спиральная модель генерации нового знания стала одной из самых популярных качественных моделей [4, 6]. В ней определены две категории знания: индивидуальное и коллективное. Каждая из этих категорий делится, в свою очередь, на две подкатегории: выраженное, или явное (*explicit*), знание и неявное (*tacit*) знание, т. е. она включает следующие четыре вида знания: выраженное и неявное индивидуальное, выраженное и неявное коллективное. Помимо этих видов знания в модели определены следующие *четыре вида* процессов трансформации знания и форм его представления: социализация, экстернализация, комбинирование, интернализация, а также понятие «виток спирали».

По определению [4, 6], каждый из витков включает последовательность: *социализация* (формирование коллективного знания на основе индивидуального) → *экстернализация* (выражение коллективного знания перцептивными формами его представления, например словами) → *комбинирование* (создание индивидуальных форм представления знания на основе коллективных) → *интернализация* (личностное понимание, т. е. трансформация перцептивных форм в индивидуальное знание) → *социализация* (начало следующего витка спирали).

Согласно Братиану [12], «спиральная модель не содержит время в качестве явной переменной. . . Модель содержит время неявно, так как любое преобразование требует времени, но это время *без какой-либо возможности его измерения* (курсив мой — ИЗ)».

Таким образом, в этой модели есть три измерения:

- (1) измерение *социализации*, на шкале которого определены два номинативных значения (индивидуальное и коллективное знание);
- (2) *темпоральное* измерение (присутствует неявно);
- (3) *эксплицитное* измерение, на шкале которого также определены два номинативных значения (неявное знание и явное, т. е. выраженное).

Ниже будет показано, что три из четырех видов процессов трансформации знания и форм его представления, а также эти три измерения, присутствующие также в WN-модели и модели Ниссена, используются в ОЗЛ-модели в явном виде и при этом первые два измерения являются числовыми. Число видов процессов трансформации в ОЗЛ-модели будет увеличено, так как она включает цифровую среду и ее границы с ментальной и информационной средами.

### 3 Модель обнаружения и заполнения лакун

Сначала сформулируем критерий новизны знания в рамках ОЗЛ-модели. Концепт знания, извлеченный из множества потенциальных источников в некоторый момент времени, считается новым, если он не был описан в онтологии до этого момента времени. Предполагается, что эксперты до начала ими процесса верификации полноты онтологии уже отразили в ней конвенциональное знание, относящееся к ее предметной области.

Пять позиций, сформулированных в первом разделе, по которым ОЗЛ-модель развивает спиральную модель, WN-модель и модель Ниссена, были получены в процессе наблюдения за работой лингвистов, исследующих смысловое содержание немецких модальных глаголов, которые могут иметь более десяти значений. Цель лингвистов заключалась в верификации полноты существующей типологии конвенциональных значений глаголов, обнаружении в ней лакун и их заполнении концептами нового знания, извлекаемого в процессе лингвистического аннотирования фрагментов текстов с модальными глаголами (см. таблицу) [13–16]. Работу лингвистов можно описать с помощью четырех симбиотических процессов:

- (1) применение рубрик типологии значений модальных глаголов при лингвистическом аннотировании [17] фрагментов параллельных текстов (конвенциональный уровень типологии был создан на основе словаря [18]);
- (2) регистрация понятийной лакуны, обнаруженной на конвенциональном уровне, с помощью тега неполноты, если в типологии нет рубрики, соответствующей значению глагола, найденному в процессе аннотирования;
- (3) извлечение из фрагмента параллельного текста и его аннотации с тегом неполноты кон-

Два из 16 268 потенциальных источников концептов нового знания

Оригинальный текст	Перевод
— «Gut», sagte ich, «soll er dich verehren, aber soviel kostbare Blumen, das ist aufdringlich. . .» <i>Геррих Бёль. Ansichten eines Clowns (1963)</i>	— Очень мило, — сказал я, — поклонник поклонником, но дарить такой большой букет дорогих цветов — значит навязываться. . . <i>Глазами клоуна. Пер. Р. Райт-Ковалева (1964)</i>
Soll Wilke sich irgendeinen Reim darauf machen! Wenn Werners Sarg de-swegen nicht fertig wird, so ist das kein Unglück — der Bankier hat Dutzende von kleinen Hausbesitzern mit Inflationsgeld um ihr bisschen-Besitz gebracht. <i>Эрих Мария Ремарк. Der schwarze Obelisk (1956)</i>	Пусть Вильке вообразит себе какую-нибудь небыллицу! Если гроб Вернера из-за всего этого не будет закончен — не беда: этот банкир, пользуясь инфляцией, лишил десятки мелких домовладельцев их жалкой собственности. <i>Черный обелиск. Пер. В. Станевич (1961)</i>

цепта нового знания о значении того модального глагола, которое встретилось во фрагменте, но отсутствует в типологии;

- (4) включение в типологию новых рубрик и соответствующих им дефиниций, создаваемых на основе обнаруженных новых значений глаголов, что позволяет заполнить обнаруженные при аннотировании понятийные лакуны.

Для формирования множества потенциальных источников концептов нового знания лингвисты использовали параллельные немецко-русские тексты, полученные из немецкого подкорпуса Национального корпуса русского языка [19]. Эти тексты включали 2,6 млн словоупотреблений (1,4 млн словоупотреблений в оригинальных текстах на немецком языке и 1,2 млн словоупотреблений в их переводах на русский язык). В результате выравнивания оригиналов и переводов было сформировано множество, которое состояло из 83 190 потенциальных источников нового знания. Из них 16 268 соответствовали цели исследования, т. е. содержали немецкие модальные глаголы (см. таблицу с двумя примерами источников), в том числе и с новыми их значениями, но число которых станет известным только после аннотирования всех 16 268 источников [14, 15].

До описания ОЗЛ-модели в форме обобщенной блок-схемы проиллюстрируем использование трех измерений (социализации, темпоральное и эксплицитное) на двух частных примерах. В первом примере (рис. 1) один лингвист анализирует источник и обнаруживает в нем концепт нового знания (обозначено как этап  $A_1$ , который завершается в момент времени  $t_1$ ), формирует для него дефиницию личностной рубрики в типологии (этап  $A_2$  завершается в момент времени  $t_2$ ), а затем согласовывает его с  $(M - 1)$  лингвистами, участвующими кроме него в пополнении типологии новыми рубриками (этап  $A_3$  завершается в момент времени  $t_3$ ) и затем  $M$  лингвистов совместно формируют для не-



Рис. 1 Четыре этапа первого примера обнаружения концептов

го дефиницию коллективной рубрики в типологии (этап  $A_4$  завершается в момент времени  $t_4$ ).

Во втором примере (рис. 2) сначала один лингвист анализирует источник и обнаруживает концепт нового знания (этап  $A_1$  завершается в момент времени  $t_1$ ), формирует для него дефиницию личностной рубрики типологии (этап  $A_2$  завершается в момент времени  $t_2$ ). Другой лингвист анализирует тот же источник, интерпретирует его иначе, чем первый (этап  $A_3$  завершается в момент времени  $t_3$ ), формирует дефиницию другой личностной рубрики типологии (этап  $A_4$  завершается в момент времени  $t_4$ ). Два лингвиста по-разному интерпретируют один и тот же источник, и поэтому две новые рубрики будут разными. Тогда они совместно формируют согласованный между ними концепт (этапы  $A_{5.1}$  и  $A_{5.2}$  завершаются в момент времени  $t_5$ ), а затем создают для него дефиницию согласованной рубрики типологии (этап  $A_6$  завершается в момент времени  $t_6$ ).

На рис. 1 и 2 показаны три измерения ОЗЛ-модели: два числовых (темпоральное и степень социализации) и одно номинативное (эксплицитность). В измерении эксплицитности показаны две номинативные подкатегории знания, определенные в спиральной модели: неявное и явное (выраженное). На рис. 2 в измерении социализации цифра 1 указывает на первого из двух лингвистов, а цифра 2 соответствует двум лингвистам, согласовавшим концепт, дефиницию и соответствующую ему коллективную рубрику типологии. Формирование коллективных концептов любым числом лингвистов обсуждается ниже. Для примера с двумя лингвистами рис. 2 показывает первые 6 точек для моментов времени окончания 7 этапов обнаружения и представления концептов нового знания ( $t_1-t_6$ ).

Из приведенных двух примеров видно, что в трех измерениях удается отобразить только последовательность этапов и моменты времени, но не процессы трансформации, поэтому на рис. 1 и 2 не



Рис. 2 Семь этапов второго примера обнаружения концептов



- (1) поиск в базе данных записи (которая может стать источником нового знания), необходимой для решения некоторой проблемы с применением онтологии (например, поиск предложения, содержащего немецкий глагол *sollen*, если целью обработки найденной записи ставится аннотирование его значения в контексте этого предложения с указанием рубрики этого значения, обозначающего понятие на конвенциональном уровне — этап *поиска*, выполняемый аналитиком, обозначен стрелкой с литерой *R*;
- (2) преобразование найденной записи из цифровой формы в перцептивную — этап обозначен словом «*Визуализация*» над стрелкой с литерой *R*;
- (3) выявление аналитиком семантического свойства найденной записи, соответствующего цели ее обработки, и его описание с использованием *рубрики* этого свойства, обозначающей *понятие (класс понятий)* на конвенциональном уровне онтологии (например, в предложении анализируется такое его семантическое свойство, как значение глагола *sollen* в этом предложении; тогда ищется рубрика, которая соответствует именно этому значению глагола; если рубрика такого понятия в онтологии найдена, то она включается в описание записи) — этап «*Концептуализация*» найденной записи;
- (4) если в процессе решения проблемы аналитиком с применением онтологии для некоторого семантического свойства найденной записи нет релевантной рубрики на конвенциональном уровне, то это может говорить о лакуне (например, в аннотируемом предложении для значения глагола *sollen*, с точки зрения аналитика, нет релевантной рубрики); тогда в описании найденной записи аналитик проставляет специальную метку неполноты онтологии (икс-тег — *x-tag*), которая фиксирует потенциальную понятийную лакуну; найденная запись и описание встретившегося в ней семантического свойства передаются для их анализа одному из экспертов, отвечающих за ведение онтологии, — этап понимания экспертом семантического свойства найденной записи обозначен словом «*Интернализация*»; результат этого процесса — *индивидуальное понимание* этого свойства экспертом;
- (5) если аналитик отразил семантическое свойство найденной записи в ее описании релевантной рубрикой конвенционального уровня (тогда в описании записи отсутствует метка неполноты онтологии, т. е. нет икс-тега), то эта найденная запись не может служить источником нового знания; ее описание с проставленной рубрикой преобразуется в цифровую форму и вводится в базу данных описаний — этот этап обозначен словом («*Оцифровка*»);
- (6) если же в описании записи аналитик проставил икс-тег для некоторого семантического свойства найденной записи, то это свойство эксперт сопоставляет с концептами рубрик онтологии, чтобы убедиться, что оно не представлено рубрикой на конвенциональном уровне; если эксперт, в отличие от аналитика, находит релевантную рубрику, то в описании найденной записи икс-тег он заменяет на эту рубрику и затем оно вводится в базу данных описаний (этап «*Экстернализация*» семантического свойства найденной записи рубрикой);
- (7) если эксперт не находит на конвенциональном уровне релевантную рубрику, то найденная запись, с его точки зрения, обладает семантическим свойством, которое не может быть представлено ни одной рубрикой; для его представления в описании новой рубрики на личностном уровне эксперт анализирует это семантическое свойство, на основе его анализа создает дефиницию новой рубрики («*Экстернализация*»), добавляет ее в свой личностный уровень онтологии (на рис. 3 добавление рубрики не показано);
- (8) в описании найденной записи икс-тег заменяется на эту новую личностную рубрику, оно вводится в базу данных описаний («*Оцифровка*») и осуществляется переход к п. 1 (поиск следующей записи, релевантной решаемой проблеме, например другое предложение с *sollen*, и так до их исчерпания).

Описание восьми перечисленных этапов включает шесть процессов трансформации (см. рис. 3): *интернализация*, *экстернализация*, *поиск*, *визуализация*, *оцифровка* и *концептуализация* (первые два есть в витке спирали), две базы данных и базу знаний с темпоральной онтологией и ретроспективой ее состояний.

Если итерационно формировать описания найденных записей, которые могут стать источниками нового знания, то в онтологии могут появляться новые индивидуальные рубрики, созданные экспертами. В общем случае для некоторых записей может быть создано по несколько разных описаний с разными рубриками. Рассмотрим в правой части рис. 3 этапы одной итерации, на которой *M* экспертов пытаются согласовать *L* описаний

одного источника с  $L$  разными индивидуальными рубриками понятий онтологии ( $1 \leq L \leq M$ , так как в общем случае не каждый из  $M$  экспертов обязательно создает свою рубрику на личностном уровне):

- (1) для некоторого источника нового знания из базы данных описаний извлекаются  $L$  описаний, созданных на его основе, но с разными новыми индивидуальными рубриками понятий онтологии (если  $L > 1$ ; при  $L = 1$  ищется одно описание, релевантное источнику, с одной новой рубрикой), а также сам источник из базы данных источников нового знания («Поиск»);
- (2) выполняется преобразование источника и  $L$  его описаний с новыми рубриками из цифровой формы в перцептивную («Визуализация»);
- (3) эксперты проводят семантический анализ личностных дефиниций рубрик онтологии, проставленных в  $L$  описаниях источника («Интернализация»);
- (4) эксперты анализируют семантическое свойство источника нового знания, сопоставляют с ним  $L$  его описаний (с разными индивидуальными рубриками) и пытаются согласовать дефиницию коллективной рубрики для семантического свойства этого источника («Социализация»);
- (5) если  $K$  экспертов ( $2 \leq K \leq M$ ) смогут согласованно создать дефиницию новой коллективной рубрики («Экстернализация»), то они вместе составляют коллективное описание источника с проставлением новой рубрики; в противном случае выполнение этой итерации прекращается («Останов»);
- (6) коллективная рубрика и ее дефиниция добавляются на коллективный уровень онтологии именно этих  $K$  экспертов, а новое описание источника с этой коллективной рубрикой вводится в базу данных описаний («Оцифровка»); затем осуществляется переход к п. 1 (извлекаются описания, созданные на основе другого источника нового знания).

В описании блок-схемы указаны аналитики и эксперты. В операции согласования («Социализация») принимают участие только эксперты. В блок-схему включено семь процессов трансформации: *интернализация, экстернализация, социализация, поиск, визуализация, оцифровка и концептуализация*. Первые три из них уже были в моделях из разд. 2. Наличие еще четырех процессов трансформации позволяет говорить о существенной степени новизны ОЗЛ-модели.

## 4 Заключение

В ОЗЛ-модели заполнение понятийных лагун на конвенциональном уровне онтологии обусловлено выполнением проблемно-ориентированной верификации ее полноты. В результате формируются еще два уровня в онтологии базы знаний. По завершении верификации появляется возможность выбирать уровень представления знания в процессе решения проблемы, а также сопоставлять ее решения, полученные на конвенциональном уровне и разными коллективами экспертов.

Отметим, что применение ОЗЛ-модели предполагает использование темпоральных онтологий, что дает возможность фиксировать и отображать динамику извлечения нового знания.

Выделим в ОЗЛ-модели те характерные черты, которые обеспечивают первые четыре направления ее развития, перечисленные в первом разделе статьи. Целенаправленность обнаружения лагун в онтологии обеспечивается процессом поиска (первое направление развития трех моделей из разд. 2). В ОЗЛ-модели процессы применения и пополнения онтологии являются *симбиотическими* (второе направление): без применения онтологии нельзя обнаружить лагуны (при концептуализации потенциальных источников) и заполнить их, а без регулярного пополнения онтологии новыми рубриками сфера ее применения всегда будет ограничена только уже существующим *конвенциональным знанием*.

В процессе поиска лагун используются потенциальные источники концептов нового знания, соответствующие проблемно-ориентированной цели применения онтологии (третье направление). Расширение числа сред до трех (четвертое направление) включает ментальную и информационную среды, а также цифровую среду компьютеров и сетей, что дает возможность проводить поиск источников нового знания в данных большого объема.

В заключение отметим, что перспективным направлением применения ОЗЛ-модели может стать проблемно-ориентированная верификация и пополнение существующих онтологий систем искусственного интеллекта на личностном и коллективном уровнях в процессе целенаправленного поиска источников нового знания в данных большого объема и их концептуализации. В настоящее время накопление таких данных наблюдается в широком спектре предметных областей. Использование данных большого объема для верификации онтологий и извлечения нового знания для их пополнения экспертами может быть реализовано на основе этой модели.

## Литература

1. *McGuinness D. L.* Ontologies come of age // Spinning the Semantic Web: Bringing the World Wide Web to its full potential / Eds. D. Fensel, J. Hendler, H. Lieberman, W. Wahlster. — Cambridge, MA, USA: MIT Press, 2003. P. 171–194.
2. *Zatsman I.* Goal-oriented creation of individual knowledge: Model and information technology // 19th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2018. Vol. 2. P. 947–956.
3. *Zatsman I.* Finding and filling lacunas in knowledge systems // 20th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2019. Vol. 2. P. 1143–1151.
4. *Nonaka I.* The knowledge-creating company // *Harvard Bus. Rev.*, 1991. Vol. 69. No. 6. P. 96–104.
5. *Nonaka I.* A dynamic theory of organizational knowledge creation // *Organ. Sci.*, 1994. Vol. 5. No. 1. P. 14–37.
6. *Nonaka И., Takeuchi Х.* Компания — создатель знания / Пер. с англ. — М.: Олимп-бизнес, 2003. 384 с. (*Nonaka I., Takeuchi H.* The knowledge-creating company. — Oxford, NY, USA: Oxford University Press, 1995. 284 p.)
7. *Wierzbicki A. P., Nakamori Y.* Basic dimensions of creative space // *Creative space: Models of creative processes for knowledge civilization age* / Eds. A. P. Wierzbicki, Y. Nakamori. — Berlin: Springer Verlag, 2006. P. 59–90.
8. *Wierzbicki A. P., Nakamori Y.* Knowledge sciences: Some new developments // *Z. Betriebswirt.*, 2007. Vol. 77. No. 3. P. 271–295.
9. *Nakamori Y.* Knowledge and systems science — enabling systemic knowledge synthesis. — London — New York: CRC Press, 2013. 234 p.
10. *Nissen M. E.* Harnessing knowledge dynamics: Principled organizational knowing & learning. — London: IRM Press, 2006. 278 p.
11. *Zatsman I., Buntman N., Coldefy-Faucard A., Nuriev V.* WEB Knowledge Base for asynchronous brainstorming // 17th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2016. Vol. 1. P. 976–983.
12. *Bratianu C.* A strategic view on the knowledge dynamics models used in knowledge management // 20th European Conference on Knowledge Management Proceedings Proceedings. — Reading: Academic Publishing International Ltd., 2019. Vol. 1. P. 185–192.
13. *Добровольский Д. О., Зализняк Анна А.* Немецкие конструкции с модальными глаголами и их русские соответствия: проект надкорпусной базы данных // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам Междунар. конф. «Диалог». — М.: РГГУ, 2018. Вып. 17(24). С. 172–184.
14. *Зацман И. М.* Стадии целенаправленного извлечения знаний, имплицитированных в параллельных текстах // *Системы и средства информатики*, 2018. Т. 28. № 3. С. 175–188.
15. *Зацман И. М.* Целенаправленное развитие систем лингвистических знаний: выявление и заполнение лакун // *Информатика и её применения*, 2019. Т. 13. Вып. 1. С. 91–98.
16. *Гончаров А. А., Зацман И. М.* Информационные трансформации параллельных текстов в задачах извлечения знаний // *Системы и средства информатики*, 2019. Т. 29. № 1. С. 180–193.
17. *Handbook of linguistic annotation* / Eds. N. Ide, J. Pustejovsky. — Dordrecht, The Netherlands: Springer Science + Business Media, 2017. 1459 p.
18. *Немецко-русский словарь: актуальная лексика* / Под ред. Д. О. Добровольского. — М.: Лексрус, 2020 (в печати).
19. Параллельный немецкий подкорпус Национального корпуса русского языка. <http://www.ruscorpora.ru/search-para-de.html>.
20. *Zatsman I.* Three-dimensional encoding of emerging meanings in AI-systems // 21st European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2020 (in press).

Поступила в редакцию 15.07.20

## PROBLEM-ORIENTED VERIFYING THE COMPLETENESS OF TEMPORAL ONTOLOGIES AND FILLING CONCEPTUAL LACUNAS

I. M. Zatsman

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

**Abstract:** An approach to verify the completeness of ontologies and fill conceptual lacunas found in them is proposed. The approach is based on the following symbiotic information processes: goal-oriented discovering new knowledge using data, its representation in an ontology, and application of the ontology to solve a problem. In the process of its solving, the completeness of the ontology is verified and its conceptual lacunas found during solving the problem are registered and filled. The personal, collective, and conventional levels of knowledge

representation in ontologies are discussed. The approach allows one to find conceptual lacunas at the conventional level of ontologies and fill them at their personal and/or collective levels, if for discovering new knowledge its potential sources are available. The purpose of the paper is to consider the model of symbiotic information processes. The developed model is a generalized flowchart that implements the proposed approach. The flowchart serves as the basis for computerization of symbiotic processes. The model description is illustrated by an example of finding conceptual lacunas in a linguistic typology and filling them with the concepts of new knowledge discovered using text data.

**Keywords:** three-level representation of knowledge; temporal ontology; conceptual lacuna; generation of new knowledge; symbiotic processes

**DOI:** 10.14357/19922264200317

## Acknowledgments

The reported study was funded by RFBR, project No. 18-07-00192.

## References

- McGuinness, D. L. 2003. Ontologies come of age. *Spinning the Semantic Web: Bringing the World Wide Web to its full potential*. Eds. D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster. Cambridge, MA: MIT Press. 171–194.
- Zatsman, I. 2018. Goal-oriented creation of individual knowledge: Model and information technology. *19th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 2:947–956.
- Zatsman, I. 2019. Finding and filling lacunas in knowledge systems. *20th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 2:1143–1151.
- Nonaka, I. 1991. The knowledge-creating company. *Harvard Bus. Rev.* 69(6):96–104.
- Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. *Organ. Sci.* 5(1):14–37.
- Nonaka, I., and H. Takeuchi. 1995. *The knowledge-creating company*. Oxford, NY: Oxford University Press. 284 p.
- Wierzbicki, A. P., and Y. Nakamori. 2006. Basic dimensions of creative space. *Creative space: Models of creative processes for knowledge civilization age*. Eds. A. P. Wierzbicki and Y. Nakamori. Berlin: Springer Verlag. 59–90.
- Wierzbicki, A. P., and Y. Nakamori. 2007. Knowledge sciences: Some new developments. *Z. Betriebswirt.* 77(3):271–295.
- Nakamori, Y. 2013. *Knowledge and systems science — enabling systemic knowledge synthesis*. London — New York: CRC Press. 234 p.
- Nissen, M. E. 2006. *Harnessing knowledge dynamics: Principled organizational knowing & learning*. London: IRM Press. 278 p.
- Zatsman, I., N. Buntman, A. Coldefy-Faucard, and V. Nuriev. 2016. WEB Knowledge Base for asynchronous brainstorming. *17th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 1:976–983.
- Bratianu, C. 2019. A strategic view on the knowledge dynamics models used in knowledge management. *20th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 1:185–192.
- Dobrovol'skiy, D. O., and Anna A. Zaluzniak. 2018. Nemetskie konstruktivnyye modal'nymi glagolami i ikh russkie sootvetstviya: proyekt nadkorporusnoy bazy dannykh [German constructions with modal verbs and their Russian correlates: A supracorpora database project]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: po mat-lam Mezhdunar. konf. "Dialog"* [Computer Linguistic and Intellectual Technologies: Conference (International) "Dialog" Proceedings]. Moscow. 17(24):172–184.
- Zatsman, I. 2018. Stadii tselenapravlennoy izvlecheniya znaniy, implitsirovannykh v parallel'nykh tekstakh [Stages of goal-oriented discovery of knowledge implied in parallel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(3):175–188.
- Zatsman, I. M. 2019. Tselenapravlennoye razvitie sistem lingvisticheskikh znaniy: vyyavlenie i zapolnenie lakun [Goal-oriented development of linguistic knowledge systems: Identifying and filling of lacunae]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):91–98.
- Goncharov, A. A., and I. M. Zatsman. 2019. Informatsionnye transformatsii parallel'nykh tekstov v zadachakh izvlecheniya znaniy [Information transformations of parallel texts in knowledge extraction]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(1):180–193.
- Ide, N., and J. Pustejovsky, eds. 2017. *Handbook of linguistic annotation*. Dordrecht, The Netherlands: Springer Science + Business Media. 1459 p.
- Dobrovol'skiy, D. O., ed. 2020 (in press). *Nemetsko-russkiy slovar': aktual'naya leksika* [German–Russian dictionary: Actual vocabulary]. Moscow: Leksrus.
- Parallel'nyy nemetskiy korpus [Parallel German corpus]. Available at: <http://www.ruscorpora.ru/search-para-de.html> (accessed July 20, 2020).

20. Zatsman, I. 2020 (in press). Three-dimensional encoding of emerging meanings in AI-systems. *21st European Con-*

*ference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd.

*Received July 15, 2020*

## 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

# ИСПОЛЬЗОВАНИЕ ТЕМАТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ ПАРНОГО СРАВНЕНИЯ КОЛЛЕКЦИЙ НАУЧНЫХ СТАТЕЙ

Ф. В. Краснов<sup>1</sup>, А. В. Диментов<sup>2</sup>, М. Е. Шварцман<sup>3</sup>

**Аннотация:** Авторами предложена новая методика для парного сравнения коллекций научных статей с помощью тематической модели. Разработанная методика получила название *сравнительного тематического анализа* (СТА). Сравнительный тематический анализ позволяет получать не только количественную оценку схожести коллекций, но и структурные различия сравниваемых коллекций как в количественном виде, так и с помощью средств визуализации, разработанных авторами. В данном исследовании проведено сравнение существующих подходов к тематическому моделированию применительно к рассматриваемой задаче сравнения коллекций научных статей. Рассмотрены вероятностные и генеративные тематические модели. Проведен анализ требований к текстовым коллекциям для корректного применения СТА. Методика СТА показала высокую эффективность на выделении структурных различий близких по тематике коллекций. Авторами разработана интегральная метрика, позволяющая сравнивать коллекции между собой: коэффициент контентной аутентичности. По результатам цифрового эксперимента наиболее информативной показала себя тематическая модель с аддитивной регуляризацией (ARTM, additive regularization of topic model).

**Ключевые слова:** сравнительный тематический анализ; анализ текстов; метрики тематической модели

**DOI:** 10.14357/19922264200318

## 1 Введение

Научные статьи представляют передовые идеи человечества в структурированной форме исследований и результатов. Но объемы производства научных текстов в мире растут экспоненциально [1]. Выбор оптимального набора журналов для наблюдения за развитием даже узкой области исследований стал сложной задачей для исследователей из-за большого потока журналов и для библиотекарей из-за постоянного роста стоимости подписки.

Анализ значительных объемов научных работ успешно выполняется средствами интеллектуального анализа текста (ИАТ, Text Mining). Роль тематического моделирования как инструмента для выделения латентных смыслов становится все более значительной [2] в ИАТ.

Тематическое моделирование в силу своих родовых особенностей обладает тремя недостатками.

1. Заранее не известно, какое число тематик задать модели, так как число тематик является свободным параметром.
2. В изначальной идее тематического моделирования нет механизмов настройки обучения на проблемный домен.

3. Полученные тематики нуждаются в интерпретации.

Интуитивная логика, которая стоит за подходом авторов, состоит в следующем: зачастую затраты на изучение коллекции научных журналов могут быть предприняты один раз, чтобы многократно получать дополнительную информацию, сравнивая уже известную коллекцию с новой.

Для реализации такого подхода необходимо иметь общий базис, в котором можно сравнивать коллекции. По гипотезе авторов, таким базисом может быть тематическая модель, обученная особым способом. Метаалгоритм для сравнения коллекций, включающий подготовку коллекций, создание и обучение тематической модели и визуализацию результатов сравнения, назван авторами методикой *сравнительного тематического анализа*.

## 2 Методика парного сравнения коллекций научных статей

Авторы рассматривают задачу парного сравнения коллекций документов в следующей постанов-

<sup>1</sup>Компания НАУМЕН, fkrasnov@naumen.ru

<sup>2</sup>Национальный электронно-информационный консорциум (NEICON), dimentov@neicon.ru

<sup>3</sup>Национальный электронно-информационный консорциум (NEICON); Российская государственная библиотека, shvar@neicon.ru

ке. Под коллекцией  $j$ , состоящей из  $M$  документов, будем понимать упорядоченную последовательность научных статей  $j = (p_0, \dots, p_M)$ . Научная статья  $p$  в дальнейшем изложении — это документ, например в формате дублинского ядра, состоящий из текста ( $d$ ) и метайнформации (авторов ( $a$ ), даты выпуска ( $y$ ), названия статьи ( $s$ ), названия журнала ( $jo$ )).

Так как коллекции обычно бывают разных размеров, как минимум содержат различное число документов, но могут иметь и разную структуру метайнформации, то возникает задача выравнивания коллекций для сравнения.

Для коллекций  $j_0$  и  $j_1$ , состоящих из  $M_0$  и  $M_1$  документов, введем пространство  $\mathbb{R}^{N \times M}$ , где  $M$  — это число документов для объединенной коллекции  $J = j_0 \cup j_1$ . Коллекция  $J$  будет содержать  $M = \dim(D_0 \cap D_1) \leq M_0 + M_1$  документов, где  $D_i = (d_0, \dots, d_{M_i})$ ,  $i \in (0, 1)$ . Словарь  $V$  для объединенной коллекции будет обладать размерностью  $N$ .

Основной недостаток такого подхода заключается в том, что в реальных коллекциях почти не встречаются одинаковые документы, поэтому матрицы  $j_i^{N \times M}$  будут иметь высокую степень разреженности. Устранить данный недостаток может более компактное представление документов в пространстве меньшей размерности. Например, с помощью выделения групп похожих документов можно выделить подпространство  $\mathbb{R}^C$  такое, что  $C \ll M$ , где  $C$  будет числом групп, на которые разбиты документы  $D$ .

## 2.1 Мягкая кластеризация текстов

Двумя основными неконтролируемыми методами обучения для группировки текстов служат кластеризация и тематическое моделирование. Задача кластеризации заключается в том, чтобы отнести каждый документ к одному определенному тематическому кластеру.

В тематическом моделировании используется вероятностная модель для определения мягкой кластеризации, в которой каждый документ имеет вероятность членства в кластере. Тематические модели можно рассматривать как процесс кластеризации с генеративной вероятностной моделью. Каждую тему можно рассматривать как распределение вероятностей по словам (матрица  $\varphi(t, V) \in \mathbb{R}^{C \times N}$ ), причем наиболее характерные слова имеют наибольшую вероятность. Каждый документ может быть выражен как вероятностная комбинация этих различных тем — матрица  $\theta(d, t) \in \mathbb{R}^{M \times C}$ . Таким образом, тема может рассматриваться как аналог кластера, а принадлежность документа к кластеру

носит вероятностный характер. Это также приводит к более элегантному представлению членства в кластере в тех случаях, когда известно, что документ содержит различные темы. В случае жесткой кластеризации иногда бывает сложно присвоить документ одному кластеру. Кроме того, тематическое моделирование элегантно связано с проблемой сокращения измерений, где каждая тема обеспечивает концептуальное измерение и документы могут быть представлены в виде линейной вероятностной комбинации этих различных тем. Таким образом, тематическое моделирование обеспечивает чрезвычайно общую структуру, которая относится как к проблемам кластеризации [3], так и к уменьшению размерности [4].

Развитие тематического моделирования начиная с исследования [5] сосредоточено на повышении информативности тематик. В работе [6] показано, что человеческая оценка интерпретируемости тематик хорошо коррелирует с автоматизированной мерой качества, называемой когерентностью тематики. В научной литературе используются следующие подходы к вычислению когерентности:  $C_{UCI}$  [7] и  $C_{UMass}$  [8]. Наиболее универсальным с методической точки зрения представляется  $C_{UCI}$ .

## 2.2 Методика сравнительного тематического анализа

Суть методики СТА, предложенной авторами, состоит в том, чтобы для группировки текстов использовать мягкую кластеризацию текстов (тематическую модель), обучение тематической модели производить на общем словаре и объединенной коллекции документов, а продуктивность сравнения измерять с помощью метрики *коэффициент контентной аутентичности* ( $\kappa$ ).

Для вычисления коэффициента контентной аутентичности авторы предложили использовать сумму модулей отклонений от равномерного распределения тематик, деленную на число тематик:

$$\kappa(j_0, j_1) = \frac{1}{C} \sum_{t \in C} \text{abs} \left[ \frac{\sum_{d \in j_0} \theta(d, t)}{\sum_{\hat{d} \in j_0} \sum_{\hat{t} \in C} \theta(\hat{d}, \hat{t})} - \frac{\sum_{d \in j_1} \theta(d, t)}{\sum_{\hat{d} \in j_1} \sum_{\hat{t} \in C} \theta(\hat{d}, \hat{t})} \right].$$

Максимальное значение  $\kappa$  равно единице, когда каждая из выявленных тематик относится только к одной из сравниваемых коллекций.

### 3 Эксперимент

#### 3.1 Подготовка данных

Для эксперимента было выбрано три коллекции научных журналов на русском языке — «Медицина-0» ( $j_0$ , 702 статьи), «Медицина-1» ( $j_1$ , 2599 статьи) и «Науки о Земле» ( $j_2$ , 1123 статьи).

В исходном виде коллекции находились в формате дублинского ядра (Dublin Core). Затем коллекции были приведены к формату Vowpal Wabbit<sup>1</sup>. По сути, это файл, в котором одна строка соответствует одному документу. Для объединенной коллекции число документов равно  $M = \dim(D) = 4424$ . Каждый документ был дополнен биграммами входящих в него слов. Такой подход к построению корпуса документов называется «Мешок слов» (Bag of Words, BoF [9]). На следующем шаге подготовки коллекции к анализу был создан словарь  $V$  объединенной коллекции, размерность которого составила  $N = \dim(V) = 2572622$ . Для очистки от шума к словарю были применены два фильтра: отброшены слова с частотой употребления меньше 40 и слова, которые встречались более чем в 60% документов. В очищенном словаре объединенной коллекции получилось  $N = \dim(V) = 37910$  слов (униграмм и биграмм). Полученный набор данных доступен в Mendeley Data.

#### 3.2 Выбор числа тематик

В исследовании [10] предложена методика для оценки числа тематик на основе качества полученных кластеров. Разработанная авторами [10] метрика качества cDBI имеет максимум при оптимальном значении  $C$ . На рис. 1 приведена зависимость метрики cDBI от  $C$  для объединенной коллекции  $J$ , из

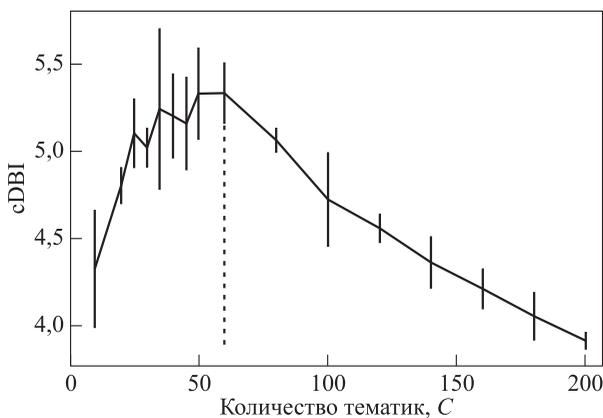


Рис. 1 Зависимость метрики cDBI от числа тематик  $C$

<sup>1</sup>[https://github.com/JohnLangford/vowpal\\_wabbit/wiki](https://github.com/JohnLangford/vowpal_wabbit/wiki).

которой видно, что максимум качества выделяемых кластеров достигается при значении  $C = 60$ .

#### 3.3 Выбор методики тематической модели

Для выбора методики тематической модели были рассмотрены методики ARTM [11], LDA (latent Dirichlet allocation) [12], BTM (biterm topic model) [13] и PLSA (probabilistic latent semantic analysis) [5]. Критерием для выбора методики модели стала метрика когерентности. Когерентность была вычислена для каждой из рассматриваемых моделей в виде зависимости от числа итераций обучения. Так как когерентность зависит от порядка документов, то для каждой итерации обучения было вычислено десять значений когерентности для коллекций с различным случайным порядком документов.

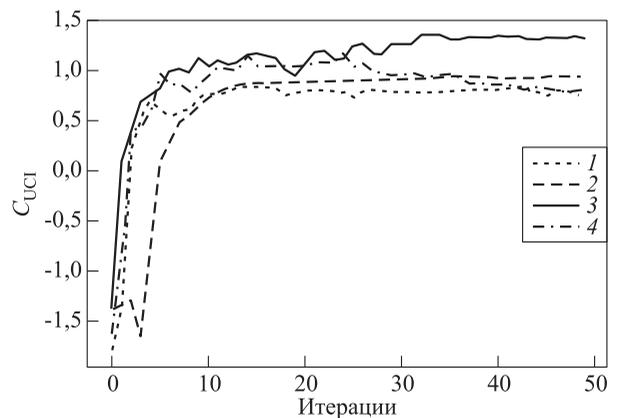


Рис. 2 Зависимость когерентности  $C_{UCI}$  от числа итераций обучения модели: 1 — LDA; 2 — BTM; 3 — ARTM; 4 — PLSA

На рис. 2 приведены зависимости  $C_{UCI}$  для рассматриваемых методик тематического моделирования. Сравнение зависимостей позволяет сделать вывод о том, что наиболее когерентные тематики могут быть получены с помощью методики ARTM.

### 4 Сравнительный анализ

Теперь, когда выбраны методика тематической модели и число тематик, можно приступить к испытанию методики сравнительного тематического анализа. Для этого обучим модель для объединенной коллекции  $J$ . В результате получатся матрицы  $\varphi(t, V)$  и  $\theta(d, t)$ .

### 4.1 Различные коллекции

Рассмотрим для сравнения две коллекции: «Медицина-1» ( $j_1$ ) и «Науки о Земле» ( $j_2$ ). При таких начальных условиях ожидается, что СТМ выделит разные наборы тематик. Это различие можно увидеть на гистограмме, изображенной на рис. 3, отображающей суммарный вес для каждой тематики по каждой из коллекций.

Из гистограммы, представленной на рис. 3, видно, что большинство столбиков имеют одинаковую штриховку — принадлежат к одной коллекции. Например, тематика *sbj9* полностью связана с «Науками о Земле». Слова с наибольшим весом в этой тематике следующие: «комплекс, порода, Урал, зона, массив, состав, базальт». С другой стороны, в тематике *sbj20* с наибольшими вероятностями будут только слова, относящиеся к «Медицине»: «препа-

рат, эффективность, эффект, доза, вес, боль, применение». Но будут и смешанные тематики, в которых с большими значениями вероятности будут присутствовать слова, относящиеся к обеим коллекциям. Количественной мерой расстояния (отличия) коллекций служит коэффициент контентной аутентичности. Для рассматриваемого случая различных по тематике коллекций  $\kappa(j_0, j_1) = 0,95$ .

### 4.2 Похожие коллекции

Результат, полученный в предыдущем разделе для заведомо различных коллекций, интересен как предельный случай. Но важно также понимать границы применимости СТМ (comparative text model) и чувствительности к изменениям в коллекциях. Поэтому для следующего эксперимента были взяты две коллекции журналов по направлению «Меди-

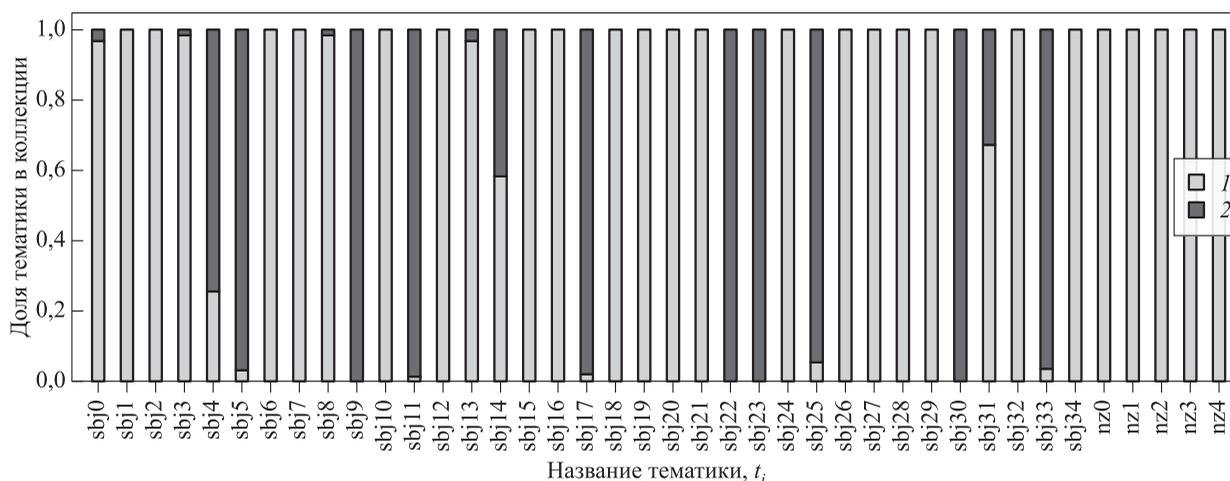


Рис. 3 Профили тематик для коллекций «Медицина-1» (1) и «Науки о Земле» (2)

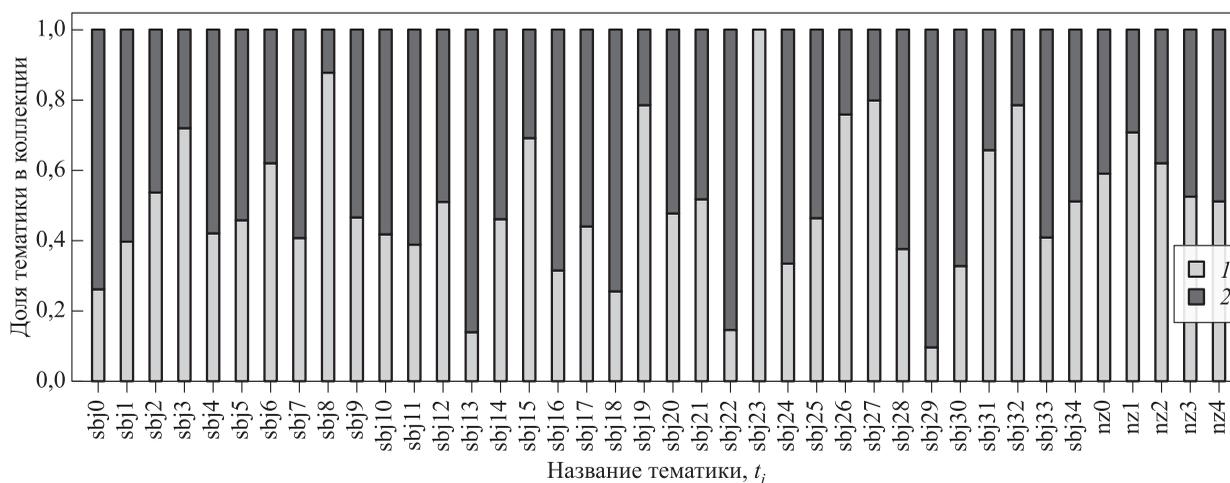


Рис. 4 Профили тематик для коллекций «Медицина-0» (1) и «Медицина-1» (2)

цина» с близкими тематиками — «Медицина-0» ( $j_0$ ) и «Медицина-1» ( $j_1$ ). При таких начальных условиях ожидается, что СТМ выделит схожие наборы тематик. Это можно увидеть на гистограмме, изображенной на рис. 4, отображающей суммарный вес для каждой тематики по каждой коллекции.

Из гистограммы, представленной на рис. 4, видно, что большинство столбиков имеют одинаковую штриховку примерно в одинаковой пропорции, другими словами, принадлежат к обеим коллекциям. Для рассматриваемого случая близких по тематике коллекций значение коэффициента контентной аутентичности  $\kappa(j_0, j_1) = 0,32$ . Так как коэффициент контентной аутентичности отражает дистанцию (разницу) между коллекциями, то чем ближе  $\kappa$  к нулю, тем более похожи коллекции.

## 5 Заключение

В данном исследовании авторы предложили методику для сравнения коллекций научных журналов (СТМ). Предложенная методика показала свою наглядность и количественную точность в ходе проведенного авторами эксперимента. В качестве метрики расстояния между коллекциями авторами предложен коэффициент контентной аутентичности. В таблице собраны значения коэффициента контентной аутентичности для трех исследуемых коллекций. Результаты, отображенные в таблице, полностью согласуются с теорией.

Коэффициент контентной аутентичности,  $\kappa$

$\kappa$	Медицина-0	Медицина-1	Науки о Земле
Медицина-0	0	0,32	0,94
Медицина-1	0,32	0	0,95
Науки о Земле	0,94	0,95	0

Авторы рассмотрели четыре наиболее популярных методики тематического моделирования: ARTM, BTM, PLSA и LDA. Проведенный эксперимент по выбору методики тематического моделирования показал, что по метрике когерентности  $C_{UCI}$  наиболее высокое значение когерентности у методики ARTM.

В процессе методического обзора авторами было обнаружено, что на основании имеющихся метрик когерентности тематических моделей не представляется возможным определить наиболее значимый параметр тематической модели — число тематик. Авторы подтвердили эту находку экспериментально. Действительно, метрика когерентности не имеет глобальных экстремумов. Важно отметить следующее: ни одна из рассмотренных метрик

не учитывает, что порядок документов влияет на тематики и на их когерентность. Авторы показали в ходе эксперимента, что метрики тематической модели необходимо считать для различного порядка документов и усреднять полученные значения. Эта методическая находка позволяет оперировать средними значениями метрик и учитывать возникающую статистическую ошибку. Также важно, что ни одна из рассмотренных метрик когерентности не учитывает вероятности тематик ( $\varphi$ ), а только выбирает топ-N тематик. Поэтому для определения числа тематик авторами была использована метрика cDBI, учитывающая качество получаемых кластеров, порядок документов и полные векторы тематик из матрицы  $\varphi$ .

## Литература

1. Verstak A., Acharya A., Suzuki H., et al. On the shoulders of giants: The growing impact of older articles. arXiv.org, 2014. arxiv:1411.0275.
2. Jelodar H., Wang Y., Yuan C., et al. Latent Dirichlet Allocation (LDA) and topic modeling: Models, applications, a survey // *Multimed. Tools Appl.*, 2019. Vol. 78. Iss. 11. P. 15169–15211. doi: 10.1007/s11042-018-6894-4.
3. Shravan K. B., Vadlamani R. LDA based feature selection for document clustering // *10th Annual ACM India Compute Conference Proceedings*. — Bhopal, India: ACM, 2017. P. 125–130.
4. Onan A., Bulut H., Korukoglu S. An improved ant algorithm with LDA-based representation for text document clustering // *J. Inform. Sci.*, 2017. Vol. 43. Iss. 2. P. 275–292.
5. Hofmann T. Probabilistic latent semantic analysis // *15th Conference on Uncertainty in Artificial Intelligence Proceedings*. — Stockholm, Sweden: Morgan Kaufmann Publishers Inc., 1999. P. 289–296.
6. Röder M., Both A., Hinneburg A. Exploring the space of topic coherence measures // *8th ACM Conference (International) on Web Search and Data Mining Proceedings*. — Shanghai: ACM, 2015. P. 399–408.
7. Newman D., Lau J. H., Grieser K., Baldwin T. Automatic evaluation of topic coherence // *Human Language Technologies: Annual Conference of the North American Chapter of the Association for Computational Linguistics*. — Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. P. 100–108.
8. Mimno D., Wallach H. M., Talley E., et al. Optimizing semantic coherence in topic models // *Conference on Empirical Methods in Natural Language Processing Proceedings*. — Edinburgh, Scotland: Association for Computational Linguistics, 2011. P. 262–272.
9. Zellig S. H. Distributional structure // *Word*, 1954. Vol. 10. Iss. 2-3. P. 146–162.

10. Krasnov F., Sen A. The number of topics optimization: Clustering approach // Machine Learning Knowledge Extraction, 2019. Vol. 1. Iss. 1. P. 416–426.
11. Vorontsov K., Potapenko A. Additive regularization of topic models // Mach. Learn., 2015. Vol. 101. Iss. 1-3. P. 303–323.
12. Blei D. M., Ng A. Y., Jordan M. I. Latent Dirichlet allocation // J. Mach. Learn. Res., 2003. Vol. 3. No. 1. P. 993–1022.
13. Cheng X., Yan X., Lan Y., Guo I. BTM: Topic modeling over short texts // IEEE T. Knowl. Data En., 2014. Vol. 26. Iss. 12. P. 2928–2941.

Поступила в редакцию 27.06.19

## USING TOPIC MODELS FOR PAIRWISE COMPARISON OF COLLECTIONS OF SCIENTIFIC PAPERS

F. V. Krasnov<sup>1</sup>, A. V. Dimentov<sup>2</sup>, and M. E. Shvartsman<sup>2,3</sup>

<sup>1</sup>NAUMEN R&D, 49A Tatishcheva Str., Ekaterinburg 620028, Russian Federation

<sup>2</sup>National Electronic Information Consortium, 5 Letnikovskaya Str., Moscow 115114, Russian Federation

<sup>3</sup>Russian State Library, 3/5 Vozdvigenka Str., Moscow 119019, Russian Federation

**Abstract:** The authors propose a new technique for pairwise comparison of collections of scientific articles via a topic model. The developed methodology is called Comparative Topic Analysis (CTA). Comparative topic analysis allows getting not only quantitative assessment of similarity of collections but also structural differences of the compared text collections. The authors developed transparent visualization for text collections distance. This study compares existing approaches to topic modeling concerning the task of comparing collections of scientific papers. The authors consider probabilistic and generative topic models. The analysis of the requirements for text collections for the correct application of CTA was carried out. The CTA methodology has shown high efficiency in identifying structural differences in related collections. The authors developed an integral metric “Content Uniqueness Ratio” which allows comparing text collections with each other. As a result of the digital experiment, the thematic model with additive regularization (ARTM) proved to be the most informative.

**Keywords:** comparative topic analysis; comparative text model; deep text analysis; topic models metrics

**DOI:** 10.14357/19922264200318

## References

1. Verstak, A., A. Acharya, H. Suzuki, *et al.* 2014. On the shoulders of giants: The growing impact of older articles. arXiv.org. Available at: <http://arxiv.org/abs/1411.0275> (accessed July 1, 2019).
2. Jelodar, H., Y. Wang, C. Yuan, *et al.* 2019. Latent Dirichlet Allocation (LDA) and topic modeling: Models, applications, a survey. *Multimed. Tools Appl.* 78(11):15169–15211. doi: 10.1007/s11042-018-6894-4.
3. Shravan, K. B., and R. Vadlamani. 2017. LDA based feature selection for document clustering. *10th Annual ACM India Compute Conference Proceedings*. ACM. 125–130.
4. Onan, A., H. Bulut, and S. Korukoglu. 2017. An improved ant algorithm with LDA-based representation for text document clustering. *J. Inform. Sci.* 43(2):275–292.
5. Hofmann, T. 1999. Probabilistic latent semantic analysis. *15th Conference on Uncertainty in Artificial Intelligence Proceedings*. Stockholm, Sweden: Morgan Kaufmann Publishers Inc. 289–296.
6. Röder, M., A. Both, and A. Hinneburg. 2015. Exploring the space of topic coherence measures. *8th ACM Conference (International) on Web Search and Data Mining Proceedings*. Shanghai: ACM. 399–408.
7. Newman, D., J.H. Lau, K. Grieser, and T. Baldwin. 2010. Automatic evaluation of topic coherence. *Human Language Technologies: Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics. 100–108.
8. Mimno, D., H. M. Wallach, E. Talley, *et al.* 2011. Optimizing semantic coherence in topic models. *Conference on Empirical Methods in Natural Language Processing Proceedings*. Edinburgh, Scotland: Association for Computational Linguistics. 262–272.
9. Zellig, H. S. 1954. Distributional structure. *Word* 10(2-3):146–162.
10. Krasnov, F., and A. Sen. 2019. The number of topics optimization: Clustering approach. *Machine Learning Knowledge Extraction* 1(1):416–426.
11. Vorontsov, K., and A. Potapenko. 2015. Additive regularization of topic models. *Mach. Learn.* 101(1-3):303–323.

12. Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. Latent Dirichlet allocation. *J. Mach. Learn. Res.* 3(1):993–1022.
13. Cheng, X., X. Yan, Y. Lan, and I. Guo. 2014. Btm: Topic modeling over short texts. *IEEE T. Knowl. Data En.* 26(12):2928–2941.

*Received June 27, 2019*

## Contributors

**Krasnov Fedor V.** (b. 1969) — Candidate of Science (PhD) in technology, Director, Department of Information Management Systems, NAUMEN R&D, 49A Tatishcheva Str., Ekaterinburg 620028, Russian Federation; fkrasnov@naumen.ru

**Dimentov Alexander V.** (b. 1975) — Head of Department, National Electronic Information Consortium, 5 Letnikovskaya Str., Moscow 115114, Russian Federation; dimentov@neicon.ru

**Shvartsman Mikhail E.** (b. 1954) — Deputy Director, National Electronic Information Consortium, 5 Letnikovskaya Str., Moscow 115114, Russian Federation; Head of Department, Russian State Library, 3/5 Vozdvigenka Str., Moscow 119019, Russian Federation; shvar@neicon.ru

**Агасандян Геннадий Аршавинович** (р. 1941) — доктор физико-математических наук, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Агеев Кирилл Анатольевич** (р. 1993) — аспирант кафедры прикладной математики и теории вероятностей Российской университета дружбы народов

**Бетелин Владимир Борисович** (р. 1946) — доктор физико-математических наук, профессор, академик РАН, научный руководитель Федерального научного центра «Научно-исследовательский институт системных исследований Российской академии наук»

**Босов Алексей Вячеславович** (р. 1969) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор кафедры «Теория вероятностей и компьютерное моделирование» Московского авиационного института (национального технического университета)

**Горшенин Андрей Константинович** (р. 1986) — кандидат физико-математических наук, доцент, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; ведущий научный сотрудник факультета вычислительной математики и кибернетики Московского центра фундаментальной и прикладной математики Московского государственного университета имени М. В. Ломоносова

**Грушо Александр Александрович** (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Грушо Николай Александрович** (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Диментов Александр Владимирович** (р. 1975) — начальник отдела Национального электронно-информационного консорциума

**Донской Владимир Иосифович** (р. 1948) — доктор физико-математических наук, профессор, заведу-

ющий кафедрой информатики Крымского федерального университета им. В. И. Вернадского

**Забезайло Михаил Иванович** (р. 1956) — доктор физико-математических наук, доцент, главный научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Зацман Игорь Моисеевич** (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Королев Виктор Юрьевич** (р. 1954) — доктор физико-математических наук, профессор, заведующий кафедрой математической статистики факультета вычислительной математики и кибернетики, главный научный сотрудник Московского центра фундаментальной и прикладной математики Московского государственного университета имени М. В. Ломоносова; ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Краснов Федор Владимирович** (р. 1969) — кандидат технических наук, директор Департамента информационных систем управления компании НАУМЕН

**Кудрявцев Алексей Андреевич** (р. 1978) — кандидат физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

**Кушниренко Анатолий Георгиевич** (р. 1944) — кандидат физико-математических наук, заведующий отделом Федерального научного центра «Научно-исследовательский институт системных исследований Российской академии наук»

**Леонов Александр Георгиевич** (р. 1961) — кандидат физико-математических наук, доцент, ведущий научный сотрудник механико-математического факультета Московского государственного университета им. М. В. Ломоносова; заведующий сектором Федерального научного центра «Научно-исследовательский институт системных исследований Российской академии наук»; профессор, старший на-

учный сотрудник Московского педагогического государственного университета

**Малашенко Юрий Евгеньевич** (р. 1946) — доктор физико-математических наук, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Мартюшова Янина Германовна** (р. 1970) — кандидат педагогических наук, старший преподаватель кафедры «Теория вероятностей и компьютерное моделирование» Московского авиационного института (национального технического университета)

**Милованова Татьяна Александровна** (р. 1977) — кандидат физико-математических наук, старший преподаватель кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

**Назарова Ирина Александровна** (р. 1966) — кандидат физико-математических наук, научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Наумов Андрей Викторович** (р. 1966) — доктор физико-математических наук, профессор кафедры «Теория вероятностей и компьютерное моделирование» Московского авиационного института (национального технического университета)

**Наумов Валерий Арсентьевич** (р. 1950) — кандидат физико-математических наук, научный руководитель Исследовательского института инноваций (Финляндия, Хельсинки)

**Новиков Даниил Андреевич** (р. 1993) — соискатель ученой степени кандидата наук Национального исследовательского университета «Высшая школа экономики»

**Попов Михаил Владимирович** (р. 1995) — аспирант Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Посыпкин Михаил Анатольевич** (р. 1974) — доктор физико-математических наук, доцент, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Разумчик Ростислав Валерьевич** (р. 1984) — кандидат физико-математических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Самуйлов Константин Евгеньевич** (р. 1955) — доктор технических наук, профессор, заведующий кафедрой прикладной информатики и теории вероятностей Российского университета дружбы народов; старший научный сотрудник Института проблем информатики Федерального исследо-

вательского центра «Информатика и управление» Российской академии наук

**Сапунова Анастасия Павловна** (р. 1998) — студентка магистратуры Московского авиационного института (национального технического университета)

**Смирнов Дмитрий Владимирович** (р. 1984) — бизнес-партнер по ИТ Департамента безопасности ПАО «Сбербанк России»

**Сопин Эдуард Сергеевич** (р. 1987) — кандидат физико-математических наук, доцент Российского университета дружбы народов; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Терехина Ирина Юрьевна** (р. 1994) — аспирантка факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

**Тимонина Елена Евгеньевна** (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Шварцман Михаил Ефремович** (р. 1954) — заместитель директора Национального электронно-информационного консорциума; начальник отдела Российской государственной библиотеки

**Шестаков Олег Владимирович** (р. 1976) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**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»

**Щербинина Анастасия Андреевна** (р. 1997) — магистр факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова Российской академии наук

**Яркина Наталья Викторовна** (р. 1979) — кандидат физико-математических наук, доцент кафедры прикладной математики и теории вероятностей Российского университета дружбы народов

---

## Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

---

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам. Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.
5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или ЛАТЭХ (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 10 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на **русском и английском языках**:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:  
[http://www.ipiran.ru/journal/issues/2013\\_07\\_01\\_rus/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01_rus/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 hozyaystvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. 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 Publ. 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. — N.Y.: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

**Описание неопубликованного документа:**

Laturov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

**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]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

**Описание ГОСТа:**

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie rashkoda 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 Publ. 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: [rust@ipiran.ru](mailto:rust@ipiran.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. *Rus. 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 Publ. 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]. PhD Thesis. Moscow: IPI RAN. 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 Publ. 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: [rust@ipiran.ru](mailto:rust@ipiran.ru) (to Svetlana Strigina)  
<http://www.ipiran.ru/english/journal.asp>